

# UC Riverside

## UC Riverside Previously Published Works

### **Title**

Causal mediation analysis with multiple mediators in the presence of treatment noncompliance

### **Permalink**

<https://escholarship.org/uc/item/6mp9m30w>

### **Journal**

Statistics in Medicine, 37(11)

### **ISSN**

0277-6715

### **Authors**

Park, Soojin  
Kürüm, Esra

### **Publication Date**

2018-05-20


### **DOI**

10.1002/sim.7632

Peer reviewed

**RESEARCH ARTICLE**

# Causal mediation analysis with multiple mediators in the presence of treatment noncompliance

Soojin Park<sup>1</sup>  | Esra Kürüm<sup>2</sup> 

<sup>1</sup>Graduate School of Education, University of California, Riverside, Riverside, CA, USA

<sup>2</sup>Department of Statistics, University of California, Riverside, Riverside, CA, USA

**Correspondence**

Soojin Park, Graduate School of Education, University of California, Riverside, Riverside, CA, USA.  
Email: soojinp@ucr.edu

Randomized experiments are often complicated because of treatment noncompliance. This challenge prevents researchers from identifying the mediated portion of the intention-to-treated (ITT) effect, which is the effect of the assigned treatment that is attributed to a mediator. One solution suggests identifying the mediated ITT effect on the basis of the average causal mediation effect among compliers when there is a single mediator. However, considering the complex nature of the mediating mechanisms, it is natural to assume that there are multiple variables that mediate through the causal path. Motivated by an empirical analysis of a data set collected in a randomized interventional study, we develop a method to estimate the mediated portion of the ITT effect when both multiple dependent mediators and treatment noncompliance exist. This enables researchers to make an informed decision on how to strengthen the intervention effect by identifying relevant mediators despite treatment noncompliance. We propose a nonparametric estimation procedure and provide a sensitivity analysis for key assumptions. We conduct a Monte Carlo simulation study to assess the finite sample performance of the proposed approach. The proposed method is illustrated by an empirical analysis of JOBS II data, in which a job training intervention was used to prevent mental health deterioration among unemployed individuals.

**KEYWORDS**

causal mediation analysis, compliers-average causal mediation effect, multiple mediators, treatment noncompliance

## 1 | INTRODUCTION

Randomized experimental data are often used to study the causal mechanism between a treatment (or an intervention) and a health disorder. Establishing this causal mechanism aids the researchers in understanding the benefits of a treatment and allows discovery of how to make a treatment more effective and cost-efficient. However, randomized experiments are often complicated by treatment noncompliance, which occurs when subjects do not adhere with the assigned treatment. To address this challenge, researchers often use a naive approach, in which they conduct causal mediation analysis with the assigned treatment as if it is the actual treatment. However, such an approach can lead to biased results.<sup>1</sup> In this paper, we develop a flexible and intuitive causal mediation framework that enables researchers to study the causal mechanism between a treatment and an outcome in the presence of treatment noncompliance and multiple mediators.

Our motivating data come from the JOBS II interventional study,<sup>2-4</sup> which was designed to help unemployed individuals in lowering their depression levels. Researchers have taken a special interest in depression among unemployed individuals

because the job searching process is known to be strenuous and can lead to the emergence of depressive symptoms. A substantive body of literature has explored the relationship between depression and unemployment. For instance, McGee and Thompson, Montgomery Jr et al, and Whooley et al<sup>5-7</sup> have established that there is a positive association between depression rate and the duration of unemployment. Price et al<sup>8</sup> suggested that interventions to help unemployed individuals to return back to work force may be one way to reduce their risk of developing depression. The JOBS II intervention studied the effectiveness of an identified intervention—in this case, a job training seminar—on lowering depressive symptoms of unemployed workers. In this study, 1801 subjects were randomly assigned to either a treatment group or to a control group. Subjects in the treatment group was encouraged to attend 5 half-day job-search seminars, which were designed to help the subjects improve their job searching skills; the control group had only received a booklet of job-searching strategies. This booklet was also mailed to the respondents that were invited to the seminar. After the job-search seminar, each group was mailed a survey 2, 6, and 24 months after the intervention. The data from this study include demographic variables, measures of depression, self-esteem, and sense of mastery. The noncompliance rate was 46%.

Two aspects of the JOBS II project motivated this study. First, to investigate mediating mechanisms through which the job training seminar impacts depressive symptoms, it is important to consider those who did not comply with the assigned treatment. Treatment noncompliance is a manageable problem in terms of identifying the intention-to-treated (ITT) effect, that is, the effect of the treatment assignment status regardless of whether the subjects actually received the treatment or not. However, it presents challenges in identifying the mediated portion of the ITT effect,<sup>1</sup> which prevents researchers from investigating mediating mechanisms in the presence of treatment noncompliance. To solve this identification problem, Yamamoto<sup>1</sup> proposed an alternative approach to identify the mediated portion of the ITT effect on the basis of average causal mediation effects (ACMEs) among compliers. The ACMEs among compliers were defined as the expected change in the outcome among compliers in response to the change in mediators from the value that would have been observed under the treatment to the value under the control condition.

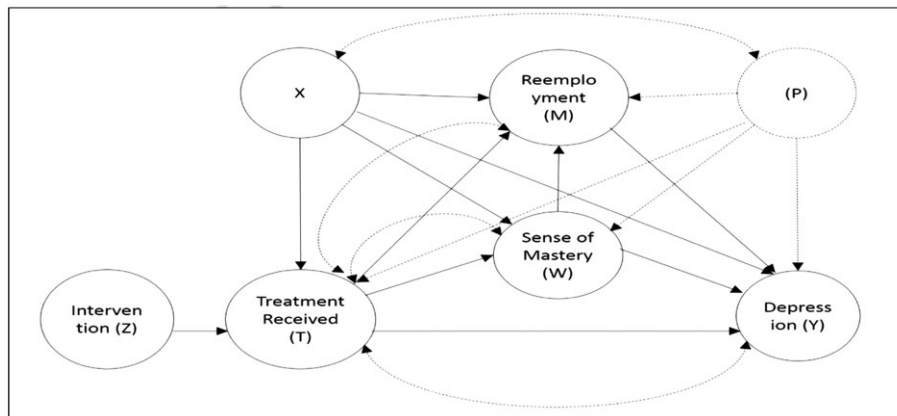
Second, considering the complexity of the causal mechanisms, it is natural to assume that multiple variables (a sense of mastery and reemployment in the JOBS II example) mediate through the causal path. Despite recent developments in causal mediation analysis, there is a gap in the literature in terms of studies on multiple causal mechanisms.<sup>9-11</sup> This may be due to the strong assumptions required to identify ACME, particularly when mediators influence one another. More specifically, when mediators influence one another, the ACME (defined when mediators are considered one at a time) is not nonparametrically identified unless one is willing to either make the no-interaction effect assumption in the treatment-mediator relationship<sup>12</sup> or it is partially identified up to sensitivity parameters.<sup>10,13</sup> In contrast, VanderWeele and Vansteelandt<sup>11</sup> proposed an approach considering all mediators jointly instead of one at a time. However, all of these approaches to identify the ACME in the presence of multiple mediators were developed assuming that all subjects complied with the treatment, which may not be a realistic scenario in randomized experiments such as the JOBS II study.

Prior analysis of JOBS II data established the effectiveness of job training on lowering the depression levels of unemployed individuals<sup>4,8,14,15</sup>; in contrast, our analysis addresses 2 challenges simultaneously: multiple mediators and treatment noncompliance. More specifically, in addition to studying the effectiveness of job training in the presence of treatment noncompliance, our interest also centers on investigating the mediating mechanisms—that is, why and how the job training seminar lowers depressive symptoms. We hypothesize that attending a job training seminar promotes sense of mastery (that is, the composite measure of job-seeking self-efficacy, self-esteem, and locus of control) and reemployment among unemployed workers, both of which affect their depressive symptoms. Diagrams of our causal structural models are presented in Figure 1. The estimation procedure in our causal mediation approach is built on Yamamoto<sup>1</sup> and VanderWeele and Vansteelandt<sup>11</sup> but is modified to accommodate the presence of treatment noncompliance and multiple mediators. In addition, we contribute to the literature by developing a sensitivity analysis for our motivating data to examine the sensitivity of effect estimates in case of violating key assumptions.

The remainder of the paper is organized as follows. In Section 2, we introduce the design of the JOBS II project and related studies. In Section 3, we explain complications in identifying a mediated ITT effect with multiple mediators in the presence treatment noncompliance and introduce our proposed approach to address this issue. In Section 3, we use a Monte Carlo simulation to study the impact of noncompliance and the effect heterogeneity on statistical power and the performance of the proposed approach. In Section 4, we apply our approach to the JOBS II study. Finally, we conclude with a discussion in Section 5. The proof and technical details can be found in the Appendix.

## 2 | JOBS II INTERVENTION PROJECT

Previous studies suggest that job loss has harmful effects on a worker's social, mental, and physical health.<sup>16-18</sup> The JOBS II project is a randomized trial that was developed at the University of Michigan to explore the impacts of an intervention



**FIGURE 1** Causal diagram for the dependent mediators. An arrow represents a causal relationship between 2 variables, and a bidirected broken arc represents confounding between 2 variables. *P* is a compliance type, and it is put in parentheses in order to indicate that the compliance type is not observable even though it may be partially observable given assumptions

for unemployed workers in preventing adverse effects of unemployment such as depression. The project recruitment consisted of a short screening questionnaire (T0) to determine eligibility (self-identified job-searcher; not being on strike, expecting to be recalled, or planning to retire in the next 2 years; within 13 weeks of job loss; no preference between control and treatment groups). Eligible participants were randomly assigned to 2 groups: treatment and control. The treatment group attended 5 half-day job-search seminars while the control group received an instructional booklet on job-search strategies. The subjects that were part of randomization were mailed a pretest (T1) questionnaire. The 1801 participants that returned their questionnaire were enrolled in the study. After the seminar, the subjects in the treatment group were also mailed the instructional booklet. Follow-up surveys were mailed after 2 months (T2), 6 months (T3), and 2 years (T4). The data collected in this project included demographic variables such as age, gender, race, and marital status; as well as measures of depression, self-esteem, job-search efficacy, internal control orientation, and reemployment status. The outcome variable, depression, was measured from responses to an 11-item list based on the Hopkins Symptom Checklist.<sup>19</sup> The self-esteem variable was obtained using an 8-item list from Rosenberg's<sup>20</sup> self-esteem scale; mastery was computed as the mean score of job-search efficacy (obtained using a 6-item list), self-esteem, and internal control orientation, which was measured from a 10-item list.<sup>21</sup>

In the JOBS II study, the intervention seminar was only offered to subjects in the treatment group, ie, participants in the control group had no way of attending the seminars. Note that among the participants that were assigned to the treatment group, about 46% of them did not attend the job training seminars.

