Multivariate Behavioral Research, 50:316–333, 2015

ISSN: 0027-3171 print / 1532-7906 online
DOI: 10.1080/00273171.2014.1003770

Routledge
Taylor & Francis Group

Bayesian Causal Mediation Analysis for Group Randomized Designs with Homogeneous and Heterogeneous Effects: Simulation and Case Study

Soojin Park and David Kaplan

Department of Educational Psychology, University of Wisconsin–Madison

A fully Bayesian approach to causal mediation analysis for group-randomized designs is presented. A unique contribution of this article is the combination of Bayesian inferential methods with G-computation to address the problem of heterogeneous treatment or mediator effects. A detailed simulation study shows that this approach has excellent frequentist properties, particularly in the case of small sample sizes with accurate informative priors. The simulation study also demonstrates that the proposed approach can take into account heterogeneous treatment or mediator effects without bias. A case study using data from a school-based randomized intervention designed to increase parent social capital leading to improved behavioral and academic outcomes in children is offered to illustrate the Bayesian approach to causal mediation in group-randomized designs.

INTRODUCTION

This article is motivated by the following problem. Consider an intervention that is administered at the school level but where the variables of interest are measured at the within-school (child, parent) level. In the example that will serve as the case study for this article, interest centers on a randomized intervention referred to as Families and Schools Together (FAST; L. McDonald & Frey, 1999; McDonald, 2002). The FAST intervention is designed to increase parent-parent and parent-school social capital among low-income families. The outcome of interest is improved behavioral and academic outcomes for the children that attend the school. The design of the intervention involves random assignment of schools to the FAST intervention or to a control. The mediator, improved social capital, is measured at the parent level, and the evaluation includes important variables that mediate between the intervention and the outcome. Thus, the implementation of the intervention follows a group-randomized design.

Two aspects of the above mentioned study serve to motivate the purposes of this article. First, in order to establish the causal impact of the treatment on the outcome via a mediator, it is important to establish if the treatment or mediator effect can be assumed to be constant or whether it is more sensible to assume that the treatment or mediator effect varies across individuals. This is the problem of *homogeneous* versus *heterogeneous* effects, respectively. Second, it is important to account for the multilevel structure of the data inherent in group-randomized experiments.

The goal of this article, therefore, is to develop causal mediation analysis for group-randomized designs that can account for homogeneous and heterogeneous treatment or mediator effects. We adopt a Bayesian framework for the specification and estimation of model parameters insofar as Bayesian inference allows for the incorporation of prior knowledge regarding the distribution of model parameters into the analysis.

The organization of this article is as follows. In the next section, we provide a brief overview of conventional causal mediation analysis following closely the work of Imai and his colleagues (Imai, Keele, & Yamamoto, 2010; Imai, Keele, & Tingley, 2010; Imai, Keele, Tingley, & Yamamoto, 2011). Next, we propose our fully Bayesian approach to multilevel causal mediation analysis, including our proposed method addressing heterogeneous treatment or mediator effects. This is followed by a simulation study. Next, we provide a case study using data from the FAST intervention. The article closes with conclusions.

Correspondence concerning this article should be addressed to Soojin Park, Department of Educational Psychology, University of Wisconsin–Madison, Madison, WI 53706. E-mail: spark256@wisc.edu

Brief Review of Single-Level Causal Mediation Analysis

The problem of estimating direct and indirect effects has been an integral part of the history of structural equation modeling going back to Goldberger and Duncan (1972), and certainly with the seminal work of Duncan (1975). The work of Duncan and others showed how one could convert the structural form of a model into the so-called *reduced form* of the model, solve issues of identification and estimation, and return the values of the structural parameters as well as the indirect and total effects. An important overview of mediating effects based on conventional path analysis can be found in MacKinnon (2008).

In their classic paper, Baron and Kenny (1986) offered a simple approach to the estimation of direct and indirect effects, which rests on the assumption of linear relationships among the variables and homogeneous mediation effects. With the advent of full information methods such as full information maximum likelihood and generalized least squares, it is, in principle, unnecessary to use the stepwise approach of Baron and Kenny (1986). However, despite the ability to estimate direct and indirect effects simultaneously, the Baron and Kenny approach remains quite popular.

Perhaps owing to the need for a more detailed "unpacking" of interventions, recent work by Imai and his colleagues (see Imai, Keele, & Yamamoto, 2010; Imai, Keele, & Tingley, 2010; Imai, Keele, Tingley, & Yamamoto, 2011), combined the classical Baron and Kenny (1986) approach to mediation analysis in the structural equation modeling framework with the potential outcomes framework of Rubin (1974), under the rubric of *causal mediation analysis*. Unlike the Baron and Kenny (1986) approach, Imai and his colleagues gave a formal causal interpretation to mediation analysis.

The estimation approaches of Imai and his colleagues (Imai, Keele, & Yamamoto, 2010; Imai, Keele, & Tingley, 2010; Imai, Keele, Tingley, & Yamamoto, 2011) have been based primarily on nonparametric or frequentist-based parametric methods. Of course, frequentist-based parametric procedures for structural equation modeling, including the Baron and Kenny approach, have been available for several decades (see Kaplan, 2009). More recently, Bayesian approaches have become available for structural modeling (see e.g., Kaplan & Depaoli, 2012; Lee, 2007; Muthén & Asparouhov, 2012; Song & Lee, 2012; Yuan & MacKinnon, 2009), owing to the development and application of Markov chain Monte Carlo estimation methods (see e.g., Geman & Geman, 1984; Gilks, Richardson, & Spiegelhalter, 1996).

Regardless of the estimation approach, any desire to unpack the mediating mechanisms of school-based interventions must account for the fact that the organizational structure of schooling consists of students and/or teachers nested in schools. Indeed, in many instances, the substantive problem concerns understanding the role that units at both levels play in some outcome of interest. In the example that we will provide below, the intervention is provided at the school level, with the outcome being measured at the within-school level. Such data collection plans are generically referred to as *clustered sampling designs*. Data from clustered sampling designs are then collected at both levels for the purposes of understanding each level separately, but also to understand the inputs and processes of student and school level variables as they predict both school-and student-level outcomes. The need to properly analyze data arising from clustered sampling designs has led to the development of *multilevel models* (see e.g., Burstein, 1980; Goldstein, 2011; Raudenbush & Bryk, 2002). Extensions of multilevel models to structural equation modeling have been given in Schmidt (1969), Muthén (1989), Goldstein and McDonald (1988), McDonald and Goldstein (1989), McDonald (1993), and McDonald (1994).

Extensions of multilevel models to causal mediation analysis have been given in Reardon and Raudenbush (2013), and VanderWeele, Hong, Jones, and Brown (2013). Reardon and Raudenbush (2013) explicated assumptions needed to identify the average treatment effect in the context of multi-site randomized trials. However, the Reardon and Raudenbush (2013) paper should be understood as a generalization of the instrumental variable approach rather than mediation analysis, as the direct effect of the treatment on the outcome is assumed to be zero by the exclusion restriction assumption. VanderWeele, Hong, Jones, and Brown (2013) explicated assumptions needed to identify the average causal mediation effect (ACME) with multiple mediators under the multilevel setting. Their study illustrates the case where the aggregated mediator is included in the between-level model, which requires the *sequential ignorability assumption for multiple mediators* to identify the ACME.

In addition to capturing the multilevel features of most educational interventions, an additionally important problem is that of capturing heterogeneous treatment or mediator effects. Specifically, as mentioned earlier, conventional approaches to mediation analysis such as the Baron and Kenny (1986) approach assume that the treatment or mediator effect is constant for all individuals. However, in many cases, it may be more reasonable to assume that the mediation effect varies with values of observed and unobserved covariates.

Potential Outcomes Framework: Notation and Definitions

In this section, we provide a brief introduction to the potential outcomes framework of Rubin (1974). This is followed by a discussion of single-level causal mediation analysis following closely the work of Imai, Keele, and Tingley (2010) and Imai et al. (2011). This is then followed by the proposed theoretical steps necessary to estimate the posterior distribution of the outcome under causal mediation analysis.

To begin, let $T_i \in \{0, 1\}$ be a binary treatment indicator for individual i, let $M_i \in M$ be a mediator, let $Y_i \in Y$ be an outcome variable, and let $X_i \in X$ be observed treatment

covariates. Then, we can write the potential mediator as $M_i(t)$ and the potential outcome as $Y_i(t, m)$. The term $M_i(t)$ denotes the potential mediator for unit i when $T_i = t$; $Y_i(t, m)$ denotes the potential outcome for unit i when $T_i = t$ and $M_i = m$.

Under the potential outcomes framework, Imai, Keele, and Tingley (2010) define the unit-level causal mediation effect under t as

$$\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)), \tag{1}$$

where $t \in \{0, 1\}$. Equation (1) represents how the outcome is changed when the mediator is changed while holding the treatment constant at $T_i = t$. Pearl (2000) denotes Equation (1) as the *unit-level natural indirect effect* because the mediator takes on the value that it would naturally assume under the assigned treatment status. In other words, the parameter $\delta_i(t)$ is the change in the potential outcome for individual i in response to a change in the mediator from the value that would have been obtained under treatment status $t = 1$ to the value that would have been obtained under treatment status $t = 0$ while fixing treatment status to t.

The unit-level causal natural direct effect under t is written as

$$\zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t)), \tag{2}$$

which represents the causal effect of the treatment on the outcome for individual i holding the mediator at its potential value when $T_i = t$.

The unit-level total causal effect represents how a treatment-induced change in the outcome is realized both through and not through the mediator. The unit-level total causal effect can be expressed as the combination of the unit-level causal mediation effect under t and the unit-level causal natural direct effect under $(1 - t)$ and is written as

$$\tau_i = \delta_i(t) + \zeta_i(1 - t) \tag{3}$$

where $t \in \{0, 1\}$. Then, the unit-level total causal effect is defined as $Y_i(1, M_i(1)) - Y_i(0, M_i(0)$ after plugging Equations (1) and (2) into Equation (3).[1]

From here, Imai, Keele, Tingley, and Yamamoto (2011) define the ACME and the average natural direct effect (ANDE) as

$$\begin{aligned}\bar{\delta}(t) &\equiv E(\delta_i(t)) = E\{Y_i(t, M_i(1)) - Y_i(t, M_i(0))\} \\ \bar{\zeta}(t) &\equiv E(\zeta_i(t)) = E\{Y_i(1, M_i(t)) - Y_i(0, M_i(t))\}\end{aligned} \tag{4}$$

where expectations are averaged over individuals.

