Causal Mediation Analysis in Multi-site Trials: An Application of Ratio-of-Mediator-Probability Weighting to the Head Start Impact Study

Xu Qin, Guanglei Hong
University of Chicago, Department of Comparative Human Development, 5736 S.
Woodlawn Ave., Chicago, IL 60637

Abstract
This study focuses on developing methods for causal mediation analysis in multisite trials and uses the national Head Start Impact Study as a motivating example. The causal effects of interest, defined in terms of potential outcomes, include the indirect effect of assignment to Head Start programs on child vocabulary learning mediated by a program-induced increase in parent reading to child and the direct effect of Head Start programs. The goal is to reveal not only the prevalent causal mechanism but also how the mechanism may vary across sites. Extending the ratio-of-mediator-probability weighting (RMPW) approach to causal mediation analysis in multi-site trials, we estimate the average direct effect, the average indirect effect, and the between-site variance and covariance of these causal effects. This strategy allows for treatment-by-mediator interaction. It greatly simplifies the outcome model specification and therefore avoids possible model misspecifications. The performance of the approach is assessed across a range of multi-site designs that differ in the number of sites and the sample size per site. We investigate the relative strengths and limitations of the RMPW strategy through simulations.

Key Words: Causal inference; multi-site experimental designs; mediation mechanism; direct effect; indirect effect; potential outcome; propensity score.

1. Introduction

Multisite randomized trials are widely used in educational and medical research. People in each site are randomly assigned to a treatment or control group. This process is replicated over multiple sites. Analyzing data from multisite randomized trials, researchers obtain evidence with regard to not only the overall average treatment effect but also the variation in the treatment effect across sites. Further research is then required, yet rarely done in the past, for understanding why the treatment effect differs by sites. One possible explanation is that the causal mechanism may not be universal and may depend on how the treatment is implemented at each site. Moreover, the impact of implementation may depend on site-specific contextual characteristics. Adopting the potential outcomes causal framework, this study focuses on developing a new method for causal mediation analysis in multisite trials with ratio-of-mediator probability weighting (RMPW). The purpose is to reveal not only the prevalent causal mechanism but also how the mechanism may vary across sites.

To illustrate the RMPW approach, we describe an application example in the next section, which is followed by a review of the existing methods for causal mediation analysis. Then we define the causal parameters, explain the theoretical rationale of the RMPW approach, clarify the identification assumptions, and demonstrate its extension to multi-site settings. After delineating the estimation procedures, we assess the performance of the RMPW approach through simulations. Finally, we discuss the strengths and limitations of this new approach and raise issues for the future study.

2. Application Example

We select the Head Start Impact Study (HSIS) (Puma, et al., 2010), a national evaluation of Head Start programs with a multisite randomized design, to examine heterogeneity in mediation mechanisms. Since 1965, the federal government has sponsored Head Start (HS) programs designed to increase school readiness of children from low-income families. Besides providing comprehensive services directly for children, HS programs also intend to improve parenting practices. However, in the absence of strong evidence from rigorous evaluations, the effectiveness of HS programs was continuously questioned. From 2002 to 2006, a national multisite randomized trial was conducted to determine the impact of Head Start on child development and parental practices.

The HSIS sample included 2,449 3-year-olds from 328 randomly selected HS centers representing the national population of newly entering 3-year-old HS applicants and the national population of HS programs, respectively. At each site, because the number of open slots in the HS center cannot meet the demand, applicants were assigned at random to the HS center or to a control group. The treatment assignment probability was decided by the ratio of open slots to the number of applicants at a given site, and was thus different across the sites. The control children could receive any other non-HS services. The sample size within each site ranges from 1 to 46, with a mean of 7.5.

In this study, we focus on the HS impact on child vocabulary ability and use the IRT (Item Response Theory) calibrated PPVT (Peabody Picture Vocabulary Test) score as the outcome. HS programs encouraged parents to read to their children at home. At the end of the treatment year for the 3-year olds, there was clear evidence that HS programs improved child vocabulary score and increased parent reading to child by a greater amount than did the control condition. We reason that the improvement in child vocabulary may not only be caused by the instruction received in HS programs, but may also be attributed to the fact that the HS parents read to their children more frequently. There are two possible pathways through which HS affects child PPVT score. We examine parent reading to child as the focal mediator and dichotomize the measure into a high reading frequency level denoted by 1 and a low level denoted by 0. We consider 16 pretreatment covariates that are associated with child vocabulary or with parent reading frequency. These include child age, gender, race, home language, pre-academic skills, and biological mother's education status.

3. Literature Review

In mediation analyses of multilevel data with individual-level outcomes, some researchers merely focused on a cluster-level treatment or mediator (Krull & Mackinnon, 1999; Pituch, Murphy, & Tate, 2009). More recently, there has been increasing attention toward individual-level treatments and mediators. Kenny, Korchmaros, & Bolger (2003) were the first to assess the between-site variation in the direct effect through multilevel path analysis. For individual *i* at site *j*, the path models are:

$$Y_{ij} = d_{0j} + c_j A_{ij} + r_{ij}$$

$$Z_{ij} = d_{1j} + a_j A_{ij} + e_{ij}$$

$$Y_{ij} = d_{2j} + c'_j A_{ij} + b_j Z_{ij} + f_{ij}$$

The site-specific treatment effect on the mediator (a_j), mediator effect on the outcome (b_j), treatment effect on the outcome (c_j), and the direct effect of the treatment on the outcome (c_j') are each assumed to have a normal distribution. However, when the multilevel path models are analyzed separately, it is difficult to estimate cov(a_j, b_j), an essential component of the average indirect effect $E(a_j b_j)$ = $E(a_j)$×$E(b_j)$ + cov(a_j, b_j). Even more challenging is the estimation of the between-site variation of the indirect effect. Bauer, Preacher, & Gil (2006) combined the mediator model and the outcome model and estimated them simultaneously as multivariate multilevel models through the use of indicator variables S_Z for the mediator and S_Y for the outcome, shown as follows, such that cov(a_j, b_j) can be directly estimated:

$$R_{ij} = S_{Z_{ij}}\left(d_{Zj} + a_j A_{ij}\right) + S_{Y_{ij}}(d_{Yj} + c_j' A_{ij} + b_j Z_{ij}) + e_{ij},$$

Still, this approach does not specify a single parameter corresponding to the average indirect effect. Thus the variance of the site-specific indirect effect and the covariance between the direct and indirect effects cannot be easily obtained. Moreover, when the mediator and sometimes the treatment are not randomized, a large number of covariates will need to be incorporated into the above model. The results may be biased due to possible misspecifications of the above model if one omits confounders of the mediator-outcome relationships, if one misspecifies nonlinear covariate-outcome relationships, or if one fails to consider treatment-by-mediator interaction, mediator-by-covariate interactions, or treatment-by-mediator-by-covariate interactions. Relying on maximum likelihood estimation, the above strategy additionally assumes that the mediator and the outcome are both normally distributed. More crucially, this line of research has not always explicated the assumptions under which the causal effects of interest can be identified from observable data. The key identifying assumptions, as we will emphasize later, are that the treatment assignment and the mediator value assignment under each treatment are both ignorable given the observed covariates and that there is no treatment-by-mediator interaction.