Analysis of the JOBS II study demonstrated the beneficial effects of the intervention, which included increased reemployment rates and enhanced mental health.<sup>14</sup> In particular, Price et al<sup>8</sup> examined the effect of the job training intervention on workers' mental health and found evidence that the intervention had beneficial effects on those who were identified as being at high risk of experiencing mental health setbacks such as episodes of depression. Vinokur and Schul<sup>4</sup> further attempted to identify the multiple mechanisms through which the job training intervention had an impact in reducing the level of depressive symptoms experienced by workers. Among the multiple mediators identified in Vinokur and Schul,<sup>4</sup> we focus on 2: the sense of mastery and reemployment status. We also discuss the possibility of extending the current identification results if there is a third mediator.

If 2 mediators are measured at the same time, it may be difficult to determine their causal ordering. Examples of this complication are situations where those who exhibit an improved sense of mastery are more likely to be reemployed, or those who are reemployed may feel that their self-esteem is enhanced. Although it may not be applicable to the above examples, our model requires the assumption that the causal ordering is known between the 2 mediators. In this article, we assume that reemployment status is affected by a sense of mastery. This problem may be prevented by using mediators that are measured at different times or if the causal ordering of 2 mediators is relatively clear from the theoretical framework.

### 3 | IDENTIFICATION OF THE MEDIATED PORTION OF THE ITT EFFECT

#### 3.1 | Complications of identifying the mediated portion of ITT effect with noncompliance

The purpose of randomizing the treatment is to make the treatment and control groups comparable, so that the difference in the outcome between the groups represents the causal effect of the treatment alone. However, in the presence of

noncompliance, treatment receipt status is no longer free from selection bias since complying with the assigned treatment is based on individual choice. For example, in the JOBS II study, those who are motivated to find a job are more likely to attend the training when they are assigned to the treatment. One way to handle this treatment noncompliance issue is to focus on the average effect of the assigned treatment regardless of whether one actually received the treatment or not, which is referred to the ITT effect.

The assumptions required to identify the ITT effect are discussed in Rubin<sup>22</sup> based on the Rubin Causal Model (RCM). Rubin Causal Model defines the individual treatment effect as the difference between the value that would have been observed under the treatment and the value under control conditions. The fundamental problem of RCM is that the potential value under the treatment and control condition for an individual will never be observable simultaneously.<sup>23</sup> Hence, the ITT effect is estimable only by invoking the following 2 assumptions.

1. **Stable Unit Treatment Value Assumption (SUTVA<sup>22</sup>):** This assumption will hold if (1) the treatment is clearly defined (no variation in the treatment) and (2) the treatment status of an individual should not affect the potential outcome of others. It will be violated if those who are assigned to the treatment interact with those who are not assigned to the treatment and influence their outcome.
2. **Randomization of the treatment<sup>22</sup>:** The treatment should be randomly assigned. This assumption enables groups (treatment and control) to be equivalent with respect to observable and unobservable variables.

Given these 2 assumptions, the ITT effect is identified as the difference in the average outcome value between those who are assigned to the treatment and those who are not. However, the mediated portion of the ITT effect is not identified even after additionally assuming that there is no unmeasured confounding in the mediator and outcome relationship. This is because there is a variable that confounds the mediator and outcome relationship and is affected by the treatment, which is referred to as *treatment-induced mediator and outcome confounding* (see, for example, Yamamoto,<sup>1</sup> VanderWeele and Vansteelandt,<sup>11</sup> and Robins<sup>12</sup>). In the presence of treatment noncompliance, the treatment receipt variable is affected by the treatment assigned and also confounds the mediator and the outcome relationship (Figure 1). This implies that the mediated portion of the ITT effect, which is obtained by conducting standard causal mediation analysis (see, for example, Imai et al,<sup>24</sup> Pearl,<sup>25</sup> and VanderWeele<sup>26</sup>) with the assigned treatment, is invalid. To solve this identification problem, Yamamoto<sup>1</sup> proposed an alternative approach to identify the mediated portion of the ITT effect on the basis of ACME among compliers. The consequences of using the mediated ITT effect while ignoring treatment noncompliance is shown in his simulation results.

### 3.2 | Complications of identifying mediation effects with multiple mediators

The same identification problem observed in the presence of treatment noncompliance also occurs with multiple mediators. That is, one mediator (in our case study, the sense of mastery) is affected by the treatment (job-training seminar) and confounds the relationship between the other mediator (reemployment) and the outcome. This treatment-induced mediator and outcome confounding issue prevents researchers from identifying the ACME with multiple mediators even without the treatment noncompliance issue.

Several approaches have been developed to identify the ACME with multiple mediators that overcome this treatment-induced mediator and outcome confounding issue. One of these approaches was developed by Imai and Yamamoto,<sup>10</sup> which presented a parametric estimation method combined with sensitivity analysis against the violation of no-interaction effect in the treatment-mediator relationship assumption. Under certain parametric assumptions (such as no-mediator-mediator interactions), the ACME is identified up to 2 sensitivity parameters. Another method to handle multiple mediators was proposed in Daniel et al,<sup>13</sup> which suggested multiple ways of decomposing the average treatment effect when 2 mediators exist. There are some special cases where the ACMEs are identified given sequential ignorability,\* but the rest are only partially identified up to sensitivity parameters. The approach shown in Daniel et al<sup>13</sup> appears to require fewer assumptions than that of Imai and Yamamoto,<sup>10</sup> since the ACME is identified as far as there are no-mediator-mediator interactions. In addition, Daniel et al<sup>13</sup> assume homogeneous effects while Imai and Yamamoto<sup>10</sup> allow heterogeneous effects, in which effects arbitrarily vary across individuals with respect to interactions between the treatment and the mediator. Both assumptions (that is, homogeneous effects and no-mediator-mediator interactions) are strong, and hence, researchers are required to make their own judgment on which assumptions are more reasonably met in the context of their study. In the JOBS II example, the difference in mediation effects under the treatment and controlled

\*The sequential ignorability assumption includes (1) ignorability of treatment and (2) ignorability of mediators within each treatment status.

conditions might be larger for some participants whereas it might be close to zero for others. In other words, it is highly unlikely that the effects are homogeneous, particularly with respect to interactions between the treatment and the mediator, across individuals. Therefore, we assume effect heterogeneity as in Imai and Yamamoto.<sup>10</sup>

Unlike the aforementioned approaches where mediators are considered one at a time, VanderWeele and Vansteelandt<sup>11</sup> proposed considering mediators jointly. This approach has 2 important advantages: (1) It forgoes the treatment-induced mediator and outcome confounding issue and (2) more than 2 mediators can be addressed. The first advantage is possible because one mediator is no longer the treatment-induced mediator-outcome confounding variable if both mediators are considered jointly. This is an important advantage since the ACME is nonparametrically identified given sequential ignorability even under heterogeneous effects. The second advantage addresses the possibility that more than 2 mediators could exist in practice. For example, the structural model suggested by Vinokur and Schul,<sup>4</sup> which is used as a reference for our case study, includes a third mediator. Importantly, the ACME is still identifiable despite the addition of a third mediator as far as the mediators are considered jointly. Therefore, we adopt the approach described in VanderWeele and Vansteelandt,<sup>11</sup> which considers mediators jointly, with some modifications in order to combine the approach with the instrumental variable approach.

### 3.3 | Considering mediators jointly

Consider the case where the SUTVA and randomization of the treatment are satisfied. Let  $Z_i \in \{0, 1\}$  and  $T_i \in \{0, 1\}$  represent the *treatment assigned* and *treatment received*, respectively, for individual  $i$ . If everyone complied with the treatment,  $Z_i$  would be equal to  $T_i$ . Let  $W_i$  and  $M_i$  be mediators and  $Y_i$  be the outcome for individual  $i$ . The term  $X_i$  is a vector of multiple observed pretreatment covariates. The support of the distributions of  $W_i, M_i, X_i$ , and  $Y_i$  are  $\mathcal{W}, \mathcal{M}, \mathcal{X}$ , and  $\mathcal{Y}$ , respectively. Under the notation for potential outcomes used by Little and Rubin,<sup>27</sup>  $W_i(z)$  and  $M_i(z)$  represent the potential mediators of  $W$  and  $M$  under  $Z_i = z$ , and  $Y_i(z, w, m)$  represents the potential outcome  $Y$  under  $Z_i = z, W_i = w$ , and  $M_i = m$  for individual  $i$  for  $z \in \{0, 1\}, w \in \mathcal{W}$ , and  $m \in \mathcal{M}$ . In our example, assignment to the job training intervention indicates a treatment assigned ( $Z$ ) with respect to attendance at the job training sessions, which serves as the treatment received ( $T$ ). The sense of mastery and reemployment status are the 2 mediators ( $W$  and  $M$ ). The dependent variable is an index of the level of depressive symptoms ( $Y$ ).