[1]To verify the relationship, note that

$$\begin{aligned}\tau_i &= \delta_i(1) + \zeta_i(0) \\ &= Y_i(1, M_i(1)) - Y_i(1, M_i(0)) \\ &\quad + Y_i(1, M_i(0)) - Y_i(0, M_i(0)) \\ &= Y_i(1, M_i(1)) - Y_i(0, M_i(0))\end{aligned}$$

Identification

In line with the fundamental problem of causal inference (Holland, 1986), we can observe $Y_i(t, M_i(t))$ but not $Y_i(t, M_i(t'))$ [see Equations (1) and (2)]. Thus, to solve this identification problem we need to invoke additional assumptions. The main assumption needed to identify the ACME is referred to as *sequential ignorability* and is an extension of the strong ignorability assumption of Rubin (1974). Formally, the sequential ignorability assumption begins by requiring that measures on observed covariates be obtained prior to treatment assignment. Then, the sequential ignorability assumption can be written as

$$\begin{aligned}&\{Y_i(t', m), M_i(t)\} \perp T_i | X_i = x \\ &Y_i(t', m) \perp M_i(t) | T_i = t, X_i = x\end{aligned} \tag{5}$$

for $t \in \{0, 1\}$ and $t' = (1 - t)$. The first part of the sequential ignorability assumption implies that given the observed indicators, there is no confounding between the treatment and mediator as well as the outcome. The second part of the sequential ignorability assumption implies that there is no confounding between the treatment and mediator given observed treatment status and observed indicators. In addition, there should not be a post treatment variable that is affected by the treatment and that also affects both the mediator and outcome.

Under these assumptions, the ACME is nonparametrically identified and given by Imai, Keele, and Yamamoto (2010), and Imai, Keele, Tingley, and Yamamoto (2011) as,

$$\begin{aligned}\bar{\delta}(t) = \int\int & E(Y_i | M_i = m, T_i = t, X_i = x) \\ & \{dF_{M_i|T_i=1, X_i=x}(m) - dF_{M_i|T_i=0, X_i=x}(m)\} dF_{X_i}(x).\end{aligned} \tag{6}$$

The average natural direct effect is identified as

$$\begin{aligned}\bar{\zeta}(t) = \int\int & \{E(Y_i | M_i = m, T_i = 1, X_i = x) \\ & - E(Y_i | M_i = m, T_i = 0, X_i = x)\} \\ & dF_{M_i|T_i=t, X_i=x}(m) dF_{X_i}(x)\end{aligned} \tag{7}$$

where $F_x(\cdot)$ and $F_{M|T}(\cdot)$ represent the distribution function of a random variable X and conditional distribution function of M given T.

An Aside: Linear Structural Equation Modeling

It is useful to examine the classical linear structural equation modeling (LSEM) approach to mediation in light of the causal mediation approach of Imai, Keele, Tingley, and Yamamoto (2011). In particular, the concern focuses on the conditions under which the LSEM approach to mediation leads, to valid causal inferences. Recall from Baron and Kenny (1986) that the mediation effect without an interaction

between the treatment and the mediator can be estimated based on following equations:

$$\begin{aligned} Y_i &= \alpha_1 + \beta_1 T_i + \xi_1' X_i + \epsilon_{i1} \\ M_i &= \alpha_2 + \beta_2 T_i + \xi_2' X_i + \epsilon_{i2} \\ Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \xi_3' X_i + \epsilon_{i3}. \end{aligned} \tag{8}$$

The question concerns whether the usual product-of-coefficients estimator $\hat{\beta}_2\hat{\gamma}$ is a proper estimate of the causal mediation effect. As pointed out by Imai, Keele, Tingley, and Yamamoto (2011), if we assume sequential ignorability as well as no interaction—namely that $\hat{\delta}(1) = \hat{\delta}(0)$, then $\hat{\beta}_2\hat{\gamma}$ is a valid estimate of the causal mediation effect under the assumption of linear relations among the variables.

Estimation

Under the sequential ignorability assumption, parametric estimation of the ACME is straightforward. One may follow the proposal of Baron and Kenny (1986), namely estimating the ACME as $\hat{\delta}(t) = \hat{\beta}_2\hat{\gamma}$ by fitting two separate linear regressions given in Equations (8). The standard error for the ACME can be calculated using the Delta method (Sobel, 1982), that is, $\text{Var}[\hat{\delta}(t)] \approx {\beta_2}^2 \, Var(\hat{\gamma}) + \gamma^2 \, Var(\hat{\beta}_2)$, or the variance formula of Goodman (1960), that is, $\text{Var}[(\hat{\delta}(t)] = {\beta_2}^2 \, Var(\hat{\gamma}) + \gamma^2 \, Var(\hat{\beta}_2) + Var(\hat{\gamma}) \, Var(\hat{\beta}_2)$. The average natural indirect and total effects and their standard errors can be estimated directly from the fitted linear regressions given in Equation (8).

Several nonparametric estimators for the ACME have also been proposed. One may use the simple nonparametric estimator given in Equation (6). By the law of large numbers, this estimator asymptotically converges to the true ACME under the sequential ignorability assumption. The second approach uses nonparametric regression to model $\mu_{tm} \equiv E(Y_i|T_i = t, M_i = m, X_i = x)$ and $\nu_{tm} \equiv Pr(M_i = m|T_i = t, X_i = x)$, and then employs the following estimator:

$$\hat{\delta}(t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{m=0}^{J-1} \hat{\mu}_{tm}(X_i)(\hat{\nu}_{1m}(X_i) - \hat{\nu}_{0m}(X_i)) \tag{9}$$

where M is discrete and takes J distinct values, that is, $M = \{0, 1, \ldots J-1\}$ and n is the number of subjects. When the mediator is not discrete, the ACME can be estimated by employing the following estimator after modeling $\hat{\mu}_{tm}(x) \equiv E(Y_i|T_i = t, M_i = m, X_i = x)$ and $\psi_t(x) = p(M_i|T_i = t, X_i = x)$.

$$\hat{\delta}(t) = \frac{1}{nL} \sum_{i=1}^{n} \sum_{l=1}^{L} \hat{\mu}_{t\tilde{m}_{1i}^{(l)}}(X_i) - \hat{\mu}_{t\tilde{m}_{0i}^{(l)}}(X_i), \tag{10}$$

where $\tilde{m}_{1i}^{l}$ is the lth Monte Carlo draw of the mediator M_i from its predicted distribution based on the fitted model $\hat{\psi}_t(X_i)$. Since there is no corresponding asymptotic variance estimator, nonparametric bootstrapping can be used to estimate the corresponding standard error.

BAYESIAN MULTILEVEL CAUSAL MEDIATION ANALYSIS

In this section, we propose a fully Bayesian approach to multilevel causal mediation analysis. We will first present extended identification results for multilevel causal mediation models for two cases: (1) when a mediator is at the student level and (2) when mediators are at both the student and school levels. Causal structural models for both cases are shown in Figures 1 and 2. Then, we provide the estimation method under the Bayesian framework.

Identification

As before, let $i = 1, 2, \ldots N$ students, and $j = 1, 2, \ldots J$ schools. In the case where the intervention is implemented at the school level, with the mediator and outcome being measured at the within-school level, the ACME is defined the same as Imai, Keele, Tingley, and Yamamoto (2011); namely,

$$\bar{\delta}(t) = E\{Y_{ij}(t, M_{ij}(1)) - Y_i(t, M_{ij}(0))\} \tag{11}$$

where $t \in 0, 1$. The ACME is identified under the following set of assumptions:

$$\begin{aligned} &\{Y_{ij}(t', m), M_{ij}(t)\} \perp T_j | X_{ij} = x \\ &Y_{ij}(t', m) \perp M_{ij}(t) | T_j = t, X_{ij} = x \end{aligned} \tag{12}$$

where $t \in 0, 1$ and $t' = 1 - t$. The definition and assumptions can be interpreted the same as those for single-level models.

Under the assumptions given in Equation (12), the ACME is identified as in Equation (6) in which the conditional distribution of M and the conditional expectation of Y are estimated based on following mediator and outcome models, respectively. Structural equations for the random intercept-only model are expressed as

$$\begin{aligned} M_{ij} &= \alpha_{2i} + \beta_{2i} T_j + \pi_{2i} X_{ij} + u_{20j} + \epsilon_{2ij} \\ Y_{ij} &= \alpha_{3i} + \beta_{3i} T_j + \gamma_i M_{ij} + \kappa_i T_j M_{ij} \\ &\quad + \pi_{3i} X_{ij} + u_{30j} + \epsilon_{3ij}. \end{aligned} \tag{13}$$

For the intercept and slope model, the second expression in Equation (13) can be written as

$$\begin{aligned} Y_{ij} = \alpha_{3i} + \beta_{3i} T_j + (\gamma_i + u_{31j}) M_{ij} + \kappa_i T_j M_{ij} + \pi_{3i} X_{ij} \\ + u_{30j} + \epsilon_{3ij}. \end{aligned} \tag{14}$$

For the random intercept and random intercept and slope model, the ACME is identified and given by $E[\beta_{2i}(\gamma_i + t\kappa_i)]$ for $t = 0$ or 1 under assumption (12). A proof for identifying

the ACME in multilevel settings is identical to the proof under single-level settings unless there exists spillover effects in which the exposure of one individual to the treatment affects the outcome of the other individual.

Equations (13) and (14) have some important differences in comparison to the structural equation models shown in Equation (8). First, the use of multilevel analysis takes into account the nested structure of the data. In group-randomized designs, a group consists of individuals and is viewed as a randomization unit, and many group-randomized designs suffer from the interference between individuals (Jo et al., 2008). Ignoring the interference between individuals may result in underestimation of standard errors yielding an inflation of the Type I error rate for the treatment effect. Here, we provide correct standard errors by using the multilevel analysis technique. The correlation among individuals within groups in the mediator as well as the outcome is controlled using the random intercept model shown in Equation (13). The correlation among individuals within groups in the relationship between the mediator and outcome is controlled using the random slope model shown in Equation (14). Second, unlike Baron and Kenny's conventional approach, our model allows heterogeneous effects of the treatment and mediator, which is represented by the subscript i in the coefficients. We introduce our approach to estimating the ACME under heterogeneous effects in the section on G-computation with Bayesian inferential methods.

In group-randomized designs, interference between groups may occur if the mediator of interest is measured at the within-school level. The interference between units may give rise to spillover effects (VanderWeele et al., 2014). For example, in the FAST example, improvement in parent social capital is hypothesized as an important variable that mediates the effect of the FAST intervention on reducing student peer problems. In addition to the mediator, the FAST effect may be reinforced by the atmosphere in the school created by the improved level of parent social capital on average. The existence of spillover effects results in violating Rubin (1974)'s Stable Unit Treatment Value Assumption (SUTVA).