The instrumental variable (IV) method has provided another alternative for analyzing data from multi-site randomized trials when treatment-by-site interactions are used as instruments (Kling, Liebman, & Katz, 2007; Raudenbush, Reardon, & Nomi, 2012; Reardon & Raudenbush, 2013; Reardon, Unlu, Zhu, & Bloom, 2014). Instead of assuming ignorability, the IV method requires (1) the exclusion restriction, that is, the direct effect of the treatment on the outcome must be zero (i.e., $c_j' = 0 \, \forall \, j$). Clearly, the IV method does not apply when the direct effect is of research interest. Additional assumptions required for multi-site cases include (2) the independence between the site-level treatment effect on the mediator and the site-level mediator effect on the outcome (i.e., cov(a_j, b_j) = 0), and (3) the treatment effect on the mediator is nonzero not only on average but also within each site (i.e., $a_j \neq 0 \, \forall \, j$).

To overcome some of the constraints of the existing methods, we extend the ratio-of-mediator-probability weighting (RMPW) approach from single-level settings to multilevel settings. For single-level data, Hong and colleagues (Hong, 2010; Hong, Deutsch, & Hill, 2011; Hong & Nomi, 2012; Hong, in press) proposed the RMPW approach that estimates the average direct and indirect effects. This approach allows for treatment-by-mediator interaction without a need to specify the functional form of the outcome model. We will show that, when applied to data from multisite trials, the RMPW approach generates estimates of not only the average direct and indirect effects but also the between-site variation in each of these causal effects and the covariance of the site-specific effects, each with a single parameter.

4. Causal parameters

Definition. We define the direct effect and the indirect effect of the HS programs on child vocabulary under the potential outcomes causal framework (Robins & Greenland, 1992; Pearl, 2001). For an eligible child at a given site, if the HS program would improve child vocabulary directly without increasing parent reading to child, the direct effect of the program would be positive. If the impact of the HS program was transmitted partly through program-induced increase in parent reading to child, the indirect effect would be positive as well. We hypothesize that the direct effect may differ by site because the quality of instruction in HS programs may not be consistent and because the availability and quality of alternative childcare for the control children may be vastly different across the sites. The indirect effect of the program may also differ by site possibly due to differential amount of program emphasis on parenting support. If a program that displayed a higher instructional quality was also generally more successful in enhancing literacy activities at home, then the direct effect and the indirect effect may be positively correlated.

Notation. Let $A_{ij} = 1$ if child i at site j is assigned to an HS program and 0 otherwise; $Z_{ij}(1)$ is the potential frequency of parent reading to child if child *i* at site *j* is assigned to HS; $Z_{ij}(0)$ is the potential frequency of parent reading to child if the same child is assigned to the control condition. For simplicity, we dichotomize the mediator such that, under treatment a for *a* = 0, 1, $Z_{ij}(a) = 1$ if the parent read with a relatively high frequency and 0 otherwise. We consider three potential outcomes denoted by $Y_{ij}(1, Z_{ij}(1))$, the vocabulary score of child i at site j if the child is assigned to HS, $Y_{ij}(0, Z_{ij}(0))$, the vocabulary score of the same child if assigned to the control condition, and $Y_{ij}(1, Z_{ij}(0))$, the vocabulary score of this child if assigned to HS yet the parent would read to the child counterfactually with the same frequency as that under the control condition. Let $\boldsymbol{X_{ij}}$ denote the individual-level observed pretreatment covariates such as parents' education status.

The direct effect of HS on the vocabulary of child *i* at site *j*, defined as the HS effect on vocabulary without changing the frequency of parent reading to child, is $\delta_{ij}^{(D)} = Y_{ij}(1, Z_{ij}(0)) - Y_{ij}(0, Z_{ij}(0))$. This is called "the natural direct effect" by Pearl (2001) and "the pure direct effect" by Robins and Greenland (1992). The indirect effect, defined as the HS effect on vocabulary mediated by the program-induced change in parent reading to child, is $\delta_{ij}^{(I)} = Y_{ij}(1, Z_{ij}(1)) - Y_{ij}(1, Z_{ij}(0))$. This is called "the natural indirect effect" by Pearl and "the total indirect effect" by Robins and Greenland. Accordingly, we define the site-specific direct effect $\delta_j^{(D)} = E\left[\delta_{ij}^{(D)}|j\right]$ and indirect effect $\delta_j^{(I)} = E\left[\delta_{ij}^{(I)}|j\right]$ for site *j*, as well as the population average direct effect $\delta^{(D)} = E\left[\delta_j^{(D)}\right]$ and the population average indirect effect $\delta^{(I)} = E\left[\delta_j^{(I)}\right]$. In addition, we are interested in estimating the between-site variation in site-specific direct effect and indirect effect, denoted by var($\delta_j^{(D)}$) and var($\delta_j^{(I)}$), respectively. We have also hypothesized a nonzero covariance between the site-specific direct effect and indirect effect denoted by cov($\delta_j^{(D)}, \delta_j^{(I)}$).

5. RMPW Rationale

The definition of the population average direct effect and indirect effect involves three population average potential outcomes. Among them, $E\left[E\left(Y_{ij}\left(0, Z_{ij}(0)\right)|j\right)\right]$ and $E\left[E\left(Y_{ij}(1, Z_{ij}(1))|j\right)\right]$ can be identified with the observed mean outcome of the control group and that of the HS group, respectively, when the treatment is randomized. However, $E\left[E\left(Y_{ij}(1, Z_{ij}(0))|j\right)\right]$ is the average of the potential outcome of high reading frequency under HS and that of low reading frequency under HS proportionally weighted by the high reading frequency rate and the low reading frequency rate, respectively, under the control condition. Suppose that not only was the treatment randomized but also children were randomized to receive either a high frequency of reading from parents with probability $pr(Z(a) = 1)$ or a low frequency of reading with probability $pr(Z(a) = 0)$ under treatment *a* for *a* = 0, 1. To identify $E\left[E\left(Y_{ij}(1, Z_{ij}(0))|j\right)\right]$, we may transform the reading frequency rates under HS to resemble those under the control condition:

$$
\begin{aligned}
&E\left[E\left(Y_{ij}(1, Z_{ij}(0))|j\right)\right] \\
&= E\left[E\left(Y_{ij}(1,\ 1)|j\right)\right] \times pr\left(Z_{ij}(0) = 1\right) + E\left[E\left(Y_{ij}(1,\ 0)|j\right)\right] \times pr\left(Z_{ij}(0) = 0\right) \\
&= pr\left(Z_{ij}(1) = 1\right) \times E\left[E\left(Y_{ij}(1,\ 1) \times \frac{pr\left(Z_{ij}(0)=1\right)}{pr\left(Z_{ij}(1)=1\right)}|j\right)\right] \\
&+ pr\left(Z_{ij}(1) = 0\right) \times E\left[E\left(Y_{ij}(1,\ 0) \times \frac{pr\left(Z_{ij}(0)=0\right)}{pr\left(Z_{ij}(1)=0\right)}|j\right)\right] \qquad (1)
\end{aligned}
$$

in which the potential outcome of high reading frequency under HS, $Y_{ij}(1,\ 1)$, is weighted by the ratio of the probability of high reading frequency under the control condition to that under HS, $pr\left(Z_{ij}(0) = 1\right)/pr\left(Z_{ij}(1) = 1\right)$; the potential outcome of low reading frequency under HS, $Y_{ij}(1,\ 0)$, is weighted by the ratio of the probability of low reading frequency under the control condition to that under HS, $pr\left(Z_{ij}(0) = 0\right)/pr\left(Z_{ij}(1) = 0\right)$. The weights are therefore named Ratio-of-Mediator-Probability Weights (RMPW).

To implement this strategy in the current study in which individuals were not assigned at random to different mediator values, we estimate for each HS child at each site the conditional probability of being read to by the parent with the actual observed reading frequency under HS: $\theta_{Z_1} = \theta_{Z_1}(\mathbf{x}) = pr\left(Z_{ij} = 1|A_{ij} = 1, \mathbf{X}_{ij} = \mathbf{x}\right)$ if HS child *i* at site *j* was read to with a high frequency and $1 - \theta_{Z_1} = pr\left(Z_{ij} = 0|A_{ij} = 1, \boldsymbol{X}_{ij} = \mathbf{x}\right)$ if the child was read to with a low frequency. We also predict, for the same child, the conditional probability of being read to by the parent with the same reading frequency under the counterfactual control condition: $\theta_{Z_0} = \theta_{Z_0}(\mathbf{x}) = pr\left(Z_{ij} = 1|A_{ij} = 0, \boldsymbol{X}_{ij} = \mathbf{x}\right)$ and $1 - \theta_{Z_0} = pr\left(Z_{ij} = 0|A_{ij} = 0, \boldsymbol{X}_{ij} = \mathbf{x}\right)$.

To estimate the direct and indirect effects, we duplicate each HS child indicated by $D = 1$ and then merge the duplicated HS sample with the original sample including HS and control children indicated by $D = 0$. The weight is 1 for the control children and the duplicate HS children, is $\theta_{Z_0}/\theta_{Z_1}$ for the HS children with high parent reading frequency, and is $\left(1 - \theta_{Z_0}\right)/\left(1 - \theta_{Z_1}\right)$ for the HS children with low parent reading frequency. The weighting scheme is summarized in Table 1.

Table 1: Weights Applied to Estimate Potential Outcomes

	$E\left[E\left(Y_{ij}(0, Z_{ij}(0))\|j\right)\right]$	$E\left[E\left(Y_{ij}(1, Z_{ij}(1))\|j\right)\right]$	$E\left[E\left(Y_{ij}(1, Z_{ij}(0))\|j\right)\right]$
A_{ij}	0	1	1
D_{ij}	0	1	0

Z_{ij}	0, 1	0, 1	0	1
W_{ij}	1	1	$\frac{pr(Z_{ij}=0 \mid A_{ij}=0,\ \mathbf{X}_{ij}=\mathbf{x})}{pr(Z_{ij}=0 \mid A_{ij}=1,\ \mathbf{X}_{ij}=\mathbf{x})}$	$\frac{pr(Z_{ij}=1 \mid A_{ij}=0,\ \mathbf{X}_{ij}=\mathbf{x})}{pr(Z_{ij}=1 \mid A_{ij}=1,\ \mathbf{X}_{ij}=\mathbf{x})}$

The RMPW strategy can be applied to multivalued mediators as well. When the mediator is continuous, one may obtain the ratio of the estimated density of a given mediator value under the control condition to that under HS where each density is estimated as a function of the pretreatment covariates.

After stacking the data from all the sites, we analyze a 2-level weighted outcome model as follows:

$$Y_{ij} = \left(\delta^{(0)} + u_j^{(0)}\right) + \left(\delta^{(D)} + u_j^{(D)}\right) A_{ij} + \left(\delta^{(I)} + u_j^{(I)}\right) D_{ij} + e_{ij} \qquad (2)$$

in which

$$\begin{pmatrix} u_j^{(0)} \\ u_j^{(D)} \\ u_j^{(I)} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{var}(\delta_j^{(0)}) & & \\ \mathrm{cov}(\delta_j^{(0)}, \delta_j^{(D)}) & \mathrm{var}(\delta_j^{(D)}) & \\ \mathrm{cov}(\delta_j^{(0)}, \delta_j^{(I)}) & \mathrm{cov}(\delta_j^{(D)}, \delta_j^{(I)}) & \mathrm{var}(\delta_j^{(I)}) \end{pmatrix}\right)$$

When the following identification assumptions hold, from analyzing the above outcome model, we can obtain unbiased estimates of the population average direct effect $\delta^{(D)}$ and indirect effect $\delta^{(I)}$, along with their standard errors. In addition, we can obtain estimates of the between-site variation in the site-specific direct effect var($\delta_j^{(D)}$) and in the indirect effect var($\delta_j^{(I)}$) as well as the covariance cov($\delta_j^{(D)}, \delta_j^{(I)}$).

6. Identification Assumptions

For the potential outcomes to be well defined and identified under the counterfactual causal framework, the following assumption is necessary:

Assumption 1: *Stable unit treatment value assumption (SUTVA).* The potential mediator and outcome of an individual are not affected by the treatment and the mediator value assigned to other individuals within the same site or at the other sites. In the HS case, it requires that a child's parent reading frequency and vocabulary score do not depend on which treatment other children were assigned to either at the same site or at another site, and a child's vocabulary score also does not depend on whether other parents at the same site or at another site read frequently to their children.