We define the ACME among compliers, which is referred as the Local ACME (LACME) by Yamamoto.<sup>1</sup> The rest of the effect is referred as the Local Average Natural Direct Effect (LANDE). The LACME is defined jointly through mediators  $M$  and  $W$  ( $\delta(z)$ ) and LANDE ( $\zeta(z)$ ) are expressed, respectively, as

$$\begin{aligned} \delta(z) &= E[Y_i(z, M_i(1), W_i(1)) - Y_i(z, M_i(0), W_i(0)) | P_i = c] \text{ and} \\ \zeta(z) &= E[Y_i(1, M_i(z), W_i(z)) - Y_i(0, M_i(z), W_i(z)) | P_i = c], \end{aligned} \tag{1}$$

where  $z \in \{0, 1\}$  and  $P \in \{c, a, n, d\}$  indicate a compliance type where  $c, a, n$ , and  $d$  represent compliers, always takers, never takers, and defiers, respectively. Compliers are those who abide by their assigned treatment and are represented as  $T_i(1) - T_i(0) = 1$ . Always takers are those who receive the treatment regardless of assignment and are represented as  $T_i(1) - T_i(0) = 0$ . Never takers are those who do not receive the treatment regardless of assignment and are represented as  $T_i(1) - T_i(0) = 0$ . Defiers are those who do not comply with the treatment protocol and are represented as  $T_i(1) - T_i(0) = -1$  (See, for example, Angrist et al<sup>28</sup>). In our example,  $\delta(1)$  indicates to what degree the level of depressive symptoms has changed among the compliers in response to the change in both the sense of mastery and reemployment status from the value that would have resulted under the training to the value that would have resulted under the control. Likewise,  $\zeta(1)$  indicates the average change in the level of depressive symptoms among compliers in response to the change in treatment status (that is, from being assigned to the job training to being assigned to no training), while holding the mediators at the natural values observed when assigned to the training. The local average treatment effect is realized by combining LACME  $\delta(z)$  and LANDE  $\zeta(z)$ .<sup>†</sup>

<sup>†2</sup>For  $z = 1$ ,

$$\begin{aligned} \tau &= \delta(z) + \zeta(z') \\ &= E[Y_i(1, M_i(1), W_i(1)) | P_i = c] - E[Y_i(1, M_i(0), W_i(0)) | P_i = c] \\ &\quad + E[Y_i(1, M_i(0), W_i(0)) | P_i = c] - E[Y_i(0, M_i(0), W_i(0)) | P_i = c]. \end{aligned} \tag{2}$$

The same also holds when  $z = 0$ .



Identifying the LACME and LANDE requires no unmeasured confounding in the mediator-outcome relationship among compliers if the compliance status was observable.

- 3. **No unmeasured confounding among compliers:** This assumption will hold if (1) there is no unmeasured confounding between  $Y$  and  $(M, W)$  and (2) there is no variable that is affected by the treatment and has an impact on both  $Y$  and  $(M, W)$ . As a formal expression,

$$Y_i(z', m, w) \perp \{M_i, W_i\} | Z_i = z, P_i = c, X_i \tag{3}$$

for  $z \in \{0, 1\}$  and  $z' = 1 - z$ .

In the context of the JOBS II example, this assumption implies that there should not be any unmeasured pretreatment covariate that confounds the relationship between depression and the 2 mediators—the sense of mastery and reemployment status. This is controversial since it does not hold even when the treatment is randomized. Although this assumption is required only among compliers, its implications necessitate careful attention of researchers regarding measuring all possible confounding variables and/or conducting a sensitivity analysis.

Given Assumptions 1 to 3, if the compliance status is observable, the LACME is identified as below (See Appendix A for proof).

$$\begin{aligned} \delta(z) &= \iint \mu_{mwzc} \{ \varphi_{w1c} \omega_{1c} - \varphi_{w0c} \omega_{0c} \} dm dw \text{ and} \\ \zeta(z) &= \iint \{ \mu_{mw1c} - \mu_{mw0c} \} \varphi_{wzc} \omega_{zc} dm dw, \end{aligned} \tag{4}$$

where  $\mu_{mwzc}$ ,  $\varphi_{wzc}$ , and  $\omega_{zc}$  are as defined in Table 1, which presents key parameters and sample statistics used in this article. For notational simplicity, we subsequently assume that we are already within stratum defined by covariates. The issue with Equation 4 is that the complier status is only partially observable, and hence, we need to invoke more assumptions to express the LACME and LANDE with observable quantities. One of the assumptions that permit identification is the no-defiers assumption.

- 4. **No defiers:**  $T_i(1) - T_i(0) \geq 0$  for all  $i = 1, \dots, N$ .

This assumption rules out defiers. In general, we cannot assume that there are no defiers by the observed data unless those who are assigned to the control condition do not have access to the treatment ( $T_i(0) = 0$  for every individuals). For example, in the JOBS II data, those who are assigned to the control condition are prohibited from access to the job training seminars because of the program protocol. In this setting, defiers as well as always takers, by definition, are unlikely to exist since those who are assigned to the control condition are not allowed to do the opposite of what they are assigned (that is, attending job training seminars). In our study, we follow this example, in which no defiers as well as always takers are assumed, but note that assuming no always takers are not crucial to identify the LACME and LANDE. The no always takers assumption can be relaxed when those who are assigned to the control condition are not prohibited from taking the treatment (See Appendix C for this extension).

**TABLE 1** Key parameters and sample statistics

Par	Description	Sample statistic
$\omega_z$	Conditional probability of $W = w$ given $Z = z$	$\hat{\omega}_z$
$\omega_{zn}$	Conditional probability of $W = w$ among never takers given $Z = z$	$\hat{\omega}_{zn}$
$\omega_{zc}$	Conditional probability of $W = w$ among compliers given $Z = z$	$\hat{\omega}_{zc}$
$\varphi_{wz}$	Conditional probability of $M = m$ given $W = w$ and $Z = z$	$\hat{\varphi}_{wz}$
$\varphi_{wzn}$	Conditional probability of $M = m$ among never takers given $W = w$ and $Z = z$	$\hat{\varphi}_{wzn}$
$\varphi_{wzc}$	Conditional probability of $M = m$ among compliers given $W = w$ and $Z = z$	$\hat{\varphi}_{wzc}$
$\mu_{mwz}$	Average outcome given $M = m, W = w, \text{ and } Z = z$	$\hat{\mu}_{mwz}$
$\mu_{mwzn}$	Average outcome among never takers given $M = m, W = w, \text{ and } Z = z$	$\hat{\mu}_{mwzn}$
$\mu_{mwzc}$	Average outcome among compliers given $M = m, W = w, \text{ and } Z = z$	$\hat{\mu}_{mwzc}$
$\pi_c$	Proportion of compliers	$\hat{\pi}_c$
$\pi_n$	Proportion of never takers	$\hat{\pi}_n$

Note: Par, parameter; Z, treatment assignment; W, first mediator; and M, second mediator.

By assuming no defiers and no always takers, compliance status is fully observable for those who are assigned to the treatment. Those who are both assigned to the treatment and received the treatment ( $Z_i = T_i = 1$ ) are compliers, and those who are assigned to the treatment but did not receive the treatment ( $Z_i = 1$  and  $T_i = 0$ ) are never takers. This implies that  $\mu_{mw1c}$ ,  $\varphi_{w1c}$ ,  $\omega_{1c}$ , and  $\pi_c$  are identified from those who are both assigned to the treatment and received it and  $\mu_{mw1n}$ ,  $\varphi_{w1n}$ ,  $\omega_{1n}$ , and  $\pi_n$  are identified from those who are assigned to the treatment but did not receive it. However, compliance status is not known for those who are assigned to the control condition ( $Z_i = 0$ ) since they are either compliers or never takers, which implies that  $\mu_{mw0c}$ ,  $\varphi_{w0c}$ , and  $\omega_{0c}$  are not identified.

Here, we need to assume the exclusion restriction assumption in order to identify  $\mu_{mw0c}$ ,  $\varphi_{w0c}$ , and  $\omega_{0c}$  using other observed quantities.

5. **Exclusion restriction:**  $W_i(z, t) = W_i(t)$ ,  $M_i(z, t) = M_i(t)$ , and  $Y_i(z, t, w, m) = Y_i(t, w, m)$  for all  $z$  and  $t$ . This assumption was discussed by Angrist et al<sup>28</sup> and Imbens and Rubin<sup>29</sup> in the absence of a mediator, and it is extended that the assigned treatment does not have an impact on the mediators or outcome for never takers and always takers. This implies that the treatment effect is only allowed for compliers while the treatment effect is zero for never takers and always takers.<sup>30</sup>

In the JOBS II study, the exclusion restriction assumption entails the effect of job training intervention on the sense of mastery, reemployment, and the level of depressive symptoms exclusively through attending the job training sessions. This assumption is violated, for example, if a subject was assigned to the job training but did not attend (never takers) yet became motivated by the assignment and improved job searching skills by reading a book.

When the exclusion restriction is assumed, the effect of treatment ( $Z$ ) on the outcome ( $Y$ ) via treatment receipt status ( $T$ ) can be viewed as a particular type of mediation, which is often referred as full (complete) mediation. An alternative approach is to consider the effect of treatment on the outcome via treatment receipt status as partial mediation, in which the treatment has its direct effect on the outcome without going through  $T$ . In this case, principal ignorability is assumed instead of exclusion restriction. The principal ignorability assumption was discussed within the framework of principal stratification suggested by Frangakis and Rubin<sup>31</sup> and further developed by Ding and Lu.<sup>32</sup> This assumption implies that conditional on pretreatment covariates, compliance status is correctly identified. Method selection should be based on researchers' discretion regarding which assumptions are likely to be met in the context of their study.<sup>33</sup> Given that the JOBS II data do not have an extensive set of pretreatment covariates, we prefer to assume exclusion restriction and conduct sensitivity analysis in case of violating this assumption.