In the event of spillover effects, the ACME is defined differently than the case where a mediator exists at the within-school level. VanderWeele, Hong, Jones, and Brown (2013) provide the statistical framework of causal mediation analysis when SUTVA is violated. The framework is similar to the case where there exists two independent mediators (Imai & Yamamoto, 2013). The ACME that is mediated through the within-school level mediator is defined as

$$\bar{\delta}_w(t) = E\{Y_{ij}(t, M_{ij}(1), W_j(t)) - Y_{ij}(t, M_{ij}(0), W_j(t))\} \quad (15)$$

for $t \in \{0, 1\}$ and W_j represents the between-school level mediator. Equation (15) expresses the difference in the outcome in response to a change in the within-school level mediator after fixing the treatment status to t and the between-school level mediator to the value that would have resulted under t. In the causal structural model of Figure 2, this is represented by the path from T to Y that is mediated through M.

The ACME that is mediated through the between-school level mediator is defined as

$$\bar{\delta}_b(t) = E\{Y_{ij}(t, M_{ij}(t), W_j(1)) - Y_{ij}(t, M_{ij}(t), W_j(0))\} \quad (16)$$

where $t \in \{0, 1\}$, and W_j represents the between-school level mediator. The term $\bar{\delta}_b(t)$ represents the difference in the outcome in response to a change in the between-level mediator W after fixing the treatment status to t and the within-school level mediator to the value that would have resulted under t. In the causal structural model of Figure 2, this is represented by the path from T to Y that is mediated through W.

The ACME is identified under the following set of assumptions:

$$\begin{aligned} &\{Y_{ij}(t', m), M_{ij}(t)\} \perp T_j | X_{ij} = x \\ &Y_{ij}(t', m) \perp M_{ij}(t) | T_j = t, X_{ij} = x \\ &Y_{ij}(t', m) \perp W_j(t) | T_j = t, X_{ij} = x \end{aligned} \quad (17)$$

where $t \in \{0, 1\}$ and $t' = (1 - t)$. The assumptions in Equation (17) can be interpreted in the same way as the sequential ignorability assumption described before except for the last assumption. The last assumption states that there is no confounding between the between-school level mediator and outcome given the treatment status and observed covariates.

Perhaps more importantly, independence is assumed between the within- and between-school level mediators. This assumption implies that there exists no causal path between the within- and between-level mediators (see Figure 2). This is a very strong assumption and may not be plausible in practice. In many cases, the interaction between individuals may generate spillover effects, and individual change in the mediator is reinforced by the spillover effects. Thus, further study is required to address this last assumption. This article focuses only on the case where spillover effects are not present.

G-Computation with Bayesian Inferential Methods

In this section, we introduce a modified G-computation approach using Bayesian inferential methods. The method of G-computation was first introduced by Robins (1986) for estimating causal effects in the presence of time-varying treatments. The method of G-computation is a causal inference technique that estimates the distribution of potential outcomes. The causal effect is estimated by computing the difference between the potential outcomes under different treatment regimes, e.g., treated and not treated. The G-computation approach can also be applied to a mediation setting as it is a special case of the time-varying treatments approach where individuals are treated or exposed at two different time points with the outcome measured at each time point. Specifically, a mediator can be viewed as the treatment

or exposure at $(t + 1)$ when the treatment is assigned at time t.

The first step in G-computation is no different than conventional regression. A distinctive feature of the G-computation approach is that it separates the estimation of the effects of interest and the effects that are not directly related to the research question by using the marginal distribution of random variables (Snowden et al., 2011). For example, in the case where the treatment effect varies with a covariate, G-computation provides a single value that is weighted by the observed frequency of the covariate in the set data. A detailed explanation of G-computation and how it differs from conventional regression can be found in Robins (1986) and Snowden, Rose, and Mortimer (2011).

We combine G-computation with Bayesian inferential methods. We argue that a fully Bayesian approach to causal mediation analysis has certain philosophical advantages over the frequentist approach. The most important advantage lies within Bayesian philosophy itself—namely that the Bayesian approach directly incorporates the analyst's degree of uncertainty into an analysis by means of the elicitation of prior distributions on all model parameters (see e.g., Kaplan, 2014). With respect to causal mediation analysis, prior knowledge can be brought to bear on the ACME.

The ACME can be estimated using a Monte-Carlo G-computation algorithm (Robins, 1989). The algorithm is not based on integration but based on a simulation technique to approximate the distribution of the causal mediation effect. We modify the algorithm to accommodate the multilevel structure of the data and combine it with Bayesian inferential methods. The steps of the algorithm are outlined below.

Algorithm Steps

1. Fit a Bayesian multilevel regression for the mediator and outcome models separately.
2. Draw coefficient values from posterior distributions of the mediator and outcome models using MCMC.
3. Generate potential mediator values $M(1)$ and $M(0)$ using the coefficients drawn from posterior distributions and the design matrix of the mediator model.
4. Generate four potential outcomes given the potential mediator values $Y(1, M(1))$, $Y(1, M(0))$, $Y(0, M(1))$ and $Y(0, M(0))$ using the coefficients drawn from posterior distributions and the design matrix of the outcome model. Potential mediator values are generated from Step 3.
5. Compute the mediation effect. The mediation effects for $t = 1$ and 0 are $\delta(1) = Y(1, M(1) - Y(1, M(0))$ and $\delta(0) = Y(0, M(1)) - Y(0, M(0))$, respectively. The combined mediation effect is computed as $(\delta(1) + \delta(0))/2$.

Several remarks on the algorithm steps are as follows. First, the number of generated potential mediator and outcome values for each individual from Step 3 and 4 is equal to the number of draws from the posterior distributions. Each drawn value of the potential mediator and outcome represents a random sample from the posterior distributions of the potential mediator and outcome. The posterior standard deviations are estimated based on the posterior distributions of $\delta(1)$ and $\delta(0)$. Second, the random effects and errors drawn from posterior distributions of the mediator and outcome models are added in generating potential mediator and outcome in order to accommodate the multilevel structure of the data. Third, note that the generated potential outcomes of $Y(0, M(1))$, and $Y(1, M(0))$ for individual i are never observed. Only the average potential outcome over individuals is identified after assuming *sequential ignorability*, and the distribution of the potential outcomes is estimated by using the G-computation method. Lastly, the simple average between $\delta(1)$ and $\delta(0)$ is used to estimate the combined mediation effect. The simple average is used because the estimated ACMEs for treated and control are defined hypothetically. The term $\delta(1)$ represents the hypothetical mediation effect where everyone received the treatment, and $\delta(0)$ represents the hypothetical mediation effect where everyone received the control. Thus, δ is defined as the hypothetical mediation effect where the treatment and control are equally distributed across the target population. However, depending on the researcher's interest, the average ACME weighted by the proportion of treated and control can be used for the combined mediation effect.

In fitting separate Bayesian multilevel regressions, we use the estimation method that was incorporated in the R package (R Development Core Team, 2013) MCMCpack (Martin et al., 2010). For the random intercept model, we fit the Bayesian multilevel models given in Equations (13) and (14), where the random effects are given the following distributions:

$$u_{20j} \sim N(0, V_{2b}),$$
$$u_{30j} \sim N(0, V_{3b}), \tag{18}$$

And the errors:

$$\epsilon_{2ij} \sim N\left(0, \sigma_2^2\right),$$
$$\epsilon_{3ij} \sim N\left(0, \sigma_3^2\right), \tag{19}$$

where V_{2b} and V_{3b} are the variances of the random effects, and σ_2^2 and σ_3^2 are error variances.

The following conjugate prior distributions are specified for Equations (13) and (14) for the parameters of the model:

$$\beta \sim N_p(\mu_\beta, V_\beta), \tag{20}$$
$$\sigma^2 \sim \text{Inverse-gamma}(v, 1/\omega), \tag{21}$$
$$V_b \sim \text{Inverse-Wishart}(r, rR), \tag{22}$$

where β represents all the fixed effects in the model, and are assumed to follow a normal distribution with prior mean (μ_β) and variance (V_β). The error variance (σ^2) is assumed

to follow the Inverse-gamma distribution with shape (ν) and scale (ω) parameters. The covariance matrix of random effects (V_b) is assumed to follow the Inverse-Wishart distribution with the shape (r) and the scale (R) parameters. For a detailed explanation of conjugate priors see Gelman, Carlin, Stern, and Rubin (2003).

The joint posterior distribution of the parameters of the random intercept model can be expressed as

$$p(\alpha_2, \beta_2, \sigma_{u_{20j}}, \sigma_{\epsilon_{2ij}}|M) \propto p(M|\alpha_2, \beta_2, \sigma_{u_{20j}}, \sigma_{\epsilon_{2ij}}) \times p(\alpha_2|\sigma_{u_{20j}}) \times p(\beta_2)p(\sigma_{u_{20j}})p(\sigma_{\epsilon_{2ij}}), \quad (23)$$

and

$$p(\alpha_3, \beta_3, \gamma, \sigma_{u_{30j}}, \sigma_{\epsilon_{3ij}}|Y) \propto p(Y|\alpha_3, \beta_3, \gamma, \sigma_{u_{30j}}, \sigma_{\epsilon_{3ij}}) \times p(\alpha_3|\sigma_{u_{30j}})p(\beta_3)p(\gamma)p(\sigma_{u_{30j}})p(\sigma_{\epsilon_{3ij}}) \quad (24)$$

where $\sigma_{u_{0j}}$ and $\sigma_{\epsilon_{ij}}$ represents the random effect for school j and the residual for student i in school j, respectively. We use MCMC via a Gibbs sampler to obtain the joint posterior distributions.

DESIGN AND RESULTS OF SIMULATION STUDIES

In this section we conduct two simulation studies to test the performance of our proposed Bayesian approach to multilevel causal mediation analysis under homogeneous and heterogeneous effects. Under homogeneous effects, the results of the proposed approach coincide with the results of Baron and Kenny's conventional regression approach under the frequentist framework. Thus, the purpose of the first simulation study is to demonstrate the Bayesian and frequentist properties of the proposed approach. Under heterogeneous effects, the proposed approach should yield different results than Baron and Kenny's approach by virtue of taking into account the heterogeneous effects using G-computation. We aim, therefore, to demonstrate the advantage of the proposed approach over the conventional approach in the second simulation study.