RMPW identifies the direct and indirect effects under the following ignorability assumptions:

Assumption 2: No confounding of treatment-outcome relationship (*Ignorability assumption 1).* Treatment assignment is independent of the potential outcomes given a set of pretreatment covariates. This assumption requires that, within levels of the observed covariates, a randomly selected child in the HS group and a randomly selected child in the control group from the same site are expected to have the same potential vocabulary score associated with HS, denoted by Y(1, Z(1)), and the same potential vocabulary score associated with the control condition, denoted by Y(0, Z(0). That is, for $a = 0, 1$,

$$Y\big(a, Z(a)\big) \coprod A \,|\mathbf{X}, j$$

Assumption 3: No confounding of treatment-mediator relationship (*Ignorability assumption 2).* Treatment assignment is independent of the potential mediators given a set of pretreatment covariates. In the HS case, among those with the same pretreatment characteristics, parents of HS children and those of control children from the same site are expected to read with the same level of frequency. Namely, for $a = 0, 1$,

$$\mathrm{Z}(a) \coprod \mathrm{A}\ |\mathbf{X}, j$$

In the Head Start study, Assumptions 2 and 3 are satisfied because of the randomized treatment assignment.

Assumption 4: No confounding of mediator-outcome relationship within a treatment (*Ignorability assumption 3)*. Under each treatment and within levels of the observed pretreatment characteristics, the mediator value assignment is independent of the potential outcomes. In other words, there are no other pretreatment confounders of the relationship between parent reading frequency and vocabulary score, given the treatment and the observed pretreatment covariates. Namely, for $a = 0, 1$

$$\mathrm{Y}\big(a, \mathrm{Z}(a)\big) \coprod \mathrm{Z}(a)\ |\mathrm{A}, \mathbf{X}, j$$

Assumption 5: No confounding of mediator-outcome relationship across treatments (*Ignorability assumption 4)*. Within levels of the observed pretreatment characteristics, the mediator value assignment under one treatment is independent of the potential outcomes associated with an alternative treatment. In other words, there are no posttreatment confounders of the relationship between parent reading frequency and child vocabulary score given the observed pretreatment covariates. Namely, for $a = 0, 1$

$$\mathrm{Y}\big(a, \mathrm{Z}(a)\big) \coprod \mathrm{Z}(a')\ |\mathrm{A}, \mathbf{X}, j$$

To satisfy Assumptions 4 and 5, the mediator value assignment should also be randomized under each treatment within levels of the pretreatment covariates.

Assumptions 2-5 constitute the “sequential ignorability” assumptions used by Imai and colleagues (Imai, Keele, and Yamamoto, 2010; Imai, Keele, & Tingley, 2010) and similarly required by the RMPW method. Yet unlike multilevel path analysis, RMPW does not require the assumption of no treatment-by-mediator interaction.

7. Estimation

Standard multilevel software programs are not suitable for the current problem because the variance-covariance matrix is not invertible. We thus develop an alternative weighted least squares procedure.

The 2-level RMPW model (2) can be rewritten as

Level 1: $\boldsymbol{Y}_j = \boldsymbol{L}_j \boldsymbol{\beta}_j + \boldsymbol{e}_j$

Level 2: $\boldsymbol{\beta}_j = \boldsymbol{\gamma} + \boldsymbol{u}_j,\ \boldsymbol{u}_j \sim N(0, \mathbf{T})$

in which $\boldsymbol{L}_j = (\mathbf{1} \quad \boldsymbol{A}_j \quad \boldsymbol{D}_j)$, $\boldsymbol{\beta}_j = \left(\delta_j^{(0)} \quad \delta_j^{(D)} \quad \delta_j^{(I)}\right)^T$, $\boldsymbol{\gamma} = (\delta^{(0)} \quad \delta^{(D)} \quad \delta^{(I)})^T$, and

$$\mathbf{T} = Var\big(\boldsymbol{\beta}_j\big) = \begin{pmatrix} \mathrm{var}(\delta_j^{(0)}) & & \\ \mathrm{cov}(\delta_j^{(0)}, \delta_j^{(D)}) & \mathrm{var}(\delta_j^{(D)}) & \\ \mathrm{cov}(\delta_j^{(0)}, \delta_j^{(I)}) & \mathrm{cov}(\delta_j^{(D)}, \delta_j^{(I)}) & \mathrm{var}(\delta_j^{(I)}) \end{pmatrix}. \qquad (3)$$

Here $\mathbf{T}$ denotes the between-site variance-covariance matrix for the site-specific causal effects.

7.1 Estimation of the Between-Site Variance and Covariance and the Average Causal Effects

We start with estimating the site-specific causal effects $\boldsymbol{\beta}_j$.

$$\widehat{\boldsymbol{\beta}}_j = \left(\boldsymbol{L}_j^T \boldsymbol{W}_j \boldsymbol{L}_j\right)^{-1} \boldsymbol{L}_j^T \boldsymbol{W}_j \boldsymbol{Y}_j.$$

The estimate has the following variability:

$$\mathbf{G} = \mathrm{Var}\left(\widehat{\boldsymbol{\beta}}_j\right) = \mathrm{Var}\left(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j + \boldsymbol{\beta}_j\right) = \mathrm{Var}\left(\boldsymbol{\beta}_j\right) + \mathrm{Var}\left(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\right) = \mathbf{T} + \mathbf{V}$$

where $\mathbf{V} = \mathrm{Var}\left(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\right)$ denotes the sampling variability of the site-specific causal effect estimates. We obtain an estimate of $\mathbf{T}$ as follows:

$$\widehat{\mathbf{T}} = \widehat{\mathbf{G}} - \widehat{\mathbf{V}}$$

$$\widehat{\mathbf{G}} = \frac{1}{J}\sum_{j=1}^{J} \frac{n_j}{\bar{n}} \left(\widehat{\boldsymbol{\beta}}_j - \widehat{\boldsymbol{\gamma}}\right)\left(\widehat{\boldsymbol{\beta}}_j - \widehat{\boldsymbol{\gamma}}\right)^T$$

$$\widehat{\mathbf{V}} = \frac{1}{J}\textstyle\sum_{j=1}^{J} \frac{n_j}{\bar{n}} \widehat{\mathbf{V}}_j = \frac{1}{J}\sum_{j=1}^{J} \frac{n_j}{\bar{n}} \left(\boldsymbol{L}_j^T \boldsymbol{W}_j \boldsymbol{L}_j\right)^{-1} \boldsymbol{L}_j^T \boldsymbol{W}_j \mathrm{Var}(\boldsymbol{Y}_j | \boldsymbol{L}_j) \boldsymbol{W}_j \boldsymbol{L}_j \left(\boldsymbol{L}_j^T \boldsymbol{W}_j \boldsymbol{L}_j\right)^{-1}. \qquad (4)$$