Under Assumptions 1 to 5, the LACME and LANDE are nonparametrically (that is, without any functional form or any distributional assumptions) identified as

$$\begin{aligned}\delta(z) &= \iint \mu_{mwzc} \times \left\{ \frac{\varphi_{w1}\omega_{1c} - \varphi_{w0}\omega_{0c}}{\pi_c} \right\} dmdw \text{ and} \\ \zeta(z) &= \iint \left\{ \mu_{mw1c} - \frac{\mu_{mw0}\varphi_{w0}\omega_{0c} - \pi_n\mu_{mw1n}\varphi_{1n}\omega_{w1n}}{\varphi_{w0}\omega_{0c} - \pi_n\varphi_{w1n}\omega_{1n}} \right\} \times \varphi_{wzc}\omega_{zc} dmdw.\end{aligned}\quad (5)$$

The term  $\mu_{mw1c}$  and  $\varphi_{w1c}\omega_{1c}$  is directly obtained from those who are assigned to and received the treatment ( $Z_i = T_i = 1$ ), while  $\mu_{mw0c}$  and  $\varphi_{w0c}\omega_{0c}$  are as identified as

$$\mu_{mw0c} = \frac{\mu_{mw0}\varphi_{w0}\omega_{0c} - \pi_n\mu_{mw1n}\varphi_{1n}\omega_{w1n}}{\varphi_{w0}\omega_{0c} - \pi_n\varphi_{w1n}\omega_{1n}} \text{ and } \varphi_{w0c}\omega_{0c} = \frac{\varphi_{w0}\omega_{0c} - \pi_n\varphi_{w1n}\omega_{1n}}{\pi_c}.$$

See Appendix A for proof. Based on these LACME and LANDE, the mediated and unmediated ITT effects are identified by multiplying the proportion of compliers, ie,  $\lambda(z) = \pi_c\delta(z)$  and  $\kappa(z) = \pi_c\zeta(z)$ , respectively.

### 3.4 | Considering mediators one at a time

In practice, it may be of primary interest to distinguish the portion of mediation effect that is attributed to each mediator from the mediation effect jointly through all mediators. To address this issue, VanderWeele and Vansteelandt<sup>11</sup> suggested a sequential approach that provides information on how much of the combined mediation effect is attributed to each mediator. According to the sequential approach, we first estimate the mediation effect attributed to  $W$  alone. Once  $\delta^W(z)$  is obtained, the proportion of  $\delta^W(z)$  can be calculated out of the LACME jointly through  $M$  and  $W$  ( $\frac{\delta^W(z)}{\delta(z)}$ ); the rest of the



effect is attributed to the LACME through  $M(1 - \frac{\delta^W(z)}{\delta(z)})$ . This procedure will provide additional information on how much of the combined LACME is attributed to each mediator.

A natural question that may arise with respect to the sequential approach is how to define the mediation effect attributed to  $W$ . As demonstrated in Daniel et al,<sup>13</sup> there are multiple ways to define a mediation effect that passes through one mediator when multiple are present. In this section, we only focus on the following definition.

$$\delta^W(z) = E[Y_i(z, M_i(z, W(1)), W_i(1)) - Y_i(z, M_i(z, W(0)), W_i(0)) | P_i = c] \tag{6}$$

for  $z \in \{0, 1\}$ . The LACME through  $W(\delta^W(z))$  states the expected change in  $Y$  among compliers in response to the change in  $W$  from the value that would have resulted under the treatment to the value under the control condition when assigned to  $Z_i = z$ . In our example,  $\delta^W(1)$  indicates how much change is expected in the level of depressive symptoms among compliers in response to the change in the sense of mastery when they are assigned to the job training. Note that  $\delta^W(z)$  includes all effects that are mediated through  $W$ , which go through (1)  $Z(= T) \rightarrow W \rightarrow M \rightarrow Y$  and (2)  $Z(= T) \rightarrow W \rightarrow Y$ .

We only focus on this definition ( $\delta^W(z)$ ) because it does not require additional assumptions to permit identification even under heterogeneous effects. Identifying  $\delta^W(z)$  requires the following assumptions: (1) There are no unobserved pretreatment confounding variables between  $W$  and  $Y$  among compliers and (2) there is no variable that is affected by the treatment and affects both  $W$  and  $Y$ . Assumption 3, in fact, implies these assumptions.

$$Y_i(z', M(z', w), w) \perp W_i | Z_i = z, P_i = c, X_i \tag{7}$$

for  $z \in \{0, 1\}$ ,  $z' = 1 - z$ , and  $w \in W$  (see Appendix B for proof).

Under Assumption (7) combined with Assumptions 1, 2, 4, and 5, the LACME through  $W$  is identified as

$$\delta^W(z) = \int \mu_{wzc} \left\{ \frac{\omega_1 - \omega_0}{\pi_c} \right\} dw, \tag{8}$$

where  $\mu_{w1c}$  is identified among those who are assigned to the treatment and received the treatment ( $Z_i = T_i = 1$ ), and  $\mu_{w0c}$  is identified as  $\mu_{w0c} = \frac{\mu_{w0}\omega_0 - \pi_n\mu_{w1n}\omega_{1n}}{\omega_0 - \pi_n\omega_{1n}}$ . Again, based on LACME attributed to  $W$ , the ITT effect mediated through  $W$  is identified as  $\lambda^W(z) = \pi_c\delta^W(z)$ .

### 4 | ESTIMATION

The LACME and LANDE shown in Equation 5 consist of the population expectation of  $Y$  and the population conditional probability of  $M$  and  $W$ . Given that we only have sample data, the unbiased estimators of the LACME and LANDE are, respectively, as

$$\begin{aligned} \hat{\delta}(z) &= \iint \hat{\mu}_{mwzc} \left\{ \frac{\hat{\phi}_{w1}\hat{\omega}_1 - \hat{\phi}_{w0}\hat{\omega}_0}{\hat{\pi}_c} \right\} dmdw \text{ and} \\ \hat{\zeta}(z) &= \iint \left\{ \hat{\mu}_{mw1c} - \frac{\hat{\mu}_{mw0}\hat{\phi}_{w0}\hat{\omega}_0 - \hat{\pi}_n\hat{\mu}_{mw1n}\hat{\phi}_{w1n}\hat{\omega}_{1n}}{\hat{\phi}_{w0}\hat{\omega}_0 - \hat{\pi}_n\hat{\phi}_{w1n}\hat{\omega}_{1n}} \right\} \times \hat{\phi}_{wzc}\hat{\omega}_{zc} dmdw. \end{aligned} \tag{9}$$

These estimators are consistent for the LACME and LANDE under both homogeneous and heterogeneous effects. The estimators for the mediated and unmediated ITT effects ( $\hat{\lambda}(z) = \hat{\pi}_c\hat{\delta}(z)$  and  $\hat{\kappa}(z) = \hat{\pi}_c\hat{\zeta}(z)$ ) are also consistent since they are obtained from the LACME and LANDE estimators.

Likewise, the unbiased moment-based estimator of the LACME through  $W$  only is

$$\hat{\delta}^W(z) = \int \hat{\mu}_{wzc} \left\{ \frac{\hat{\omega}_1 - \hat{\omega}_0}{\hat{\pi}_c} \right\} dw. \tag{10}$$

This estimator is consistent for the LACME through  $W$  under both heterogeneous and homogeneous effects. The estimate for the mediated ITT effect through  $W(\hat{\lambda}(t) = \hat{\pi}_c\hat{\delta}(t))$  is also consistent.

The method of moments estimators shown in Equations 9 and 10 is purely nonparametric, which provide an intuitive understanding of how LACME and LANDE are estimated. However, the direct employment of these nonparametric estimators may not be always feasible if the sample size is small, as there will not be enough observations for each combination of  $M = m$ ,  $W = w$ , and  $Z = z$ . Therefore, we fit parametric regressions to obtain the sample expectation of  $Y$  and

sample conditional probabilities of  $M$ ,  $W$ , and  $T$ . As a simplified example, suppose that 2 mediators are binary and the outcome is continuous. By fitting logistic and normal regressions, we have

$$\begin{aligned} E[T|Z] &= \alpha_0 + \alpha_1 Z, \\ P(W = 1|Z, T) &= \text{logit}^{-1}(\beta_0 + \beta_1 Z + \beta_2 T), \\ P(M = 1|Z, T, W) &= \text{logit}^{-1}(\gamma_0 + \gamma_1 Z + \gamma_2 T + \gamma_3 W + \gamma_4 ZW + \gamma_5 TW), \text{ and} \\ E[Y|Z, T, W, M] &= \kappa_0 + \kappa_1 Z + \kappa_2 T + \kappa_3 W + \kappa_4 M + \kappa_5 ZW + \kappa_6 ZM + \kappa_7 TW + \kappa_8 TM, \end{aligned} \quad (11)$$

where  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\kappa$  are regression coefficients. From the first regression model, we obtain the probability of compliers and never takers, respectively, by  $\hat{\pi}_c = \hat{\alpha}_1$  and  $\hat{\pi}_n = 1 - \hat{\alpha}_1$ . From the second regression model, we obtain the conditional probability of  $W = w$  given  $Z = 1$  among compliers and never takers, respectively, by  $\hat{\omega}_{1c} = \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2)$  and  $\hat{\omega}_{1n} = \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1)$ . From the third regression model, we obtain the probability of  $M = m$  given  $W = w$  and  $Z = 1$  among compliers and never takers, respectively, by  $\hat{\phi}_{w1c} = \text{logit}^{-1}(\hat{\gamma}_0 + \hat{\gamma}_1 + \hat{\gamma}_2 + w(\hat{\gamma}_3 + \hat{\gamma}_4 + \hat{\gamma}_5))$  and  $\hat{\phi}_{w1n} = \text{logit}^{-1}(\hat{\gamma}_0 + \hat{\gamma}_1 + w(\hat{\gamma}_3 + \hat{\gamma}_4))$ . Lastly, the expected values of  $Y$  given  $M = m$ ,  $W = w$ , and  $Z = 1$  among compliers and never takers are  $\hat{\mu}_{mw1c} = \hat{\kappa}_0 + \hat{\kappa}_1 + \hat{\kappa}_2 + w(\hat{\kappa}_3 + \hat{\kappa}_5 + \hat{\kappa}_7) + m(\hat{\kappa}_4 + \hat{\kappa}_6 + \hat{\kappa}_8)$  and  $\hat{\mu}_{mw1n} = \hat{\kappa}_0 + \hat{\kappa}_1 + w(\hat{\kappa}_3 + \hat{\kappa}_5) + m(\hat{\kappa}_4 + \hat{\kappa}_6)$ , respectively. Once sample expectations and sample conditional probabilities are obtained, empirical integration was used with respect to different values of mediators after plugging them into Equations 9 and 10. The 95% confidence intervals are obtained from the nonparametric bootstrap. The same estimation procedure was used in Yamamoto.<sup>1</sup> All analyses were done in R<sup>34</sup> and the R code will be available upon request.