Simulation Study 1: Homogeneous Effects

Using the proposed approach, the ACME between a random intercept only and a random intercept and slope model is estimated under the following study design conditions: (a) non-informative and informative priors on coefficients, (b) inaccurate and accurate priors, (c) varying precisions of inaccurate and accurate priors, (d) different values of the intra-class correlation (ICC), and (e) different group sample sizes with a constant within-group sample size.

A non-informative prior refers to vague or diffused information on coefficients whereas informative prior refers to specific and definite information on coefficients. Under the non-informative prior condition we set μ_β to zero along with precisions ($\frac{1}{V_\beta}$) of 0.[2] Under the informative prior condition, we consider the situation where we have inaccurate or accurate priors on coefficients. An accurate prior refers to information that is close to the true value whereas an inaccurate prior refers to information that is far from the true value. We set μ_β arbitrarily to zero in order to mimic the case where the researcher has inaccurate prior information while accurate priors are obtained from true parameter values used in the simulation study. For accurate and inaccurate priors, we examine three levels of precision: 1, 10, and 100. One may argue that a researcher should not use an inaccurate prior with a high level of precision. However, we used the same three levels of precision for the inaccurate prior situation to study the case in which a researcher believes that the wrong prior is correct.

In multilevel Bayesian models, it is important to incorporate appropriate priors on the covariance matrix of random effects (V_b) and the error variance (σ_2). In our simulation study, we used non-informative inverse-Wishart and inverse-gamma priors to mimic the case where a researcher lacks knowledge on the covariance matrix of random effects (V_b) and the error variance (σ_2). The inverse-Wishart (1, 0.1) and the inverse-Wishart (2, R) are used for the random intercept only and random intercept and slope models, respectively, where R is a diagonal matrix of (0.1, 0.1). For the error variance, inverse-gamma (0.001, 0.001) is used as a non-informative prior.

Gelman (2006) noted that serious problems can occur with the inverse-gamma family of non-informative prior distributions, and recommends to use the uniform prior when the number of groups is small and in other settings where weakly informative priors are necessary. Bearing this caveat in mind, the simulation study was conducted with non-informative inverse-gamma priors due to the limitation of the program *MCMChregress* (in *MCMCpack*) that only allows one to incorporate the inverse-gamma distribution as the prior distribution for the error variance.

Multilevel data were generated as follows. First, a cluster identification number was assigned to every 10 students. Random effects and residuals of the mediator and outcome were generated from a multivariate normal distribution of the τ and σ^2 matrix, respectively. The τ and σ^2 matrix are determined by varying the ICC (0.1, 0.2 and 0.3). For the random slope model, we set the covariance of the random intercept and the slope to the same value as the variance of the intercept. The treatment indicator was generated as a vector of ones and zeros.

For the random intercept model, the mediator and outcome values for individual i who attended school j were generated as follows.

$$M_{ij} = 8.1 + 1.2T_j + u_{20j} + \epsilon_{2ij}$$

[2]Setting $\frac{1}{V_\beta} = 0$ tells *MCMCpack* to set a improper uniform prior.

$$Y_{ij} = -1.0 - 1.4T_j - 2.6M_{ij} - 0.5T_jM_{ij} + u_{30j} + \epsilon_{3ij} \quad (25)$$

For the random slope model, the mediator and outcome values are generated as

$$Y_{ij} = -1.0 - 1.4T_j + (-2.6 + u_{31j})M_{ij} - 0.5T_jM_{ij} + u_{30j} + \epsilon_{3ij}. \quad (26)$$

The coefficients are drawn by applying the same model to the FAST data. However, some adjustments are made to the coefficients in order to avoid the small effect size issue (the effect size of the ACME in the FAST data is –0.04). To ensure that the effect size is at least moderate, the coefficients are multiplied by 10. Data consists of three variables T, M, Y, and the cluster ID. Nine data sets were generated with varying group sample size 5, 10, and 30 while the within-group sample size is set to 10, crossed with varying values of the ICC (0.1, 0.2, and 0.3). These conditions were also crossed with the inaccurate and accurate informative priors conditions.

The R package *MCMChregress* utilizes a Gibbs sampler. For this study, results were based on 2,000 MCMC iterations with a thinning interval of 2 after 1,000 burn-in for both the mediator and outcome models. In general, 2,000 MCMC iterations with a thinning interval of 2 is a relatively small number of iterations but we assessed convergence of the sampler by examining the posterior density plots and trace plots. Convergence of the Gibbs sampler was obtained for all cells of the study design. These plots, as well as the R code, are available at http://bise.wceruw.org/publications.html.

The outcomes of interest for the simulation study portion of this article are the %Bias and Root Mean Square Error (hereafter, RMSE) for the ACME. The %Bias and RMSE are computed based on 500 replications as

$$\%Bias = \frac{\bar{\delta} - \delta}{\delta} * 100 \quad (27)$$

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{K}(\delta_i - \delta)^2}{K}} \quad (28)$$

where $\bar{\delta}$ is the mean of the estimated mediation effect, δ_i is the estimated mediation effect for ith replication and K is number of replications ($K = 500$). In essence, this design offers an evaluation of the frequentist properties of a Bayesian estimate (see Little, 2006, 2011).

In addition to the measures for the posterior mean, we also present the standard deviation of the posterior distribution. This measure offers a measure of the width of the credible interval for each iteration. The posterior standard deviation is also based on 500 replications and obtained as

$$P.S.D. = \sqrt{\frac{\Sigma_{i=1}^{K}\sigma_i^2}{K}} \quad (29)$$

where K is the number of replications and σ_i^2 is the variance of the estimated mediation effect for iteration i.

Results of Simulation Study 1

Results for the random intercept models are shown in Tables 1, 2, and 3 and the results for the random intercept and slope models are shown in Tables 4, 5, and 6. The first column displays the results for the control [$\delta(0)$] and for the treated [$\delta(1)$] conditions. The second column denotes ICC conditions followed by the level of precision ($\frac{1}{V_\beta}$). The remainder of the tables display the results under the inaccurate prior and accurate prior conditions. The first column of each ICC represents the non-informative uniform prior case.

As shown in Table 1, the bias under non-information priors is in general 3 to 4% when the sample size is 50 (five schools), and it does not differ much by ICC levels. Under informative priors, the bias and variance differ as a function of the accuracy of the priors. Using inaccurate priors, the bias increases up to 100% when the precision level is 100, while it decreases to less than 1% with the same precision using accurate priors. The variance becomes smaller as the level of precision increases. The pattern shown in bias and variance is because the posterior estimates are shifted toward the priors as precision increases. This result implies that accurate priors with greater precision can help in solving the small sample size issue, while wrong priors with greater precision could lead to severely distorted results. This finding is completely consistent with the logic of Bayesian inference.

With a sample size of 100 (10 schools), the bias is less than 2% under the non-informative priors condition and also does not depend on the ICC levels. Under the informative priors condition, results show the same patterns as found in sample size of 50 condition in which the bias depends on the accuracy of the priors. With a sample size of 300 (30 schools), results show that the bias is less than 1% in most cases except where wrong priors are specified with high levels of precision (10 or 100).

Results from the random intercept and slope models are shown in Tables 4, 5, and 6. Unlike the random intercept models, the bias depends on the ICC levels in which bias increases as the ICC level increases. For example, with a sample size of 50, the bias under the non-informative priors case is only 0.9 when the ICC level is 0.1, while it is 8.1 when the ICC level is 0.3. This change is due to the fact that the effective sample size is lower as the ICC level increases. Under the informative priors case, the bias also changes according to the accuracy of priors the same as with the random intercept models.

With sample sizes of 100 and 300, the pattern of results is similar to the random slope models with a sample size of 50. The only difference is that the bias is smaller when compared to a sample size of 50 within the same conditions.

As shown across all of the tables, posterior standard deviations follow consistent patterns: (1) greater precision of the

TABLE 1
Random Intercept Model with Sample Size of 50 (5 Schools)

			Inaccurate Prior				Accurate Prior			
True Value	ICC	Prec.	Est.	%Bias	RMSE	P.S.D.	Est.	%Bias	RMSE	P.S.D.
$\hat{\delta}(0) = -3.12$	0.1	0	−3.029	2.9	0.826	0.828	−3.029	2.9	0.826	0.828
		1	−2.837	9.1	0.793	0.791	−3.029	2.9	0.741	0.771
		10	−0.799	74.4	2.422	0.503	−3.057	2.0	0.465	0.578
		100	−0.000	99.9	3.120	0.014	−3.097	0.7	0.121	0.268
	0.2	0	−3.036	2.7	0.784	0.741	−3.036	2.7	0.784	0.741
		1	−2.891	7.4	0.747	0.711	−3.035	2.7	0.716	0.699
		10	−1.196	61.7	2.060	0.565	−3.059	1.9	0.479	0.544
		100	−0.000	99.9	3.120	0.015	−3.097	0.8	0.136	0.264
	0.3	0	−3.043	2.5	0.739	0.652	−3.043	2.5	0.739	0.652
		1	−2.940	5.8	0.700	0.628	−3.042	2.5	0.686	0.622
		10	−1.632	47.7	1.633	0.584	−3.060	1.9	0.490	0.504
		100	−0.000	99.9	3.120	0.017	−3.096	0.7	0.154	0.259
$\hat{\delta}(1) = -3.72$	0.1	0	−3.568	4.1	0.991	1.0	−3.568	4.1	0.991	1.000
		1	−3.495	6.0	0.934	0.972	−3.603	3.2	0.869	0.902
		10	−1.115	70.0	2.733	0.647	−3.643	2.1	0.544	0.673
		100	−0.001	99.9	3.719	0.020	−3.692	0.8	0.143	0.311
	0.2	0	−3.582	3.7	0.940	0.894	−3.582	3.7	0.940	0.894
		1	−3.558	4.4	0.883	0.872	−3.610	2.9	0.838	0.816
		10	−1.594	57.2	2.298	0.709	−3.645	2.0	0.560	0.633
		100	−0.002	99.9	3.718	0.021	−3.691	0.8	0.159	0.306
	0.3	0	−3.595	3.4	0.885	0.787	−3.595	3.4	0.885	0.787
		1	−3.614	2.8	0.832	0.769	−3.618	2.8	0.801	0.727
		10	−2.105	43.4	1.800	0.722	−3.647	1.9	0.572	0.586
		100	−0.002	99.9	3.718	0.024	−3.690	0.8	0.180	0.300

Note. (1) ICC= Intra-Class Correlation; RMSE = Root Mean Square Error; Prec. = Precision; Est. = Estimate; and P.S.D. = Posterior Standard Deviation. (2) When precision is 0, results are based on non-informative priors. (3) Accurate priors are obtained from true parameter values while inaccurate priors are set arbitrarily to zero.

prior leads to smaller standard deviations regardless of using accurate or inaccurate priors; (2) standard deviations associated with models that have small sample sizes are larger; and (3) standard deviations in the random slope models are larger than in the random intercept models across all other conditions.