We use $n_j^{(E)}$ and $n_j^{(C)}$ to respectively represent the sampled number of experimental units and that of control units at site *j*. Including the duplicates, there are totally $2n_j^{(E)}+ \ n_j^{(C)}$ units at site *j*. Therefore, $\boldsymbol{L}_j$ is a $\left[2n_j^{(E)} + n_j^{(C)}\right] \times 3$ matrix with two sub-matrices: the one for the experimental group is $2n_j^{(E)} \times 3$ in dimensions, in which every two rows represent an original unit and its duplicate; the other for the control group is $n_j^{(C)} \times 3$ in dimensions. $\boldsymbol{W}_j$ denotes a $\left[2n_j^{(E)} + n_j^{(C)}\right] \times \left[2n_j^{(E)} + n_j^{(C)}\right]$ diagonal matrix for the weights applied to the observations at site *j*; $\boldsymbol{W}_j \mathrm{Var}(\boldsymbol{Y}_j | \boldsymbol{L}_j) \boldsymbol{W}_j$ is a $\left[2n_j^{(E)} + n_j^{(C)}\right] \times \left[2n_j^{(E)} + n_j^{(C)}\right]$ matrix with two primary sub-matrices along the diagonal. One is for the experimental units and their duplicates, which is $2n_j^{(E)} \times 2n_j^{(E)}$ in dimensions with $n_j^{(E)}$ 2×2 sub-sub-matrices along the diagonal each taking the form

$$\begin{pmatrix} W_{ij}^2 \sigma_{j,E}^2 & W_{ij}\sigma_{j,E,D} \\ W_{ij}\sigma_{j,E,D} & \sigma_{j,D}^2 \end{pmatrix} \qquad (5)$$

for the ith experimental unit at site j. Here W_{ij} is the RMPW for the unit; $\sigma_{j,E}^2$ is the error variance of the RMPW adjusted outcome for the original experimental units at site *j*; $\sigma_{j,D}^2$ is the error variance for the duplicate experimental units at site *j*; $\sigma_{j,E,D}$ denotes the covariance between the two errors for an experimental unit at site *j*. We estimate the covariance as $\rho_{j,E,D} \times \hat{\sigma}_{j,E} \times \hat{\sigma}_{j,D}$, in which $\rho_{j,E,D}$ is the corresponding correlation. The other sub-matrix of $\boldsymbol{W}_j \mathrm{Var}(\boldsymbol{Y}_j | \boldsymbol{L}_j) \boldsymbol{W}_j$ is for the observations in the control group, which is an $n_j^{(C)} \times n_j^{(C)}$ diagonal matrix with $\sigma_{j,C}^2$ on the diagonal, which is the error variance for the control units at site *j*.

When the site size is too small as is often the case in the HSIS data, the information at each site is too limited for estimating the error variance in either the treatment or the control group at a site. Instead of estimating the heterogeneous model-based variance site by site, we estimate the homogeneous model-based variance by pooling the units from all the sites and obtaining the average σ_E^2, σ_D^2, $\sigma_{E,D}$ and σ_C^2.

The potential difficulty in estimating the above model-based variance is that $\rho_{E,D}$ is unknown. Our preliminary simulation shows that the estimation results are sensitive to the values of $\rho_{E,D}$. Hence, we further develop a robust estimator by replacing $\hat{\sigma}_{j,E}^2, \hat{\sigma}_{j,D}^2, \hat{\sigma}_{j,E,D}$ and $\hat{\sigma}_{j,C}^2$ with $\hat{r}_{ij,E}^2$, $\hat{r}_{ij,D}^2$, $\hat{r}_{ij,E} \times \hat{r}_{ij,D}$ and $\hat{r}_{kj,C}^2$, respectively. Here $\hat{r}_{ij,E} = \sqrt{W_{ij}}\left(Y_{ij} - \boldsymbol{L}_{ij}\widehat{\boldsymbol{\beta}}_j\right)$ for RMPW adjusted original experimental unit *i* at site *j*; $\hat{r}_{ij,D} = Y_{ij} - \boldsymbol{L}_{ij}\widehat{\boldsymbol{\beta}}_j$ for duplicate experimental unit *i;* $\hat{r}_{kj,C} = Y_{kj} - \boldsymbol{L}_{kj}\widehat{\boldsymbol{\beta}}_j$ for control unit *k* at site *j*.

Having obtained $\hat{\mathbf{V}}_j$, we then estimate the variance-covariance matrix of the site-specific effects:

$$\hat{\mathbf{T}} = \frac{1}{J}\sum_{j=1}^{J}\frac{n_j}{\bar{n}}\left[\left(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\gamma}}\right)\left(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\gamma}}\right)^T - \hat{\mathbf{V}}_j\right] \tag{6}$$

where

$$\hat{\boldsymbol{\gamma}} = \left(\sum_{j=1}^{J} n_j\right)^{-1}\sum_{j=1}^{J} n_j\hat{\boldsymbol{\beta}}_j \tag{7}$$

which is the estimate of the population average causal effects. For the Heywood cases, in which $\hat{\mathbf{T}}$ is negative definite (i.e. there are negative variances or some correlations are greater than one in magnitude), we replace the negative variances and the corresponding covariances with 0.

With $\hat{\mathbf{T}}$ at hand, we can further obtain the model-based estimator of the sampling variability of the point estimate of the population average causal effects:

$$\widehat{Var}(\hat{\boldsymbol{\gamma}}) = \left(\sum_{j=1}^{J}\boldsymbol{n}_j\right)^{-\mathbf{2}}\sum_{j=1}^{J}\boldsymbol{n}_j^2\left(\hat{\mathbf{T}} + \hat{\boldsymbol{V}}_j\right) \tag{8}$$

7.2 Hypothesis testing for the Between-Site Variance and Covariance

We propose the following hypothesis testing procedure for the variance of the site-specific direct and indirect effects. $H_0\colon \mathrm{var}\left(\delta_j^{(D)}\right) = 0$ can be tested by means of the statistic

$$\sum_{j=1}^{J}\left(\hat{\boldsymbol{\beta}}_{j,2} - \hat{\boldsymbol{\gamma}}_2\right)^2 / \left(\boldsymbol{V}_j\right)_{22} \tag{9}$$

which is distributed χ^2 with J-1 degrees of freedom. Similarly, $H_0\colon \mathrm{var}\left(\delta_j^{(I)}\right) = 0$ can be tested by means of the statistic

$$\sum_{j=1}^{J}\left(\hat{\boldsymbol{\beta}}_{j,3} - \hat{\boldsymbol{\gamma}}_3\right)^2 / \left(\boldsymbol{V}_j\right)_{33} \tag{10}$$

which is also distributed χ^2 with J-1 degrees of freedom.