Another approach is to fit nonparametric regressions to obtain each sample expectation and conditional probability. In this case, a kernel regression or a local linear regression could be performed. Compared to a kernel regression, a local linear regression would provide more consistent and accurate estimates (See Fan and Gijbels<sup>35</sup> for more details on these approaches).

## 5 | A SIMULATION STUDY

Since the estimators of the mediated and unmediated portion of ITT effects are consistent under both homogeneous and heterogeneous effects, it would be of interest to know how large the sample size and how high the compliance rate should be to (1) obtain unbiased estimates of the mediated and unmediated ITT effect and (2) achieve sufficient statistical power. A Monte Carlo simulation was used to test the performance and statistical power of our proposed approach in estimating (1) the mediated ITT effects jointly through 2 mediators ( $\lambda(t)$ ) and (2) the mediated ITT effects through mediator  $W(\lambda^W(t))$  in various settings. We used the same data generation technique as in Yamamoto,<sup>1</sup> given its similarity to identification of the mediated portion of the ITT effect based on LACME.

The simulation results presented in this section are based on 2000 replications. To study the performance of our method under various potential data conditions, we used the following design conditions: (1) different sample sizes (200, 1000, and 2000), (2) varying compliance rates (0.3, 0.5, and 0.8), and (3) varying degrees of the effect heterogeneity (0, 1 standard deviation [SD], and 2 SD). As the JOBS II data have a sample size of 1802 and a compliance rate of 0.46 among those who are assigned to the treatment, this simulation study is beneficial in judging our method's performance regarding our motivating data.

As mentioned above, in addition to the sample size and compliance rate, we considered the degree of the effect heterogeneity as one of the key conditions. Although the proposed definitions of the mediated and unmediated portion of ITT effects are identified under both homogeneous and heterogeneous effects, the statistical power will vary depending on how heterogeneous the effects are across individuals. For example, even if the mean difference is constant between complier and never taker distributions, the statistical power will vary by the variance of each distribution. Therefore, we allow the SD of the effects to vary from 0 to 2 SD, which represent homogeneous to heterogeneous effects.

The treatment variable ( $Z$ ) is randomly assigned to 0 or 1 in order to mimic a complete randomization. Two hypothetical populations are generated: compliers and never takers. The treatment received ( $T$ ) is determined by the treatment assignment ( $Z$ ) and the compliance type. The proportion of compliers is manipulated according to the compliance rate. Never takers are generated via the proportion  $(1 - \pi_c)$ , where  $\pi_c$  is the compliance rate. Two dependent mediators ( $W$  and  $M$ ) and the outcome ( $Y$ ) are generated based on the following models:

$$\begin{aligned} Pr(W_i(t) = 1) &= \text{logit}^{-1}(\alpha_{1i} + \beta_{1i} T_i), \\ Pr(M_i(t) = 1) &= \text{logit}^{-1}(\alpha_{2i} + \beta_{2i} T_i + \beta_{3i} W_i + \gamma_{1i} T_i W_i), \text{ and} \\ Y_i(t, m, w) &= \alpha_{3i} + \beta_{4i} T_i + \beta_{5i} M_i + \beta_{6i} W_{2i} + \gamma_{2i} T_i M_i + \gamma_{3i} T_i W_i + \gamma_{4i} W_i M_i + e_i, \end{aligned} \quad (12)$$

where  $\alpha_i, \beta_i,$  and  $\gamma_i$  are heterogeneous individual effects that vary by a compliance type and at random. The random variation of the heterogeneous effects ranges from 0 to 2 SD. The error term  $e_i$  follows a standard normal distribution. (See Appendix D for parameters for compliers and never takers.) This data generation ensures that all the assumptions described in Section 3 are satisfied. The true value of LACME jointly through mediators when assigned to the treatment is 1 regardless of the conditions, and the true value of the mediated portion of the ITT effect changes accordingly with the compliance rate. The true value of LACME through  $W$  is 0.2, and the true value of the mediated portion of the ITT effect through  $W$  also changes accordingly with the compliance rate. For an intuitive understanding of this data-generating model, suppose that there is no interaction effect (ie, all  $\gamma$ s are zero). Then the LACME is expressed as  $\delta = E[(\beta_{2i} + \beta_{1i}\beta_{3i})\beta_{5i} + \beta_{1i}\beta_{6i}|P_i = c]$  and LANDE is  $\zeta = E[\beta_{4i}|P_i = c]$ . The estimation in the simulated data is performed as explained in Section 4 using parametric regressions.

The simulation results are summarized in terms of bias, root mean square error (RMSE), coverage rate, and statistical power. Bias is defined as the difference between the average of the LACME estimates over replications and the true LACME. The RMSE is defined as the average difference between the LACME estimates and the true LACME over 2000 replications. The coverage rate is the proportion of replications out of 2000 where the true value falls within the nominal level of the 95% confidence interval. Statistical power is obtained by the proportion of replications out of 2000 where the 95% confidence interval does not cover zero.

Table 2 shows the performance of the proposed method in estimating the mediated ITT effects. The columns represent the following: sample size, bias, RMSE, and 95% coverage rate for each of the 80%, 50 %, and 30% compliance rates. The top half of the table presents the estimates for  $\hat{\lambda}(1)$ , and the bottom half of the table presents the estimates for  $\hat{\lambda}^W(1)$ . The results of the mediated ITT effects for the not-treated group are similar to those for the mediated ITT effects for the treated group.

The proposed estimator for the mediated ITT effects jointly through 2 mediators recovers the true value with a moderate sample size and a minimum 50% compliance rate. For instance, if the compliance rate is 50% and the sample size is 1000 or greater, the bias in  $\hat{\lambda}(1)$  is less than -0.04, and the 95% confidence interval coverage rate reaches the nominal level. The result is also similar with the proposed estimator for the mediated ITT effect via mediator  $W$ . When the compliance rate is 50% and the sample size is 1000, the bias in  $\hat{\lambda}^W(1)$  is less than 0.02 and the 95% coverage rate reaches the nominal level.

Although they display similar patterns, the bias and variance are generally smaller for  $\hat{\lambda}^W(1)$  than  $\hat{\lambda}(1)$ . Additionally,  $\hat{\lambda}^W(1)$  reaches nominal coverage more quickly than  $\hat{\lambda}(1)$ . This phenomenon can be explained by the fact that estimating  $\lambda^W(1)$  requires only 1 mediator instead of 2.

**TABLE 2** The performance of the mediated intention-to-treated effect estimates

	N	80% Compliance rate			50% Compliance rate			30% Compliance rate		
		Bias	RMSE	95 Cov	Bias	RMSE	95 Cov	Bias	RMSE	95 Cov
$\hat{\lambda}(1)$	200	Inf	Inf	95.71	Inf	Inf	99.72	Inf	Inf	100.00
	1000	-0.01	0.12	95.25	-0.03	0.14	95.68	Inf	Inf	97.50
	2000	0.00	0.08	95.73	-0.01	0.09	95.25	-0.04	0.12	94.20
$\hat{\lambda}^W(1)$	200	-0.00	0.18	94.83	Inf	Inf	96.82	Inf	Inf	99.85
	1000	-0.00	0.08	94.75	-0.01	0.08	94.90	-0.02	0.09	93.80
	2000	-0.00	0.05	94.60	-0.00	0.05	95.40	-0.01	0.05	94.15

Note: (1) N, Sample size; and 95 Cov, 95% confidence interval coverage rate; (2) the results for  $\hat{\lambda}(0)$  and  $\hat{\lambda}^W(0)$  are not shown in the table, but the pattern is similar to what is shown in Table 2; and (3) the result is based on when the variance of the effect heterogeneity is 1.

**TABLE 3** Statistical power of the mediated intention-to-treated effects estimates

	N	80% Compliance rate			50% Compliance rate			30% Compliance rate		
		$\sigma=0$	$\sigma=1$	$\sigma=2$	$\sigma=0$	$\sigma=1$	$\sigma=2$	$\sigma=0$	$\sigma=1$	$\sigma=2$
$\hat{\lambda}(1)$	200	11.8	19.0	7.1	0.2	0.0	0.0	0.0	0.0	0.0
	1000	100.0	99.9	95.7	77.4	63.9	26.7	7.0	2.5	0.2
	2000	100.0	100.0	100.0	99.2	96.2	76.1	43.8	25.7	6.6
$\hat{\lambda}^W(1)$	200	44.7	24.1	5.1	2.7	0.9	0.3	0.2	0.0	0.0
	1000	100.0	97.5	69.0	81.1	51.1	14.4	12.7	4.8	0.7
	2000	100.0	100.0	95.0	99.1	93.0	51.3	50.2	25.4	4.7

Note: (1) N, Sample size; and  $\sigma$ , variance of the effects heterogeneity; and (2) the results for  $\hat{\lambda}(0)$  and  $\hat{\lambda}^W(0)$  are not shown in the table, but the pattern is similar to what is shown in Table 3.