We also find that bias is a function of the sample size and accuracy of priors for the random intercept models, while it is a function of the ICC, sample size and accuracy of priors for the random intercept and slope models. Because the estimated ACME is derived from two parameters, i.e., β_2 and γ, the ACME is affected only by the random effect of the slope but not the intercept.

Simulation Study 2: Heterogeneous Effects

As shown in Simulation Study 1, the bias for the estimated ACME is less than 1% when the ICC is 0.1 and the sample size is 300 with 30 schools. We conducted a small simulation study under this condition, aiming to compare the ACME estimated under the proposed approach and the conventional Baron and Kenny (1986)'s approach in the presence of heterogeneous treatment and mediator effects.

The ACME estimated under both approaches is compared under the following study design conditions: (a) the difference in the probability of being in the treatment group for those with $X = 1$ and $X = 0$, (b) the difference in the average mediator value for those with $X = 1$ and $X = 0$, and (c) the difference between the random intercept only and random intercept and slope models.

The conventional Baron and Kenny approach to causal mediation analysis assumes a linear and constant effect with the ACME estimated as $\beta_2(\gamma + \kappa t)$ where $t \in \{0, 1\}$ and β_2, γ and κ are obtained from the mediator and outcome regression models under the frequentist framework. In contrast, the proposed approach assumes that the mean response is also linearly related to X but with varying slope when $T = 1$ and $T = 0$. The ACME is estimated as $E[\beta_{2i}(\gamma_i + \kappa_i t)]$ where $t \in \{0, 1\}$ using Monte Carlo G-computation algorithm (Robins, 1989).

To generate data with heterogeneous treatment effects, an interaction effect is included between the treatment and covariate X in the mediator model shown in Equations (25) and (26). The interaction effect between the treatment and covariate X creates the heterogeneous treatment effect on the mediator depending on the value of X ($1.5T_jX_{ij}$). For the sake of simplicity, X was generated as a binary variable that takes on values of 1 or 0 that follows the Bernoulli distribution. Likewise, we included the interaction effect between the mediator and covariate X in the outcome model ($1.7M_{ij}X_{ij}$) to generate heterogeneous mediator effects.

TABLE 2
Random Intercept Model with Sample Size of 100 (10 Schools)

			Inaccurate Prior				Accurate Prior			
True Value	ICC	Prec.	Est.	%Bias	RMSE	P.S.D.	Est.	%Bias	RMSE	P.S.D.
$\hat{\delta}(0) = -3.12$	0.1	0	−3.111	0.3	0.547	0.562	−3.111	0.3	0.547	0.562
		1	−3.115	0.2	0.523	0.550	−3.117	0.1	0.519	0.545
		10	−2.180	30.1	1.025	0.519	−3.117	0.1	0.391	0.463
		100	−0.002	99.9	3.118	0.020	−3.116	0.1	0.128	0.249
	0.2	0	−3.112	0.3	0.519	0.504	−3.112	0.3	0.519	0.504
		1	−3.148	0.9	0.500	0.491	−3.118	0.1	0.495	0.492
		10	−2.340	24.9	0.880	0.476	−3.118	0.1	0.391	0.428
		100	−0.002	99.9	3.118	0.023	−3.116	0.1	0.141	0.243
	0.3	0	−3.113	0.2	0.488	0.444	−3.113	0.2	0.488	0.444
		1	−3.172	1.7	0.471	0.433	−3.119	0.0	0.469	0.435
		10	−2.500	19.8	0.738	0.426	−3.119	0.0	0.388	0.389
		100	−0.003	99.9	3.117	0.027	−3.117	0.1	0.245	0.235
$\hat{\delta}(1) = -3.72$	0.1	0	−3.760	1.1	0.654	0.713	−3.760	1.1	0.654	0.713
		1	3.842	3.2	0.649	0.689	−3.734	0.4	0.609	0.658
		10	−2.714	27.0	1.126	0.649	−3.723	0.1	0.457	0.548
		100	−0.007	99.8	3.713	0.030	−3.718	0.0	0.146	0.292
	0.2	0	−3.760	1.1	0.618	0.639	−3.760	1.1	0.618	0.639
		1	−3.879	4.3	0.626	0.616	−3.736	0.4	0.580	0.594
		10	−2.910	21.8	0.953	0.594	−3.724	0.1	0.457	0.507
		100	−0.009	99.8	3.711	0.034	−3.719	0.0	0.161	0.285
	0.3	0	−3.760	1.1	0.580	0.562	−3.760	1.0	0.580	0.562
		1	−3.902	4.9	0.596	0.543	−3.738	0.5	0.549	0.527
		10	−3.106	16.5	0.788	0.532	−3.726	0.2	0.453	0.461
		100	−0.012	99.7	3.708	0.039	−3.720	0.0	0.179	0.275

Note. (1) ICC= Intra-Class Correlation; RMSE = Root Mean Square Error; Prec. = Precision; Est. = Estimate; and P.S.D. = Posterior Standard Deviation. (2) When precision is 0, results are based on non-informative priors. (3) Accurate priors are obtained from true parameter values while inaccurate priors are set arbitrarily to zero.

In the presence of heterogeneous treatment effects on the mediator, the difference in the probability of being in the treatment group for those with $X = 1$ and $X = 0$ is a critical condition to see the difference in the estimated ACME between the conventional and proposed approaches. Schafer and Kang (2008) noted that the conventional regression approach is sensitive to the assumption that the mean response is linearly related to X with identical slopes when $T = 1$ and $T = 0$, particularly if the distribution of X in the two groups are very different. Thus, three different probabilities of being in the treatment group for those with $X = 1$ and $X = 0$ are considered:

(1) $P(T = 1|X = 1) - P(T = 1|X = 0) = 0.6$,
(2) $P(T = 1|X = 1) - P(T = 1|X = 0) = 0.3$, and
(3) $P(T = 1|X = 1) - P(T = 1|X = 0) = 0$.

Likewise, in the presence of heterogeneous mediator effects on the outcome, three different levels of the average mediator value for those with $X = 1$ and $X = 0$ are considered:

(1) $(\bar{M}|X = 1) - (\bar{M}|X = 0) = 3$,
(2) $(\bar{M}|X = 1) - (\bar{M}|X = 0) = 1$, and
(3) $(\bar{M}|X = 1) - (\bar{M}|X = 0) = 0$.

All other conditions of the simulation study remain the same as the non-informative mediation model used in Simulation Study 1. Outcomes of interest are also consistent with Simulation Study 1 as shown in equations (27)–(29).

We used non-informative priors on the coefficients, the covariance matrix of the random effects, and the error variance. Assuming informative priors can be elicited, it is relatively straightforward to incorporate these priors into the heterogeneous treatment effect model.

Results of Simulation Study 2

The comparison of the ACME estimated by the conventional and proposed approaches under the heterogeneous treatment is shown in Table 7. The ACME estimated by the conventional approach is biased when the difference in the probability of being in the treatment group when $X = 1$ and $X = 0$ is larger than 0. The bias is more than 8% when the difference in the probability for those with $X = 1$ and $X = 0$ is 0.3, and the bias is more than 15% when the difference is 0.6. In contrast, the bias in the ACME estimated by our proposed approach is less than 1% regardless of the difference in the probability for those with $X = 1$ and $X = 0$.

TABLE 3
Random Intercept Model with Sample Size of 300 (30 Schools)

			Inaccurate Prior				Accurate Prior			
True Value	ICC	Prec.	Est.	%Bias	RMSE	P.S.D.	Est.	%Bias	RMSE	P.S.D.
$\hat{\delta}(0) = -3.12$	0.1	0	−3.108	−0.395	0.3	0.309	−3.108	0.4	0.342	0.309
		1	−3.154	1.1	0.334	0.305	−3.108	0.4	0.336	0.306
		10	−2.800	10.3	0.441	0.305	−3.108	0.4	0.302	0.288
		100	−0.014	99.5	3.106	0.045	−3.110	0.3	0.155	0.203
	0.2	0	−3.108	0.4	0.325	0.279	−3.108	0.4	0.325	0.279
		1	−3.146	0.83	0.318	0.275	−3.108	0.4	0.321	0.276
		10	−2.871	7.9	0.384	0.274	−3.108	0.4	0.294	0.263
		100	−0.027	99.1	3.095	0.053	−3.110	0.3	0.164	0.193
	0.3	0	−3.109	0.3	0.306	0.246	−3.109	0.3	0.306	0.246
		1	−3.139	0.6	0.300	0.243	−3.109	0.3	0.303	0.245
		10	−2.935	5.9	0.335	0.242	−3.109	0.3	0.283	0.235
		100	−0.099	96.8	3.039	0.073	−3.111	0.3	0.173	0.181
$\hat{\delta}(1) = -3.72$	0.1	0	−3.697	−0.6	0.416	0.377	−3.697	0.6	0.416	0.377
		1	−3.834	3.1	0.421	0.373	−3.700	0.5	0.399	0.366
		10	−3.455	7.1	0.457	0.377	−3.705	0.4	0.354	0.340
		100	−0.062	98.3	3.658	0.063	−3.709	0.3	0.178	0.237
	0.2	0	−3.698	0.6	0.395	0.339	−3.698	0.6	0.395	0.339
		1	−3.816	2.6	0.399	0.337	−3.701	0.5	0.381	0.331
		10	−3.540	4.8	0.401	0.338	−3.706	0.4	0.344	0.310
		100	−0.092	97.5	3.631	0.076	−3.710	0.3	0.189	0.226
	0.3	0	−3.700	0.5	0.372	0.299	−3.700	0.5	0.372	0.299
		1	−3.799	2.1	0.373	0.297	−3.702	0.5	0.360	0.293
		10	−3.616	2.8	0.358	0.298	−3.706	0.4	0.331	0.277
		100	−0.196	94.7	3.545	0.103	−3.710	0.3	0.200	0.212

Note. (1) ICC= Intra-Class Correlation; RMSE = Root Mean Square Error; Prec. = Precision; Est. = Estimate; and P.S.D. = Posterior Standard Deviation. (2) When precision is 0, results are based on non-informative priors. (3) Accurate priors are obtained from true parameter values while inaccurate priors are set arbitrarily to zero.