8. Estimation Procedure

To estimate the causal effects and their variations across sites for the HS sample, we first impute the missing data in the outcome and the covariates, and generate five imputed data sets. We estimate each data set separately, and combine the estimation results over the five. The estimation procedures are as follows:

Step 1: Specify the propensity score model. We specify the propensity score model for the mediator under each treatment condition, conditional on the observed pretreatment covariates. Analyzing data from the HS sample, we estimate the propensity scores of high and low reading frequency under the HS condition for an HS unit, θ_{Z_1} and $1 - \theta_{Z_1}$. After fitting another logistic regression model to the control group, we apply the model to predict the propensity scores of high and low reading frequency under the counterfactual control condition for an HS unit, θ_{Z_0} and $1 - \theta_{Z_0}$.

Step 2: Identify the common support. Under each treatment condition, among the children who have the same propensity scores, those with high reading frequency are expected to have their counterparts with low reading frequency. Therefore, we compare the distribution of the logit of $\hat{\theta}_{Z_1}$ and $\hat{\theta}_{Z_0}$ across the four treatment-by-mediator groups, and exclude the cases in which the propensity scores do not overlap across the four groups. One may add 20% of a standard deviation of the logit of each propensity score at each end to expand the range of the common support (Austin, 2011). The following analysis will be based on the cases within the common support.

Step 3: Check balance in covariate distribution across the treatment-by-mediator combinations. When the sequential ignorability holds, inverse-probability-of-treatment weighting (IPTW, Robins, 2000) transforms the data to approximate a sequential randomized design. We assign the weight $pr(Z=1|A=1)/\theta_{Z_1}$ to the HS children with high reading frequency, $pr(Z=0|A=1)/\left(1-\theta_{Z_1}\right)$ to the HS children with low reading frequency, $pr(Z=1|A=0)/\theta_{Z_0}$ to the untreated children with high reading frequency, and $pr(Z=0|A=0)/\left(1-\theta_{Z_0}\right)$ to the untreated children with low reading frequency. For each observed covariate, we conduct pairwise comparisons between these four subgroups. All of the covariates show a standardized difference less than .25 after weighting.

Step 4: Create a duplicate, and estimate $\boldsymbol{W}_j$. We then duplicate the HS sample and assign RMPW as shown in Table 1.

Step 5: Estimate $\widehat{\boldsymbol{\beta}}_j$ *within each site, and then estimate* $\widehat{\boldsymbol{\gamma}}$ by taking the weighted average of $\widehat{\boldsymbol{\beta}}_j$ as shown in (7)

Step 6: Estimate $\widehat{\boldsymbol{T}}$ *and test its significance.* We estimate $\widehat{\boldsymbol{V}}_j$ according to (4) and estimate $\widehat{\boldsymbol{T}}$ according to (6). We then test the significance of the variance of the site-specific direct and indirect effects based on (9) and (10). Since the correlation between the observed and the counterfactual outcome for the experimental children is unknown, we use the robust estimator.

Step 7: Obtain standard error of $\widehat{\boldsymbol{\gamma}}$. Based on the estimates of $\widehat{\boldsymbol{T}}$ and $\widehat{\boldsymbol{V}}_{\boldsymbol{j}}$, we can estimate the standard error of $\widehat{\boldsymbol{\gamma}}$ according to (8).

By combining the estimation results over the five imputed Head Start data sets, we obtain the average estimates of the causal effects and the variance estimates of the site-specific effects. The estimated average direct effect is 3.81 (SE=1.64, *t*=2.32, *p*=0.01), about 11% of a standard deviation of the outcome, with an estimated variance of 251.96. The estimated average indirect effect is 0.13 (SE=0.21, *t*=0.61, *p*=0.27), with an estimated variance of 4.15, and the estimated covariance between the site-specific direct and indirect effects is -4.42. According to these results, the HS programs improved child vocabulary directly rather than through increasing parent reading to child. The HS-induced change in parent reading to child alone does not significantly improve child vocabulary. Only the direct effect appears to vary across the sites.

9. Simulations

We conduct a series of Monte Carlo simulations to assess the performance of the above procedure in estimating the population average direct effect and indirect effect and the joint distribution of the site-specific direct effect and indirect effect. We focus on the case of a binary randomized treatment, a binary mediator, and a continuous outcome.

9.1 Sample Size

We select four different sets of sample size: 1) $J=100$ and $n_j=150$, which represents a large sample size within each site; 2) $J=300$ and $n_j=10$, which represents a small sample size within each site; 3) $J=100$ and n_j ranges from 100 to 200; and 4) $J=328$ and n_j ranges from 1 to 46, with a mean of 7.5. The last set resembles the Head Start data. For the first two sets of sample size, we specify the site-specific probability of treatment assignment $\bar{A}_j$ to be constantly 0.5. For the latter two sets of sample size, we specify that $\bar{A}_j \sim Beta(14,10)$, which is similar to the Head Start data. For each case, we generate 1,000 random samples.

9.2 Data Generation

We generate three independent covariates X_1, X_2, and X_3 with identical distributions: $X_{ij} = \bar{X}_j + e_{X_{ij}}$, in which $\bar{X}_j \sim N(0,1)$ and $e_{X_{ij}} \sim N(0,1)$. The treatment assignment is random with probability $\bar{A}_j$ within each site. We then specify the distribution of the site-specific direct effect and indirect effect,

$$\begin{pmatrix} \delta_j^{(D)} \\ \delta_j^{(I)} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta^{(D)} \\ \delta^{(I)} \end{pmatrix}, \begin{pmatrix} \mathrm{var}(\delta_j^{(D)}) & \mathrm{cov}(\delta_j^{(D)}, \delta_j^{(I)}) \\ \mathrm{cov}(\delta_j^{(D)}, \delta_j^{(I)}) & \mathrm{var}(\delta_j^{(I)}) \end{pmatrix} \right).$$

To examine how true variances of the causal effects affect the estimation results, we compare two different distributions of the site-specific direct effect and indirect effect:

(1) $\begin{pmatrix} \delta_j^{(D)} \\ \delta_j^{(I)} \end{pmatrix} \sim N\left(\begin{pmatrix} 3 \\ 0.1 \end{pmatrix}, \begin{pmatrix} 6.25 & 4.5 \\ 4.5 & 9 \end{pmatrix} \right)$, which represents relatively large variances;

(2) $\begin{pmatrix} \delta_j^{(D)} \\ \delta_j^{(I)} \end{pmatrix} \sim N\left(\begin{pmatrix} 3 \\ 0.1 \end{pmatrix}, \begin{pmatrix} 0.5 & -0.1 \\ -0.1 & 0.3 \end{pmatrix} \right)$, which represents relatively small variances.