Table 3 demonstrates the statistical power of the proposed estimators for the mediated ITT effects. The top half of the table presents statistical power for  $\lambda(1)$ , and the bottom half of the table presents the estimates for  $\lambda^W(1)$ . The columns represent the following: 0, 1, and 2 SD of the effect heterogeneity for the 80%, 50%, and 30% compliance rates.

The statistical power of estimating the mediated ITT effects is negatively affected by the variance of the effect heterogeneity. To achieve sufficient power, a sample size of 1000 or greater is required with at least a 50% compliance rate when effects are homogeneous ( $\sigma = 0$ ). When effects are largely heterogeneous ( $\sigma = 2$ ), a sample size of 1000 is insufficient to achieve adequate power with a 50% treatment noncompliance rate. The pattern is similar with the mediated ITT effect that goes through  $W$ . When effects are largely heterogeneous ( $\sigma = 2$ ), a greater sample size is required to achieve a similar level of statistical power in estimating  $\lambda^W(1)$  than  $\lambda(1)$ . This is due to the fact that true effect of  $\lambda^W(1)$  is closer to zero than  $\lambda(1)$ .

## 6 | APPLICATION TO JOBS II STUDY

In this section, we apply our method to the aforementioned JOBS II study. In this analysis, we focus on estimating the mediated portion of the job training intervention effects on reducing an individual's level of depressive symptoms through an improved sense of mastery and reemployment. Assignment to the job training intervention ( $Z$ ) is a binary variable that is 1 when a participant is assigned to the intervention or 0 otherwise. Attendance at the job training seminar ( $T$ ) is a binary variable that is 1 when the participant attended or 0 otherwise.

For the mediators and outcome, we used the same constructs as those measured by Vinokur and Schul.<sup>4</sup> All of the constructs were assessed with multiple items measured at T2 or T3 and most had a Cronbach alpha value in the range .7 to .9. The sense of mastery measure utilizes the Likert scale (ie, "Not at all," "A little," "Somewhat," and "Pretty much"). Reemployment, another mediator ( $M$ ), is a binary variable that is 1 if a participant is reemployed or 0 otherwise. The dependent variable is the index of the level of depressive symptoms ( $Y$ ), which is measured at T3 using a subscale of 11 items. The pretreatment covariates ( $X$ ), which include gender, age, marital status, race, education, family income, perceived economic hardship, and level of depressive symptoms, were measured at T0 or T1.<sup>‡</sup>

In the regression models that are fitted for treatment, mediator  $M$ , and outcome are same as Equations 11. Given that the type of mediator  $W$  is ordinary, the following regression is fitted as

$$P(W = j | j < J, Z, T) = \Phi(\Gamma_j - e_1Z - e_2T) - \Phi(\Gamma_{j-1} - e_1Z - e_2T),$$

$$P(W = J | Z, T) = 1 - \Phi(\Gamma_J - e_1Z - e_2T),$$

where  $e_1$  and  $e_2$  are regression coefficients,  $\Phi$  denotes the distribution function of a standard normal random variable, and  $\Gamma$  is the categorical threshold for the ordinal category of  $W_i = j$  for  $j = 1, \dots, J$ . An ordered probit regression model is fitted for estimating the conditional probability of the sense of mastery measure, as the variable has an ordinal scale with 4 categories.

Based on the assumptions presented in Section 3, the results shown in Table 4 can be given a causal interpretation. The upper half of the table presents the local effect estimates for the compliers, and the lower half presents the ITT effect estimates. The patterns of the mediated and unmediated ITT effects do not differ from those of the LACME and LANDE. The only difference is that the ITT effect estimates are smaller than the analogous local effect estimates. The 95% confidence intervals are directly obtained from the nonparametric bootstrap.

The mediated ITT effect estimated jointly through an enhanced sense of mastery and reemployment for those who received the training ( $\hat{\lambda}(1)$ ) is -0.038, and the 95% confidence interval indicates that  $\hat{\lambda}(1)$  is significantly different from zero. The mediated ITT effect estimate for those who did not receive the training ( $\hat{\lambda}(0)$ ) is also -0.043 but is not significant at the 95% confidence level. The proportion of the estimated LACME through an enhanced sense of mastery is  $-0.029 / -0.038 = 0.76$ . That indicates that about 76% of the mediation effect can be attributable to just an enhanced sense of mastery. Given these assumptions, we conclude that the mediating mechanisms through which the job training intervention has its effects on the level of depressive symptoms include an enhanced sense of mastery and reemployment. The average proportion of mediated effects through these mediators combined is  $-0.038 / -0.063 = 0.60$  for the treatment group.

<sup>‡</sup>The average missing rate of these construct measures is about 20%. We performed the predictive mean matching imputation under the assumption that (1) missing values are random given the pretreatment covariates, treatment, and attendance status for those who are treated and (2) missing values are random given the pretreatment covariates and treatment for those who are not treated. This approach involves a strong assumption where missing values do not depend on the compliance status. The consequences of this strong assumption are discussed in Frangakis and Rubin.<sup>36</sup>

**TABLE 4** The estimates of the LACME and LANDE

Complier effects				ITT effects			
Parameter	Est	95% CI	90% CI	Parameter	Est	95% CI	90% CI
$\delta(1)$	-0.073	[-0.122, -0.021]	[-0.114, -0.030]	$\lambda(1)$	-0.038	[-0.055, -0.002]	[-0.061, -0.016]
$\delta(0)$	-0.082	[-0.215, 0.126]	[-0.182, 0.032]	$\lambda(0)$	-0.043	[-0.076, 0.000]	[-0.097, 0.017]
$\delta^W(1)$	-0.056	[-0.104, -0.006]	[-0.098, -0.013]	$\lambda^W(1)$	-0.029	[-0.055, -0.003]	[-0.052, -0.007]
$\delta^W(0)$	-0.067	[-0.143, 0.001]	[-0.129, -0.014]	$\lambda^W(0)$	-0.037	[-0.076, 0.001]	[-0.068, -0.007]
$\zeta(1)$	-0.038	[-0.299, 0.130]	[-0.206, 0.097]	$\kappa(1)$	-0.020	[-0.159, 0.069]	[-0.109, 0.052]
$\zeta(0)$	-0.048	[-0.170, 0.065]	[-0.160, 0.060]	$\kappa(0)$	-0.025	[-0.090, 0.035]	[-0.076, 0.026]
local $\tau$	-0.120	[-0.229, -0.014]	[-0.241, -0.001]	ITT $\tau$	-0.063	[-0.131, 0.001]	[-0.122, -0.007]

Note: Est, estimates; CI, confidence interval; LACME, Local Average Causal Mediation Effect; LANDE, Local Average Natural Direct Effect.

### 6.1 | Sensitivity analysis

These analytical results can be given a causal interpretation only under Assumptions 1 to 5. Previously, we discussed that 2 assumptions are controversial: (1) the exclusion restriction and (2) no unmeasured confounding in the mediator-outcome relationship among compliers. First, we examine whether our conclusion will change when the exclusion restriction assumption is violated while maintaining the other assumptions. The exclusion restriction will be violated if there are some participants who were assigned to the job training but did not participate in the training (never takers) yet became motivated by the assigned treatment and improved their job-searching skills by reading a book. For instance, suppose that the improved job-searching skills, in return, enhanced a sense of mastery, reemployment rate, and/or decreased the level of depression. In this case, because of the violation of the exclusion restriction, the average level of sense of mastery and reemployment rate for those who are assigned to the job training will be higher than those who are not assigned to the job training among never takers. Conversely, the violation of the exclusion restriction assumption leads to a lower average level of depression for those who are assigned to the job training than those who are not assigned to the job training among never takers. Suppose that

$$\frac{\omega_{0n}}{\omega_{1n}} = \rho_1, \frac{\varphi_{w0n}}{\varphi_{w1n}} = \rho_2, \text{ and } \frac{\mu_{mw0n}}{\mu_{mw1n}} = \rho_3$$

for every value of  $M = m$  and  $W = w$ , and where  $\rho_s$  are sensitivity parameters. If  $\rho_s$  equal 1, the exclusion restriction assumption is satisfied. We use the reference value of  $\rho_1 = 0.89$ ,  $\rho_2 = 0.79$ , and  $\rho_3 = 1.09$  based on the fact that  $\mu_{mw1n} - \mu_{mw0n}$  is unlikely to exceed the average effect for compliers (CACE), which is -0.11 with respect to the outcome. Likewise, we assume that  $\varphi_{w1n} - \varphi_{w0n}$  and  $\omega_{1n} - \omega_{0n}$  are unlikely to exceed CACE with respect to mediators, which are 0.25 and 0.10, respectively. Table 5 presents the sample estimates of  $\mu_{mw1n}$ ,  $\varphi_{w1n}$ , and  $\omega_{1n}$ . Based on Table 5, the maximum value of  $\rho_3$  is 1.09 because  $|\mu_{mw1n} - \rho_3 \mu_{mw0n}| < 0.11$ . Likewise, minimum values of  $\rho_1$  and  $\rho_2$  are 0.89 and 0.79, respectively. For fixed values of  $\rho_s$ , the LACMEs for those who are treated, for instance, are identified as

$$\delta(1) = \iiint \mu_{mw1c} \times \left\{ \varphi_{w1c} \omega_{1c} - \frac{\varphi_{w1} \omega_1 - \rho_1 \rho_2 \varphi_{w1n} \omega_{1n}}{\pi_c} \right\} dmdw \text{ and}$$

$$\delta^W(1) = \iint \mu_{w1c} \times \left\{ \omega_{1c} - \frac{\omega_1 - \rho_1 \omega_{1n}}{\pi_c} \right\} dw.$$

Interestingly, the LACME through both an enhanced sense of mastery and reemployment for those who are treated is even larger (-0.75) and the 95% interval does not contain zero (-0.86, -0.68). This pattern is consistent with the LACME through a sense of mastery in that it is -0.27 and the 95% interval does not contain zero (-0.34, -0.23). This result implies that our conclusion is robust (and even stronger) for this particular scenario of violating the exclusion restriction assumption.