The same pattern is shown under the heterogeneous mediator effect. As shown in Table 8, the ACME estimated by the conventional approach is biased when the difference in the average mediator value for those with $X = 1$ and $X = 0$ is larger than 0. The bias is more than 4.9% when the difference in average mediator value is 3, and the bias is more than 7% when the difference is 0.6. In contrast, the bias in the ACME estimated by the proposed approach is less than 1% regardless of the difference in the average mediator value for those with $X = 1$ and $X = 0$. The same pattern holds for both the random intercept only and random intercept and slope models.

Note that the unbiased results of our proposed approach are only possible when models are correctly specified. For example, our approach will provide the same results as the conventional approach if the interaction effect between T and X (or the interaction effect between M and X) is not included in the model.

CASE STUDY

We apply our proposed approach for estimating the ACME in group randomized designs under the Bayesian framework using the FAST example described in the introduction.

Data

FAST is a program that aims to develop social capital, especially relations of trust and shared expectations, among parents, school staff, and children. The program includes an 8-week session of weekly group meetings and 2 years of monthly follow-up parent-led meetings. The program was originally designed to increase social capital among minority families, and it demonstrated that the active engagement of parents in the program led to improved behavioral and academic outcomes of their children. The program was implemented in 47 states and 16 countries (McDonald & Frey 1999; McDonald 2002).

The FAST intervention was implemented following a group-randomized design. Twenty six schools were assigned to participate in FAST and the other 26 schools were assigned to continue with business as usual without FAST. Parent social capital was measured both before and after the treatment period. Student behavioral outcome was reported by teachers after the treatment period.

Turley and her colleagues (Turley et al., 2012) examined the causal effect of the FAST intervention on student peer problems, and found evidence that the FAST intervention has a significant effect on reducing student peer problems. They went one step further, and attempted to investigate the mechanisms of the effect. Their interest focussed on whether

TABLE 4
Random Slope Model with Sample Size of 50 (5 Schools)

			Inaccurate Prior				Accurate Prior			
True Value	ICC	Prec.	Est.	%Bias	RMSE	P.S.D.	Est.	%Bias	RMSE	P.S.D.
$\hat{\delta}(0) = -3.12$	0.1	0	−3.149	0.9	0.841	0.836	−3.149	0.9	0.841	0.836
		1	−2.940	5.8	0.790	0.807	−3.147	0.9	0.774	0.780
		10	−1.597	48.8	1.607	0.687	−3.139	0.6	0.517	0.591
		100	−0.245	92.2	2.876	0.297	−3.133	0.4	0.256	0.312
	0.2	0	−3.182	1.9	1.066	1.652	−3.192	2.3	1.226	1.672
		1	−2.992	4.0	0.774	0.726	−3.149	0.9	0.771	0.708
		10	−1.786	42.8	1.442	0.655	−3.142	0.7	0.557	0.557
		100	−0.287	90.8	2.834	0.296	−3.135	0.5	0.311	0.305
	0.3	0	−3.391	8.7	2.660	11.311	−3.374	8.1	2.633	11.433
		1	−3.040	2.6	0.759	0.644	−3.151	0.9	0.763	0.632
		10	−1.994	36.1	1.266	0.611	−3.145	0.8	0.592	0.517
		100	−0.345	88.9	2.777	0.298	−3.137	0.5	0.362	0.297
$\hat{\delta}(1) = -3.72$	0.1	0	−3.753	0.9	1.027	1.050	−3.753	0.9	1.027	1.050
		1	−3.629	2.5	0.940	0.995	−3.745	0.7	0.902	0.923
		10	−1.979	46.8	1.850	0.849	−3.736	0.4	0.588	0.688
		100	−0.285	92.3	3.436	0.345	−3.728	0.2	0.257	0.343
	0.2	0	−3.852	3.5	2.366	2.788	−3.864	3.8	2.603	2.819
		1	−3.689	0.8	0.919	0.894	−3.747	0.7	0.891	0.837
		10	−2.214	40.4	1.645	0.808	−3.739	0.5	0.628	0.648
		100	−0.341	90.8	3.381	0.350	−3.731	0.3	0.315	0.337
	0.3	0	−4.606	23.8	6.890	18.335	−4.600	23.6	6.939	18.460
		1	−3.743	0.6	0.898	0.791	−3.749	0.8	0.875	0.747
		10	−2.473	33.5	1.426	0.754	−3.742	0.6	0.663	0.601
		100	−0.415	88.8	3.308	0.358	−3.733	0.3	0.368	0.329

Note. (1) ICC= Intra-Class Correlation; RMSE = Root Mean Square Error; Prec. = Precision; Est. = Estimate; and P.S.D. = Posterior Standard Deviation. (2) When precision is 0, results are based on non-informative priors. (3) Accurate priors are obtained from true parameter values while inaccurate priors are set arbitrarily to zero.

FAST effects on reducing student peer problems were mediated through improvement in parent social capital, that is, intergenerational closure. Intergenerational closure is based on parent self report about whether parents know the parents of their children's friends.

The FAST intervention serves as the treatment (T) and is a dichotomous variable that takes on the value 1 when a school is assigned to the intervention or 0 otherwise. Intergenerational closure, which serves as the mediator (M), is a numeric variable that counts the number of parents of their children's friends that a participating parent knows. The dependent variable is the index of peer problems (Y), and is a continuous variable. The pretreatment measures of intergenerational closure and design effects that are used in Turley, Gamoran, Turner, and Fish (2012) are also included in the model. The sample size is 1,833. According to the simulation studies in the previous section, the sample size is large enough to provide a reliable result even with non-informative priors under the sequential ignorability assumption as long as models are correctly specified.

Assumptions and Estimation

The *sequential ignorability* assumption that is needed to identify and estimate the ACME when a mediator is at the parent level is shown in Equaton (12). The first part of the assumption implies that there is no confounding between the FAST intervention and the number of parents of their children's friends (intergenerational closure) as well as the peer problems. The first part of the assumption is achieved by the randomization of the FAST intervention. However, note that the benefit of randomization vanishes if there exist parents who did not comply with the treatment protocol. In that case, the ACME is no longer identified even after satisfying the sequential ignorability assumption and a different approach is required to identify and estimate the ACME (Yamamoto, 2013).

This case study assumes that parents and students are all affected by the FAST intervention regardless of their participation in the FAST program, and also assumes perfect compliance to the treatment protocol. The second assumption implies that there is no confounding between intergenerational closure and student peer problems given the treatment status and covariates. This assumption is not guaranteed to be met even after conditioning on covariates, and is not empirically testable. Thus, conducting a sensitivity analysis would be desirable if it is available. However, the assumption may be plausible because the pretreatment measure of parent social capital is controlled. In the case where there are other observed covariates that are believed to confound the mediator and outcome relationship, they should be included in addition to the pretreatment measure of parent social capital.

TABLE 5
Random Slope Model with Sample Size of 100 (10 Schools)

			Inaccurate Prior				Accurate Prior			
True Value	ICC	Prec.	Est.	%Bias	RMSE	P.S.D.	Est.	%Bias	RMSE	P.S.D.
$\hat{\delta}(0) = -3.12$	0.1	0	−3.122	0.1	0.598	0.569	−3.122	0.1	0.598	0.569
		1	−3.095	0.8	0.572	0.556	−3.119	0.0	0.570	0.550
		10	−2.159	30.8	1.054	0.525	−3.118	0.1	0.432	0.469
		100	−0.440	85.9	2.682	0.283	−3.121	0.0	0.203	0.268
	0.2	0	−3.098	0.7	0.714	2.614	−3.098	0.7	0.709	2.614
		1	−3.132	0.4	0.557	0.498	−3.122	0.1	0.559	0.497
		10	−2.314	25.8	0.919	0.482	−3.120	0.0	0.448	0.434
		100	−0.524	83.2	2.600	0.289	−3.122	0.1	0.239	0.261
	0.3	0	−2.976	4.6	1.152	12.489	−2.977	4.6	1.146	12.489
		1	−3.162	1.4	0.543	0.439	−3.124	0.1	0.546	0.442
		10	−2.470	20.8	0.789	0.433	−3.122	0.1	0.461	0.396
		100	−0.636	79.6	2.489	0.297	−3.122	0.1	0.273	0.252
$\hat{\delta}(1) = -3.72$	0.1	0	−3.705	0.4	0.729	0.684	−3.705	0.4	0.729	0.684
		1	−3.802	2.2	0.702	0.677	−3.712	0.2	0.666	0.645
		10	−2.682	27.9	1.166	0.648	−3.715	0.1	0.498	0.547
		100	−0.538	85.5	3.185	0.346	−3.719	0.0	0.209	0.303
	0.2	0	−3.618	2.7	1.561	4.221	−3.621	2.6	1.518	4.222
		1	−3.842	3.3	0.684	0.605	−3.714	0.2	0.649	0.582
		10	−2.875	22.7	1.001	0.593	−3.717	0.1	0.512	0.506
		100	−0.643	82.7	3.081	0.354	−3.720	0.0	0.247	0.295
	0.3	0	−3.203	13.9	3.355	20.168	−3.218	13.5	3.260	20.171
		1	−3.871	4.0	0.664	0.532	−3.716	0.1	0.630	0.517
		10	−3.068	17.5	0.845	0.532	−3.719	0.0	0.522	0.460
		100	−0.784	78.9	2.943	0.364	−3.721	0.0	0.283	0.285

Note. (1) ICC= Intra-Class Correlation; RMSE = Root Mean Square Error; Prec. = Precision; Est. = Estimate; and P.S.D. = Posterior Standard Deviation. (2) When precision is 0, results are based on non-informative priors. (3) Accurate priors are obtained from true parameter values while inaccurate priors are set arbitrarily to zero.

Our Bayesian estimation method used to estimate the ACME and ANDE are based on following equations.

$$IC_{ij} = \alpha_{2i} + \beta_{2i} FAST_j + \pi_{2i} X_{ij} + \xi_{2i} FAST_j \cdot X_{ij} + u_{20j} + \epsilon_{2ij},$$

$$Peer\ problems_{ij} = \alpha_{3i} + \beta_{3i} FAST_j + \gamma_i IC_{ij} + \kappa_i FAST_j \cdot IC_{ij} + \pi_{3i} X_{ij} + \xi_{3i} FAST_j \cdot X_{ij} + u_{30j} + \epsilon_{3ij} \quad (30)$$

where *IC* is intergenerational closure and X is a vector consisting of the pretreatment measure of intergenerational closure and design effects. Note that the interaction effect is included in the model between the FAST intervention and pretreatment measure of intergenerational closure. As shown Simulation Study 2, the proposed method provides the unbiased estimates of the ACME and ANDE when the interaction effect exists between the FAST intervention and pretreatment measure of intergenerational closure.