We then generate the observed outcome and mediator from the following model, allowing for an interaction between the treatment and the mediator:

$$logit\{P(Z_{ij} = 1 | A_{ij}, \boldsymbol{X}_{ij})\} = \alpha_{0j} + \alpha_{1j} A_{ij} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j$$

$$Y_{ij} = \theta_{0j} + \theta_{1j} A_{ij} + \theta_{2j} Z_{ij} + \theta_{3j} A_{ij} Z_{ij} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\theta}_j + e_{ij}$$

in which $e_{ij} \sim N(0, \sigma_e^2)$.

Valeri & VanderWeele (2013) derived the expressions of the direct and indirect effects from the above model:

$$\delta_j^{(D)} = \theta_{1j} + \theta_{3j} \cdot \frac{exp(\alpha_{0j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)}{1 + exp(\alpha_{0j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)}$$

$$\delta_j^{(I)} = (\theta_{2j} + \theta_{3j}) \cdot \left\{ \frac{exp(\alpha_{0j} + \alpha_{1j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)}{1 + exp(\alpha_{0j} + \alpha_{1j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)} - \frac{exp(\alpha_{0j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)}{1 + exp(\alpha_{0j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)} \right\}$$

We first set values for $\alpha_{0j}, \alpha_{1j}, \boldsymbol{\alpha}_j, \theta_{0j}, \theta_{2j}, \boldsymbol{\theta}_j$ and σ_e^2, which are invariant across simulations:

$$\alpha_{0j} \sim N(0.18, 0.0001), \alpha_{1j} \sim N(-0.25, 0.0001)$$

$$\alpha_j^{(1)} \sim N(-0.1, 0.0025), \alpha_j^{(2)} = -0.2, \alpha_j^{(3)} = 0.25$$

$$\theta_{0j} \sim N(0.5, 0.49), \theta_{2j} \sim N(-0.17, 0.01)$$

$$\theta_j^{(1)} \sim N(0.4, 0.64), \theta_j^{(2)} = 0.6, \theta_j^{(3)} = 0.9$$

$$\sigma_e^2 = 0.09$$

With $\delta_j^{(D)}$ and $\delta_j^{(I)}$ already specified, we then determine θ_{3j} and θ_{1j}:

$$\theta_{3j} = \delta_j^{(I)} \Big/ \left\{ \frac{exp(\alpha_{0j} + \alpha_{1j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)}{1 + exp(\alpha_{0j} + \alpha_{1j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)} - \frac{exp(\alpha_{0j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)}{1 + exp(\alpha_{0j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)} \right\} - \theta_{2j}$$

$$\theta_{1j} = \delta_j^{(D)} - \theta_{3j} \frac{exp(\alpha_{0j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)}{1 + exp(\alpha_{0j} + \boldsymbol{X}_{ij}{}^T \boldsymbol{\alpha}_j)}$$

Finally, we generate the outcome and the mediator for each case.

9.3 Evaluation Criteria

The evaluation criteria include: (1) bias in the point estimate; (2) mean squared error (MSE); and (3) CI (Confidence Interval) coverage rate.

9.4 Simulation Results

The bias, MSE, and coverage rates for the CIs of the estimates of the population average direct and indirect effects are presented in Table 2, with a comparison among different sample sizes, treatment assignment probabilities and true variances of the effects.

Table 2: Comparison of the average estimates across different sample sizes

	Bias		MSE		% CI coverage	
True Variance of Direct and Indirect Effects	Big	Small	Big	Small	Big	Small
Direct effect estimate ($\hat{\delta}^{(D)}$)						
$\mathrm{J} = 100,\ n_j = 150$[a]	0.01	0.00	0.13	0.01	96.1	93.5
$\mathrm{J} = 100,\ n_j = 100 \sim 200$[b]	-0.01	0.00	0.13	0.01	95.8	92.8
$\mathrm{J} = 300,\ n_j = 10$[a]	0.01	0.02	0.49	0.02	95.4	95.8
$\mathrm{J} = 328,\ n_j = 1 \sim 46$[b]	0.05	0.02	0.57	0.02	96.2	96.2
Indirect effect estimate ($\hat{\delta}^{(I)}$)						
$\mathrm{J} = 100,\ n_j = 150$	0.02	0.00	0.10	0.00	94.7	92.7
$\mathrm{J} = 100,\ n_j = 100 \sim 200$	-0.02	0.00	0.10	0.00	94.2	95.4
$\mathrm{J} = 300,\ n_j = 10$	-0.02	-0.02	0.03	0.00	93.9	74.9
$\mathrm{J} = 328,\ n_j = 1 \sim 46$	-0.03	-0.03	0.07	0.00	89.9	78.3

Note: [a] $\bar{A}_j = 0.5$ *for these two cases.* [b] $\bar{A}_j \sim Beta(14, 10)$ *for these two cases.*

As Table 2 shows, the estimates of the population average direct and indirect effects are almost always unbiased and have a high CI coverage rate under different settings. When the site size is small, and when $\bar{A}_j$ and n_j vary across sites, there is a minor increase in the bias and MSE and a reduction in the CI coverage rate. When the true variances of the site-specific effects increase, the MSE of the average effect estimates also tends to increase.

Table 3 presents the simulation results for the model-based and robust estimators of the variance and covariance of the site-specific effects. We compare across different sample sizes, treatment assignment probabilities, and true variances of the effects. When the site size is relatively small, the information at each site is limited for estimating the site-specific error variance. Hence, we use the homogeneous model-based variance estimator here. Since the true direct and indirect effects are set to be constant across the sites, the correlation between the observed and the counterfactual potential outcome for the experimental units, namely $\rho_{E,D}$ in Section 7, is 1.

According to the results summarized in Table 3, when the site size is relatively big, the robust and model-based estimators perform similarly well. When the site size is relatively small, in comparison with the model-based estimators, the robust estimators have smaller bias but bigger MSE for $\widehat{var}(\delta_j^{(I)})$ and $\widehat{cov}(\delta_j^{(D)}, \delta_j^{(I)})$, and have much bigger bias and MSE for $\widehat{var}(\delta_j^{(D)})$. However, these simulation results are based on the fact that $\rho_{E,D}$ is correctly specified. When $\rho_{E,D}$ is unknown, which is typically the case in real data analyses, the robust estimator is more reliable.