We next examine whether our conclusions will change when there are unmeasured pretreatment covariates between the mediators and outcome. In other words, we consider the violation of the no-unmeasured confounding among compliers assumption while assuming that the rest of the assumptions are satisfied. A bias can originate from unmeasured covariates that confound the  $W - Y$  relationship,  $M - Y$  relationship, or both. We assume that there might exist an unmeasured composite  $U$  that confounds both the  $W - Y$  and  $M - Y$  relationships. For simplicity, suppose that  $U$  is binary where  $U_i \in \{0, 1\}$ .

According to VanderWeele,<sup>37</sup> the impact of bias originating from confounding in the  $M - Y$  relationship depends on the following 2 paths in which the associations are obtained among compliers: (1) the path between the assigned treatment



**TABLE 5** Sensitivity parameters regarding exclusion restriction

Values of $M$	Values of $W$	$\hat{\mu}_{mw1n}$	$\hat{\phi}_{w1n}$	$\hat{\omega}_{1n}$
0	1	2.30	0.44	2.37
0	2	1.95	0.39	
0	3	1.49	0.49	
0	4	1.49	0.46	
1	1	2.11		
1	2	1.80		
1	3	1.52		
1	4	1.29		
CACE		-0.11	0.25	0.10
max or min values of $\rho$		1.09	0.75	0.89

Note:  $\hat{\mu}_{mw1n}$ , Sample expectation of  $Y$  given  $M = m$ ,  $W = w$ , and  $Z = 1$  among never takers;  $\hat{\phi}_{w1n}$ , sample conditional probability of  $M = 1$  given  $W = w$  and  $Z = 1$  among never takers; and  $\hat{\omega}_{1n}$ , sample conditional probability of  $W = 1$  given  $Z = 1$  among never takers.

and unobserved covariates via mediator  $M$  (that is,  $Z \rightarrow M \rightarrow U$  or  $Z \rightarrow W \rightarrow M \rightarrow U$  as represented in bold in Figure 2A) and (2) the path between unmeasured covariates and the outcome (that is,  $U \rightarrow Y$  as represented in bold in Figure 2C). The first path between  $Z$  and  $U$  can be denoted as  $\beta_m$  if the association between  $Z$  and  $U$  among compliers is constant across the strata of  $W = w$ ,  $M = m$ , and  $X = x$ . The second path between  $U$  and  $Y$  can be denoted as  $\alpha$  if the association between  $U$  and  $Y$  among compliers is constant across the strata of  $W = w$ ,  $M = m$ , and  $X = x$ . By taking into account the unmeasured composite  $U$  that confounds the relationship between  $M$  and  $Y$ , the biased estimate is expressed as  $\hat{\delta}(t) = \delta(t) + \alpha\beta_m$ . Additionally, by considering the bias due to the confounding relationship between  $W$  and  $Y$  among compliers, the biased estimate is expressed as

$$\hat{\delta}(t) = \delta(t) + \alpha(\beta_m + \beta_w),$$

where  $\beta_w$  is the path between  $U$  and  $Z$  via mediator  $W$  only (that is,  $Z \rightarrow W \rightarrow U$  as represented in bold in Figure 2B).<sup>§</sup>

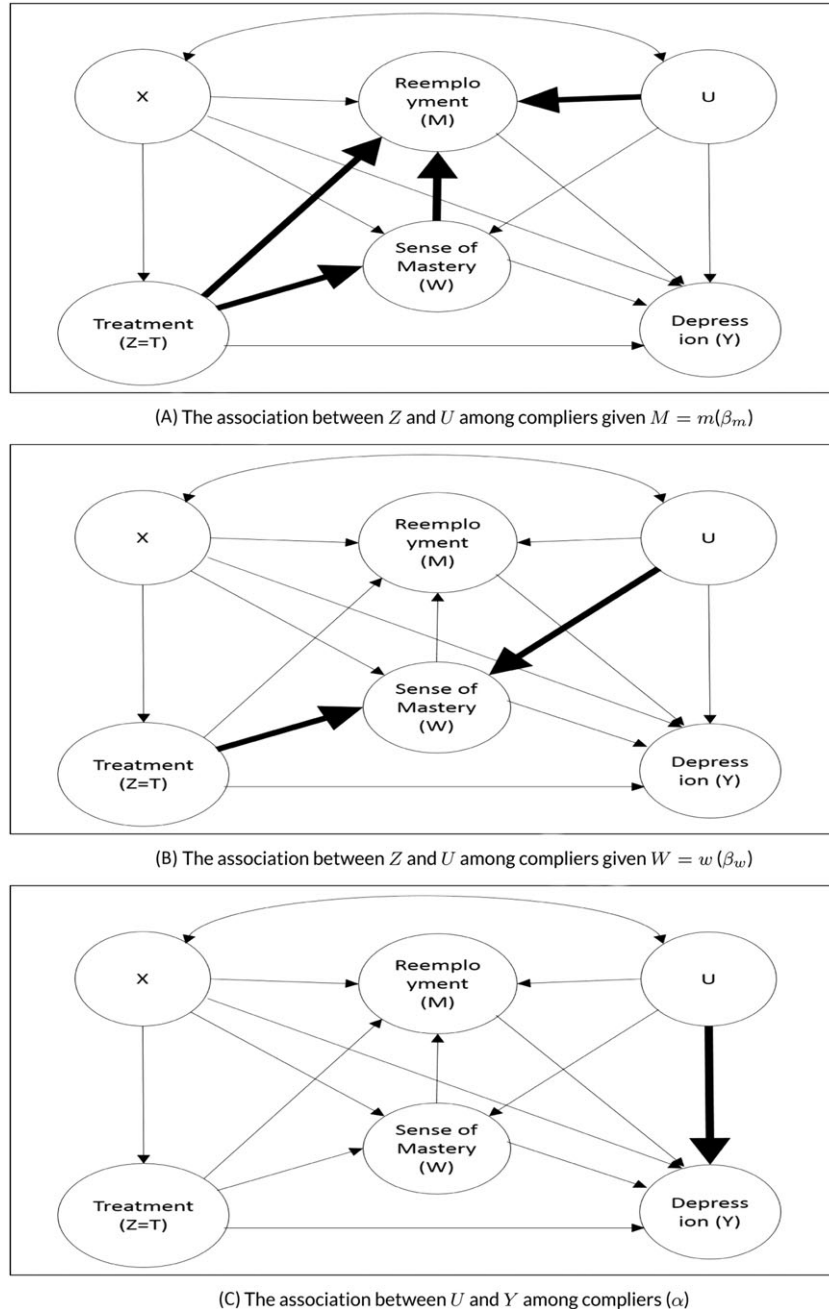
If the impact of hidden bias due to  $U$ , which is represented as  $\alpha(\beta_m + \beta_w)$ , is the same as our estimate, the true value ( $\delta(t)$ ) might be in fact zero. For example, suppose that an unmeasured health condition confounds the relationships between the 2 mediators and the level of depressive symptoms. Suppose further that an individual in good health had on average 0.3 SD lower depression levels ( $\alpha = -0.3$ ). The reference value of 0.3 SD is chosen based on the covariate that demonstrated the strongest effect on the level of depressive symptoms—that is, the previously measured lagged outcome variable—after controlling existing covariates and mediators. When conditioned on covariates and either mediator, if individuals who received job training had a 0.23 higher probability of having better health ( $\beta_m + \beta_w = 0.23$ ), the true value is zero ( $-0.07 - \{-0.3 \times 0.23\} = 0$ , where  $\hat{\delta}(1) = -0.07$ ). This amount of confounding can be considered moderate or high in social sciences.

Although the estimate for  $\delta(1)$  was considered relatively insensitive to an unmeasured composite, the 95% confidence interval may contain zero after incorporating an unmeasured covariate. Suppose that the same reference value for  $\alpha$  is used as before (ie,  $\alpha = -0.3$ ). When the upper bound of the estimate was used (the upper bound of  $\hat{\delta}(1)$  is  $-0.01$ ), the probability of 0.03 can make the true value zero ( $-0.01 - \{-0.3 \times 0.03\} = 0$ ). Thus, the LACME via 2 mediators can be altered if there is an unmeasured covariate that has an impact on the outcome as strong as the previously measured lagged outcome. The mediated portion of the ITT effect, which is obtained by multiplying the inverse of the proportion of compliers, will be in turn altered.

Our results show that the mediating mechanisms include an enhanced sense of mastery and reemployment between the causal relationship between the job training seminar and depression levels if Assumptions 1 to 5 are satisfied. Our

<sup>§</sup>The extension of VanderWeele's<sup>37</sup> bias formula to multiple mediators can be easily verified by using DAG rule, *Conditional Independence in Colliders*.<sup>38</sup> The path denoted as  $\beta_m$ , expressed as  $P(U_i = 1|Z_i = 1, M_i = m, X_i = x, P_i = c) - P(U_i = 1|Z_i = 0, M_i = m, X_i = x, P_i = c)$ , is identified by conditioning on collider  $M$ . A collider is the variable influenced by 2 parental variables and conditioning on a collider makes the 2 parental variables dependent. With the same logic, the path denoted by  $\beta_w$ , expressed as  $P(U_i = 1|T_i = 1, W_i = w, X_i = x, P_i = c) - P(U_i = 1|T_i = 0, W_i = w, X_i = x, P_i = c)$ , is identified by conditioning on collider  $W$ . Lastly, the path denoted by  $\alpha$ , expressed as  $E(Y_i|U_i = 1, W_i = w, M_i = m, X_i = x, P_i = c) - E(Y_i|U_i = 0, W_i = w, M_i = m, X_i = x, P_i = c)$ , is identified by blocking every back-door path.