We specified non-informative priors on the coefficients, random effects, and residuals as shown below.

$$\beta \sim N_p(0, 1000000), \quad (31)$$

$$\sigma^2 \sim \text{Inverse-gamma}(1, 0.1), \quad (32)$$

$$V_b \sim \text{Inverse-Wishart}(0.001, 0.001), \quad (33)$$

where β represents all the fixed effects in the model, and is assumed to follow the normal distribution with prior mean of 0 and variance of 1000000. The error variance (σ^2) is assumed to follow the Inverse-gamma distribution with a shape parameter of 1 and rate parameter of 0.1. The covariance matrix of random effects (V_b) is assumed to follow the Inverse-Wishart distribution with shape and scale parameters of 0.001.

For the sake of comparison, we also conducted causal mediation analysis with informative priors. For the informative case, priors based on a frequentist analysis of these data were used. We recognize that in practice researchers would not use elicited priors of this sort. Rather, elicited priors should be obtained more or less objectively from prior research or expert opinion (see e.g., O'Hagan et al., 2006).

We now turn our attention to the case where spillover effects are present. Suppose the effect of the FAST interven-

TABLE 6
Random Slope Model with Sample Size of 300 (30 Schools)

			Inaccurate Prior				Accurate Prior			
True Value	ICC	Prec.	Est.	%Bias	RMSE	P.S.D.	Est.	%Bias	RMSE	P.S.D.
$\hat{\delta}(0) = -3.12$	0.1	0	−3.108	0.4	0.342	0.309	−3.108	0.4	0.342	0.309
		1	−3.154	1.1	0.334	0.305	−3.108	0.4	0.336	0.306
		10	−2.800	10.3	0.441	0.305	−3.108	0.4	0.302	0.288
		100	−0.014	99.6	3.106	0.045	−3.110	0.3	0.155	0.203
	0.2	0	−3.108	0.4	0.325	0.279	−3.108	0.4	0.325	0.279
		1	−3.146	0.8	0.318	0.275	−3.108	0.4	0.321	0.276
		10	−2.871	7.9	0.384	0.274	−3.108	0.4	0.294	0.263
		100	−0.027	99.1	3.095	0.053	−3.110	0.3	0.164	0.193
	0.3	0	−3.109	0.3	0.306	0.246	−3.109	0.3	0.306	0.246
		1	−3.139	0.6	0.300	0.243	−3.109	0.3	0.303	0.245
		10	−2.935	5.9	0.335	0.242	−3.109	0.3	0.283	0.235
		100	−0.099	96.8	3.039	0.073	−3.111	0.3	0.173	0.181
$\hat{\delta}(1) = -3.72$	0.1	0	−3.697	0.6	0.416	0.377	−3.697	0.6	0.416	0.377
		1	−3.834	3.1	0.421	0.373	−3.700	0.5	0.399	0.366
		10	−3.455	7.1	0.457	0.377	−3.705	0.4	0.354	0.340
		100	−0.062	98.3	3.658	0.063	−3.709	0.3	0.178	0.237
	0.2	0	−3.698	0.6	0.395	0.339	−3.698	0.6	0.395	0.339
		1	−3.816	2.6	0.399	0.337	−3.701	0.5	0.381	0.331
		10	−3.540	4.8	0.401	0.338	−3.706	0.4	0.344	0.310
		100	−0.092	97.5	3.631	0.076	−3.710	0.3	0.189	0.226
	0.3	0	−3.700	0.5	0.372	0.299	−3.700	0.5	0.372	0.299
		1	−3.799	2.1	0.373	0.297	−3.702	0.5	0.360	0.293
		10	−3.616	2.8	0.358	0.298	−3.706	0.4	0.331	0.277
		100	−0.196	94.7	3.545	0.103	−3.710	0.3	0.200	0.212

Note. (1) ICC= Intra-Class Correlation; RMSE = Root Mean Square Error; Prec. = Precision; Est. = Estimate; and P.S.D. = Posterior Standard Deviation. (2) When precision is 0, results are based on non-informative priors. (3) Accurate priors are obtained from true parameter values while inaccurate priors are set arbitrarily to zero.

tion on reducing student peer problems is mediated through parent improved social capital as an individual, and is also mediated through school atmosphere created by the level of improved social capital on average. The assumptions also apply here and were discussed in the previous case where there is a within-group mediator. In addition to the assumptions, independence is assumed between the between-group mediator and outcome. This assumption implies that there is no confounding between the between-level mediator and outcome given the treatment status and covariates. This as-

TABLE 7
Comparison Between the Conventional and Proposed Approaches When the Heterogeneous Treatment Effects on the Mediator are Present

		Conventional Approach				Proposed Approach			
True Value	Dif.	Est.	%Bias	RMSE	P.S.D.	Est.	%Bias	RMSE	P.S.D.
Random Intercept									
$\hat{\delta}(0) = -5.07$	0.6	−4.289	15.4	0.862	0.409	−5.081	0.2	0.415	0.407
	0.3	−4.650	8.3	0.532	0.364	−5.093	0.4	0.336	0.340
	0	−5.068	0.0	0.326	0.348	−5.077	0.1	0.324	0.317
$\hat{\delta}(1) = -6.05$	0.6	−5.113	15.4	1.030	0.409	−6.063	0.3	0.502	0.477
	0.3	−5.530	8.5	0.647	0.364	−6.062	0.3	0.403	0.395
	0	−6.051	0.1	0.388	0.348	−6.066	0.4	0.387	0.353
Random Slope									
$\hat{\delta}(0) = -5.07$	0.6	−4.252	16.1	0.923	0.422	−5.044	0.5	0.468	0.414
	0.3	−4.661	8.1	0.541	0.381	−5.096	0.5	0.367	0.347
	0	−5.074	0.1	0.385	0.369	−5.080	0.2	0.386	0.325
$\hat{\delta}(1) = -6.05$	0.6	−5.076	16.0	1.088	0.422	−6.031	0.2	0.539	0.485
	0.3	−5.543	8.3	0.649	0.381	−6.070	0.4	0.425	0.405
	0	−6.053	0.1	0.415	0.369	−6.068	0.4	0.418	0.360

Note. (1) P.S.D. = Posterior Standard Deviation; and RMSE = Root Mean Square Error. (2) Dif. = Difference in probability of being in the treatment group for those with $X = 1$ and $X = 0$ ($P(T|X = 1)$–$P(T|X = 0)$).

TABLE 8
Comparison Between the Conventional and Proposed Approaches When the Heterogeneous Mediator Effects on the Outcome are Present

		Conventional Approach				Proposed Approach			
True Value	Dif.	Est.	%Bias	RMSE	P.S.D.	Est.	%Bias	RMSE	P.S.D.
Random Intercept									
$\hat{\delta}(0) = -5.07$	3	−4.289	15.4	0.862	0.409	−5.081	0.2	0.415	0.407
	1	−4.650	8.3	0.532	0.364	−5.093	0.4	0.336	0.340
	0	−5.068	0.0	0.326	0.348	−5.077	0.1	0.324	0.317
$\hat{\delta}(1) = -6.05$	3	−5.113	15.4	1.030	0.409	−6.063	0.3	0.502	0.477
	1	−5.530	8.5	0.647	0.364	−6.062	0.2	0.403	0.395
	0	−6.051	0.1	0.388	0.348	−6.066	0.4	0.387	0.353
Random Slope									
$\hat{\delta}(0) = -5.07$	3	−4.252	16.1	0.923	0.422	−5.044	0.5	0.468	0.414
	1	−4.661	8.1	0.541	0.381	−5.096	0.5	0.367	0.347
	0	−5.074	0.1	0.385	0.369	−5.080	0.2	0.386	0.325
$\hat{\delta}(1) = -6.05$	3	−5.076	16.0	1.088	0.422	−6.031	0.2	0.539	0.485
	1	−5.543	8.3	0.649	0.381	−6.070	0.4	0.425	0.405
	0	−6.053	0.1	0.415	0.369	−6.068	0.4	0.418	0.360

Note. (1) P.S.D. = Posterior Standard Deviation; and RMSE = Root Mean Square Error. (2) Dif. = Difference in the average mediator value for those with $X = 1$ and $X = 0$ $((\bar{M}|X = 1)–(\bar{M}|X = 0))$.

sumption is not guaranteed to be met even after conditioning on covariates, and is not empirically testable. However, the assumption may be plausible because the average pretreatment measured parent social capital is adjusted.

The assumption of independence between within-group social capital and spillover effects is unlikely to hold in practice. The level of the parent social capital is likely to be influenced by the school atmosphere that is created by the improved parent social capital on average as parents engage in social interactions inside and outside of their school. Due to the violation of this assumption, the ACMEs through within-level social capital and spillover effects are not estimated.

Results of Case Study

The FAST intervention effects on student peer problems through parent improved intergenerational closure, i.e., whether parents know the parent of their children's friends is estimated by our proposed approach. The estimated ACME and ANDE are presented in Table 9 and 10.

With non-informative priors, the estimated ACME for the treated group is –0.02 and the 95% posterior probability interval (PPI) indicates that the ACME is within the interval of [–0.04, –0.01] . The estimated ANDE for the treated group is –0.09 and the 95% PPI indicates that ANDE is within the interval of [–0.27, 0.07]. These results are similar to estimates for the control group except for the fact that the effect size is slightly larger than the estimate for the treatment group. Specifically, the estimated ACME for the control group is –0.03 and the estimated ANDE for the control group is –0.10.

The results of the case study do not differ much when incorporating informative priors. In fact, it is shown in the first simulation study that informative priors play a limited role when the sample size is large. The only difference lies in the width of the PPI. The width of the PPIs after incorporating informative priors tends to be smaller than intervals with non-informative priors. For example, the interval of the ANDE for the treatment group with non-informative prior is [–0.27, 0.02] while the interval with informative prior is [–0.20, 0.02]. This implies that incorporating correct priors provides a precise estimate.