Table 3: Comparison of the variance estimates among different samples

	Robust Estimator				Model-based Estimator			
	Bias		MSE		Bias		MSE	
True Variance	Big	Small	Big	Small	Big	Small	Big	Small
Variance of direct effect estimate ($\widehat{var}(\delta_j^{(D)})$)								
$\mathrm{J} = 100,\ n_j = 150$[a]	-0.85	-0.04	7.36	0.02	-1.22	-0.12	8.08	0.03
$\mathrm{J} = 100,\ n_j = 100\sim200$[b]	-0.72	-0.03	6.74	0.01	-1.08	-0.12	7.43	0.03
$\mathrm{J} = 300,\ n_j = 10$[a]	41.99	1.57	2370.78	3.19	3.61	-0.45	216.94	0.24
$\mathrm{J} = 328,\ n_j = 1\sim46$[b]	64.65	2.65	4791.52	7.77	3.94	-0.46	239.06	0.25
Variance of indirect effect estimate ($\widehat{var}(\delta_j^{(I)})$)								
$\mathrm{J} = 100,\ n_j = 150$	0.57	0.02	8.50	0.01	0.48	0.01	8.24	0.01
$\mathrm{J} = 100,\ n_j = 100\sim200$	0.40	0.02	8.07	0.01	0.31	0.02	7.88	0.01
$\mathrm{J} = 300,\ n_j = 10$	-0.76	-0.03	17.19	0.02	-2.03	-0.09	16.52	0.02
$\mathrm{J} = 328,\ n_j = 1\sim46$	-0.13	0.00	25.11	0.03	-1.26	-0.05	21.41	0.02
Covariance between direct and indirect effect estimates ($\widehat{cov}(\delta_j^{(D)}, \delta_j^{(I)})$)								
$\mathrm{J} = 100,\ n_j = 150$	0.22	0.01	3.34	0.01	0.19	0.01	3.29	0.01
$\mathrm{J} = 100,\ n_j = 100\sim200$	0.01	0.00	2.74	0.01	0.00	0.00	2.69	0.01
$\mathrm{J} = 300,\ n_j = 10$	0.82	0.07	18.68	0.03	-2.19	0.10	13.99	0.01
$\mathrm{J} = 328,\ n_j = 1\sim46$	-0.57	0.04	20.82	0.03	-2.99	0.10	16.50	0.01

Note: [a] $\bar{A}_j = 0.5$ *for these two cases.*

[b] $\bar{A}_j \sim Beta(14, 10)$ *for these two cases.*

The simulation for the case that is similar in structure to the Head Start data reveals that there are considerably large bias and MSE in $\widehat{var}(\delta_j^{(D)})$. In general, when the site size is relatively small, the variance estimate of the site-specific direct effect becomes particularly unreliable.

To compare the RMPW approach with path analysis, we also conduct simulations in the case of a binary randomized treatment, a continuous mediator, and a continuous outcome. This is because standard multilevel path analysis does not apply when the mediator is binary. Our results show that, in comparison with multilevel path analysis, the RMPW approach is much less demanding computationally, and the bias in $\widehat{var}(\delta_j^{(D)})$ and $\widehat{var}(\delta_j^{(I)})$ is much smaller if the outcome model is misspecified, especially when the true variances are relatively big. Due to the space limit, we do not display the detailed results here.

10. Conclusions

The RMPW extension to data from multi-site trials provides an important alternative to the existing methods for multilevel mediation analysis. This approach relaxes the assumption of no treatment-by-mediator interaction and simplifies the outcome model specification. More importantly, the RMPW approach enables researchers to investigate possible heterogeneity of mediation mechanisms across sites. Such empirical information is essential for improving future intervention designs. However, the RMPW approach faces some potential challenges often shared by other competing methods. For example, when selection mechanisms vary across sites, site-specific propensity score models may be overfitted; omitted pretreatment covariates may cause bias; and there is a lack of adjustment strategies for post-treatment covariates. In light of these potential constraints, increasing the sample size at each site and collecting comprehensive pretreatment information are essential. Conceptualizing and investigating the relationships among multiple mediators may provide a solution for handling observed post-treatment covariates. Sensitivity analysis may be conducted to assess the impact of omitted pretreatment and posttreatment confounders.

References

Austin, P. C. (2011). Optimal caliper widths for propensity‐score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical statistics, 10(2), 150-161.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. Psychological Methods, 11(2), 142–163.

Hong, G. (2010). Ratio of mediator probability weighting for estimating natural direct and indirect effects. In JSM Proceedings, Biometrics Section. Alexandria, VA: American Statistical Association, pp.2401-2415.

Hong, G., Deutsch, J., & Hill, H. D. (2011). Parametric and non-parametric weighting methods for estimating mediation effects: An application to the National Evaluation of Welfare-to-Work Strategies. In JSM Proceedings, Social Statistics Section. Alexandria, VA: American Statistical Association, pp.3215-3229.

Hong, G. (in press). Causality in a Social World: Moderation, Mediation, and Spill-over. West Sussex, UK: Wiley-Blackwell.

Hong, G., & Nomi, T. (2012). Weighting methods for assessing policy effects mediated by peer change. Journal of Research on Educational Effectiveness special issue on the statistical approaches to studying mediator effects in education research, 5(3), 261–289.Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. Psychological Methods, 15(4), 309–334.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. Statistical Science, 25(1), 51–71.

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. Psychological Methods, 8(2), 115–128.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. Econometrica, 75(1), 83–119.

Krull, J. L., & Mackinnon, D. P. (1999). Multilevel Mediation Modeling in Group-Based Intervention Studies. Evaluation Review, 23(4), 418–444.

Pearl, J. (2001). Direct and indirect effects. In Proceedings of the seventeenth conference on uncertainty in artificial intelligence (pp. 411–420).

Pituch, K. A., Murphy, D. L., & Tate, R. L. (2009). Three-Level Models for Indirect Effects in School- and Class-Randomized Experiments in Education. The Journal of Experimental Education, 78(1), 60–95.

Raudenbush, S. W., Reardon, S. F., & Nomi, T. (2012). Statistical Analysis for Multisite Trials Using Instrumental Variables With Random Coefficients. Journal of Research on Educational Effectiveness, 5(3), 303–332.

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.

Reardon, S. F., Unlu, F., Zhu, P., & Bloom, H. S. (2014). Bias and Bias Correction in Multisite Instrumental Variables Analysis of Heterogeneous Mediator Effects. Journal of Educational and Behavioral Statistics, 39(1), 53-86.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In Statistical models in epidemiology, the environment, and clinical trials (pp. 95-133). Springer New York.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology, 3(2), 143–155.

Puma, M., et al. (2010). Head Start impact study: Final report. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.