Overall, we conclude from the case study that the causal effect of the FAST intervention on peer problems does

TABLE 9
Multilevel Bayesian Causal Mediation Effects with Non-Informative Priors

Parameter	EAP	95 Lower PPI	95 Upper PPI
Treated			
Mediation Effect	−0.02	−0.04	−0.01
Direct Effect	−0.09	−0.27	0.07
Proportion via Mediation	0.15	−1.70	1.72
Not-treated			
Mediation Effect	−0.03	−0.05	−0.01
Direct Effect	−0.10	−0.27	0.06
Proportion via Mediation	0.20	−2.30	2.37
Average			
Mediation Effect	−0.02	−0.04	−0.01
Direct Effect	−0.09	−0.27	0.06
Proportion via Mediation	0.17	−1.90	2.29
Total Effect	−0.12	−0.29	0.04

Note. (1) Multilevel structure of the data is taken into account using Bayesian multilevel regression models with random intercept and slope. (2) EAP = Expected A Posteriori; and PPI = Posterior Probability Interval.

TABLE 10
Multilevel Bayesian Causal Mediation Effects with Informative Priors

Parameter	EAP	95 Lower PPI	95 Upper PPI
Treated			
Mediation Effect	−0.02	−0.04	−0.01
Direct Effect	−0.09	−0.20	0.02
Proportion via Mediation	0.17	0.02	0.87
Not-treated			
Mediation Effect	−0.03	−0.04	−0.01
Direct Effect	−0.10	−0.21	0.01
Proportion via Mediation	0.22	0.05	1.13
Average			
Mediation Effect	−0.02	−0.04	−0.01
Direct Effect	−0.09	−0.21	0.02
Proportion via Mediation	0.19	0.05	0.99
Total Effect	−0.12	−0.23	−0.01

Note. (1) Multilevel structure of the data is taken into account using Bayesian multilevel regression models with random intercept and slope. (2) EAP = Expected A Posteriori; and PPI = Posterior Probability Interval.

mediate through parent-improved intergenerational closure. The average proportion of the total effect that is via mediation is 17%, which is obtained by dividing the ACME by the average total effect.

SUMMARY AND CONCLUSION

Causal mediation analysis represents an important extension of conventional mediation analysis allowing for the identification and estimation of causal effects under the potential outcomes framework. As with all methods of causal inference, causal mediation analysis rests on strong assumptions, in particular the assumption of sequential ignorability. This article offers a fully Bayesian approach to causal mediation analysis for group-randomized designs which directly encodes uncertainty about the causal effect into the analysis through the specification of a prior distribution on the causal effect.

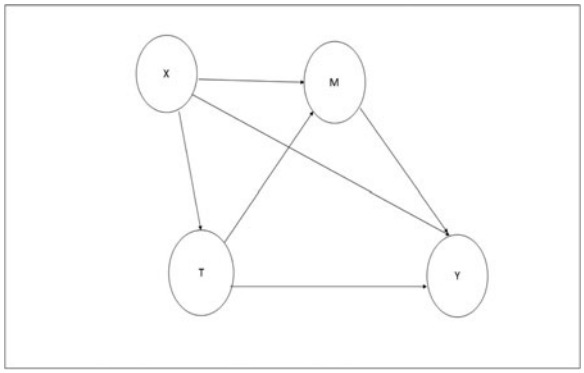

FIGURE 1 The causal structural model when spillover effects are not present. X = covariate; T= treatment; M = Mediator; Y = outcome.

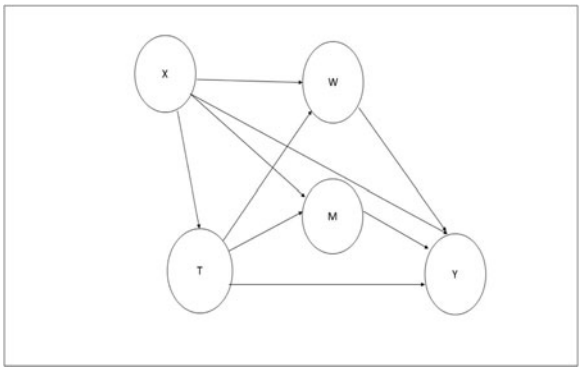

FIGURE 2 The causal structural model when spillover effects are present. X = covariate; T = treatment; M = within-school mediator; W = between school mediator; Y = outcome.

In addition, our approach provides a method for addressing homogeneous versus heterogeneous treatment and mediator effects. The results of the first simulation study demonstrate that our proposed Bayesian approach to causal mediation analysis satisfies Bayesian as well as frequentist properties. The second simulation study suggests that the proposed approach provides unbiased results even under heterogeneous treatment or mediator effects by using G-computation as long as models are correctly specified.

An essential component in the application of Bayesian methods is the elicitation of priors on all model parameters. In cases where prior knowledge is lacking, non-informative prior distributions, such as the uniform distribution, can be employed. If prior knowledge is available, then informative prior with varying levels of precision can be employed. In many applications there may be a mix of non-informative and informative priors on model parameters, and this is particularly true for complex multiparameter models. In the context of causal mediation analysis, we demonstrate the consequences of using non-informative or informative priors on the parameters of the mediation model. In the case of informative priors, we used the generating coefficients and, of course, this would not occur in a real data setting. Bayesian methodology does allow for model comparison based on different elicitations of priors using such statistics as the Bayes factors or posterior predictive checks (see e.g., Kaplan, 2014). The development of these model comparison methods for Bayesian causal mediation analysis was beyond the scope of this article. Nevertheless, as background knowledge accumulates, the Bayesian approach to causal mediation advocated in this article should lead to more accurate estimates of causal effects.

ARTICLE INFORMATION

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors

reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was supported by Grant R305D110001 from the Institute of Educational Sciences. Case study data for this article were collected under a grant (1R01HD051762-01A2) from the National Institute of Child Health and Human Development.

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors would like thank Peter Steiner and anonymous reviewers of *Multivariate Behavioral Research* for their comments on prior versions of this manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institution, the Institute of Education Sciences, or the National Institute of Child Health and Human Development is not intended and should not be inferred.

REFERENCES

Baron, R. M., & Kenny, D. A. (1986). A moderator-mediator variable distinction in social psychological research. *Journal of Personality and Social Psychology*, *51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, *8*, 158–233. doi:10.2307/1167125

Duncan, O. D. (1975). *Introduction to structural equation models*. New York, NY: Academic.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. doi:10.1214/06-BA117A

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd ed. London, UK: Chapman and Hall.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images). *Pattern Analysis and Machine Intelligence, IEEE Transactions*, *6*, 721–741. doi:10.1080/02664769300000058

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. London, UK: Chapman and Hall.

Goldberger, A. S., & Duncan, O. D. (1972). *Structural equation methods in the social sciences*. New York, NY: Seminar.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). New York, NY: Wiley.

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53*, 455–467. doi:10.1007/bf02294400

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960. doi:10.1080/01621459.1986.10478354

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*, 309–334. doi:10.1037/a0020761

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, *105*, 765–789. doi:10.1017/s0003055411000414

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*, 51–71. doi:10.1214/10-sts321

Imai, K., & Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing esperiments. *Political Analysis*, *21*, 141–171. doi:10.1093/pan/mps040

Jo, B., Asparouhov, T., Muthén, B. O., Ialongo, N. S., & Brown, C. H. (2008). Cluster randomized trials with treatment noncompliance. *Psychological Methods*, *13*(1), 1. doi:10.1037/1082-989x.13.1.1

Kaplan, D. (2009). *Structural Equation Modeling: Foundations and Extensions*. (2nd ed.). Newbury Park, CA: Sage.

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York, NY: Guilford.

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 650–673). New York: Guilford.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. New York, NY: Wiley.

Little, R. J. (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *The American Statistician*, *60*, 213–223. doi:10.1198/000313006X117837

Little, R. J. (2011). Calibrated Bayes for statistics in general and missing data in particular. *Statistical Science*, *26*, 162–174. doi:10.1214/10-STS318

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Routledge.

Martin, A. D., Quinn, K. M., & Park, J. H. (2010 May 10). *Markov chain Monte Carlo (MCMC) Package*. http://mcmcpack.wustl.edu

McDonald, L. (2002). *Evidence-based, family strengthening to reduce delinquency: FAST Families and Schools Together*. Social workers' desk reference. New York, NY: Oxford University.

McDonald, L., & Frey, H. E. (1999). *Families and schools together: Building relationships*. U.S. Department of Justice, Office of Justice Programs, Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

McDonald, R. P. (1993). A general model for two level data with responses missing at random. *Psychometrika*, *58*, 575–585. doi:10.1007/bf02294828

McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods and Research*, *22*, 399–413. doi:10.1177/0049124194022003007

McDonald, R. P., & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology*, *42*, 215–232. doi:10.1111/j.2044-8317.1989.tb00911.x

Muthén, B. (1989). Latent variable modeling in heterogenous populations. *Psychometrika*, *54*, 557–585. doi:10.1007/bf02296397

Muthén, B., & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. doi:10.1037/a0026802

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., . . . Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. West Sussex, England: Wiley.

Pearl, J. (2000). *Causality: Models, reasoning and inference* (Vol. 29). Cambridge, MA: MIT Press.

R Development Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousands Oaks, CA: Sage.

Reardon, S. F., & Raudenbush, S. W. (2013). Under what assumptions do site-by-treatment instruments identify average causal effects? *Sociological Methods & Research*, *42*(2), 143–163. doi:10.1177/0049124113494575

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period – application to control of the healthy worker survivor effect. *Mathematical Modeling*, *7*, 1393–1512. doi:10.1016/0270-0255(86)90088-6

Robins, J. M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: A focus on AIDS*, *113*, 159.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701. doi:10.1037/h0037350

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, *13*, 279–313. doi:10.1037/a0014268

Schmidt, W. H. (1969). *Covariance structure analysis of the multivariate random effects model*. Unpublished doctoral dissertation, University of Chicago.

Snowden, J. M., Rose, S., & Mortimer, K. M. (2011). Implementation of g-computation on a simulated data set: Demonstration of a causal inference technique. *American Journal of Epidemiology*, *173*(7), 731–738. doi:10.1093/aje/kwq472

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *13*(1982), 290–312. doi:10.2307/270723

Song, X.-Y., & Lee, S.-Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. New York, NY: John Wiley & Sons.

Turley, R. N. L., Gamoran, A., Turner, A., & Fish, A. (2012). Causal effects of social capital on child outcomes. Unpublished manuscript.

VanderWeele, T., Hong, G., Jones, S., & Brown, J. (2013). Mediation and spillover effects in group-randomized trials: A case study of the 4r's educational intervention. *Journal of the American Statistical Association*, *108*, 469–482. doi:10.1080/01621459.2013.779832

Yamamoto, T. (2013). Identification and estimation of causal mediation effects with treatment noncompliance. Unpublished manuscript.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation Analysis. *Psychological Methods*, *14*, 301–322. doi:10.1037/a0016972