**FIGURE 2** A, The association between  $Z$  and  $U$  among compliers given  $M = m(\beta_m)$ . B, The association between  $Z$  and  $U$  among compliers given  $W = w(\beta_w)$ . C, The association between  $U$  and  $Y$  among compliers ( $\alpha$ )

finding is robust to a potential violation to the exclusion restriction assumption. This is particularly true in the case of participants who were assigned to the job training but did not attend yet improved their job searching skills by other means such as reading a book. However, our conclusion might be changed if there is an unmeasured pretreatment covariate that is as strong as the previously measured lagged outcome.

## 7 | SUMMARY AND CONCLUSIONS

In the presence of treatment noncompliance, the mediated portion of the ITT effect is not identified nonparametrically, unless one is willing to make a strong assumption such as no treatment-mediator interactions. Yamamoto<sup>1</sup> suggested identifying the mediated portion of the ITT effects on the basis of the ACME among compliers. This dates back to

Frangakis and Rubin<sup>36</sup> and Jo et al<sup>30</sup> in the context of identifying the ITT effects with subsequent missing values and in cluster randomized trials, respectively. Building on this, the approach presented here contributes to the literature by solving the identification problem in the presence of 2 challenges simultaneously: multiple mediators and treatment non-compliance. Understanding how much the ITT effect is attributed to 2 mediators jointly or to only 1 mediator will enable researchers to identify relevant mediators and thus make informed decisions on how to strengthen the intervention effect.

To avoid bias in the estimation of the mediated ITT effects, the approach of obtaining the mediated portion of the ITT effects estimate was obtained on the basis of the ACME among compliers. One limitation of this approach is that the mediated ITT effects will be biased if any of the assumptions required to identify the ACME among compliers is violated. Therefore, any method that better handles the estimation of the ACME among compliers is likely to improve the estimation of the mediated ITT effects. For example, alternative estimation procedures for estimating the ACME among compliers in the presence of multiple mediators and treatment noncompliance maybe developed under the Bayesian framework. Compared to the current frequentist approach, the Bayesian approach is known to be more efficient in terms of standard errors, and missing values can be simultaneously imputed in the Gibbs sampling algorithm (See Little and Yau<sup>2</sup> and Frangakis and Rubin<sup>36</sup>). Thus, it may be an interesting future research topic to establish a Bayesian estimation procedure.

We proposed a sensitivity analysis for key assumptions that might be violated in the context of the Jobs II example. In our sensitivity analysis, we examined the robustness of our estimates when key assumptions were individually violated. In other words, in each sensitivity analysis, we assessed the effect estimates against the potential violation of one assumption while the others were assumed to be satisfied. This kind of sensitivity analysis would be useful in a setting where the exclusion restriction is supported by design (for instance, double-blind design) and only the unconfoundness among compliers needs to be assessed (or vice versa). In practice, multiple assumptions might be simultaneously violated and have possible interactions between these assumptions. Further studies are needed to develop a sensitivity analysis that considers simultaneous violations of multiple assumptions and better address this complexity. Another drawback of the proposed sensitivity analysis is that it requires the specification of unknown sensitivity parameters. In our study, sensitivity parameters were obtained by assuming a worst-case scenario. Developing a sensitivity analysis that can narrow down ranges of the sensitivity parameters is thus an important area for future research.

## ACKNOWLEDGMENTS

This study was supported by a seed grant from the University of California, Riverside. The content is solely the responsibility of the authors and does not necessarily represent the official reviews of this organization. We thank Teppei Yamamoto for sharing his simulation code. We thank David Kaplan and Peter M. Steiner for their valuable comments. We also thank the Editor and 2 anonymous referees for their comments, which significantly improved the paper. Lastly, we would like to thank Beth C. Tamayose for her assistance in editing this paper.

## ORCID

Soojin Park  <http://orcid.org/0000-0003-0288-5589>

Esra Kürüm  <http://orcid.org/0000-0003-1767-1671>

## REFERENCES

1. Yamamoto T. Identification and estimation of causal mediation effects with treatment noncompliance Unpublished manuscript; 2013.
2. Little RJ, Yau LHY. Statistical techniques for analyzing data from prevention trials: treatment of no-shows using rubin's causal model. *Psychological Methods*. 1998;3(2):147-159.
3. Vinokur AD, Price RH, Schul Y. Impact of the jobs intervention on unemployed workers varying in risk for depression. *Am J Community Psychology*. 1995;23(1):39-74.
4. Vinokur AD, Schul Y. Mastery and inoculation against setbacks as active ingredients in the jobs intervention for the unemployed. *J Consulting Clin Psychology*. 1997;65(5):867-877.
5. McGee RE, Thompson N. J. Peer reviewed: unemployment and depression among emerging adults in 12 states, behavioral risk factor surveillance system, 2010. *Prev Chronic Dis*. 2015;12.
6. Montgomery Jr JA, Frisch MJ, Ochterski JW, Petersson GA. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J Chem Phys*. 1999;110(6):2822-2827.
7. Whooley MA, Kiefe CI, Chesney MA, Markovitz JH, Matthews K, Hulley SB. Depressive symptoms, unemployment, and loss of income: the cardia study. *Arch Internal Med*. 2002;162(22):2614-2620.

8. Price RH, Van Ryn M, Vinokur AD. Impact of a preventive job search intervention on the likelihood of depression among the unemployed. *J Health Social Behav.* 1992;33:158-167.
9. Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies. *Am Political Sci Rev.* 2011;105:765-789.
10. Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. *Political Anal.* 2013;21:141-171.
11. VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. *Epidemiologic Methods.* 2014;2(1):95-115.
12. Robins JM. Semantics of causal dag models and the identification of direct and indirect effects. *Highly Struct Stochastic Syst.* 2003;70-81.
13. Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. *Biometrics.* 2015;71(1):1-14.
14. Caplan RD, Vinokur AD, Price RH, Van Ryn M. Job seeking, reemployment, and mental health: a randomized field experiment in coping with job loss. *J Appl Psychology.* 1989;74(5):759-769.
15. Yau L. HY, Little RJ. Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *J Am Stat Assoc.* 2001;96(456):1232-1244.
16. Catalano R. The health effects of economic insecurity. *Am J Public Health.* 1991;81(9):1148-1152.
17. Catalano R, Dooley C. D. Economic predictors of depressed mood and stressful life events in a metropolitan community. *J Health Social Behav.* 1977;18:292-307.
18. Cobb S, Kasl SV. Termination; the consequences of job loss. NIOSH; 1977.
19. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The hopkins symptom checklist (HSCL): a self-report symptom inventory. *Syst Res Behav Sci.* 1974;19(1):1-15.
20. Rosenberg M. *Society and the Adolescent Self-Image.* New Jersey: Princeton University Press; 2015.
21. Gurin P, Gurin G, Morrison BM. Personal and ideological aspects of internal and external control. *Social Psychology.* 1978;41:275-296.
22. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat.* 1978;6:34-58.
23. Holland PW. Statistics and causal inference. *J Am Stat Assoc.* 1986;81:945-960.
24. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychological Methods.* 2010;15:309-334.
25. Pearl J. The causal mediation formula—a guide to the assessment of pathways and mechanisms. *Prev Sci.* 2012;13(4):426-436.
26. VanderWeele T. *Explanation in Causal Inference: Methods for Mediation and Interaction.* New York: Oxford University Press; 2015.
27. Little RJA, Rubin DB. *Statistical Analysis with Missing Data.* New York: J. Wiley & Sons; 1987.
28. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc.* 1996;91(434):444-455.
29. Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann Stat.* 1997;25:305-327.
30. Jo B, Asparouhov T, Muthén BO. Intention-to-treat analysis in cluster randomized trials with noncompliance. *Stat Med.* 2008;27(27):5565-5577.
31. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics.* 2002;58(1):21-29.
32. Ding P, Lu J. Principal stratification analysis using principal scores. *J R Stat Soc Series B (Stat Methodology).* 2017;79(3):757-777.
33. Stuart EA, Jo B. Assessing the sensitivity of methods for estimating principal causal effects. *Stat Methods Med Res.* 2015;24(6):657-674.
34. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0; 2013.
35. Fan J, Gijbels I. *Local Polynomial Modelling and its Applications.* London: Chapman and Hall; 1996.
36. Frangakis CE, Rubin DB. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika.* 1999;86(2):365-379.
37. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiol (Cambridge, Mass.)* 2010;21(4):540-551.
38. Pearl J. *Causality: Models, Reasoning, and Inference,* 2nd ed. Cambridge: Cambridge University Press; 2009.

**How to cite this article:** Park S, Kürüm E. Causal mediation analysis with multiple mediators in the presence of treatment noncompliance. *Statistics in Medicine.* 2018;37:1810–1829. <https://doi.org/10.1002/sim.7632>

## APPENDIX A: PROOF OF $\delta(\mathbf{Z})$

First, note that the expected potential outcome of  $Y_i$  can be expressed as observable quantities under Assumption 1. We have

$$\begin{aligned}
& E(Y_i(z, M_i(z', W(z')), W_i(z')) | P_i = c) \\
&= \iint E(Y_i(z, m, w) | M_i(z', w) = m, W_i(z') = w, P_i = c) P(M_i(z', w) = m | W_i(z') = w, P_i = c) P(W_i(z') = w | P_i = c) dm dw \\
&= \iint E(Y_i(z, m, w) | P_i = c) P(M_i(z', w) = m | W_i(z') = w, P_i = c) P(W_i(z') = w | P_i = c) dm dw \\
&= \iint E(Y_i(z, m, w) | Z_i = T_i = z, M_i = m, W_i = w, P_i = c) P(M_i(z', w) = m | Z_i = z', W_i = w, P_i = c) P(W_i(z')) \\
&= w | Z_i = z', P_i = c) dm dw \\
&= \iint E(Y_i | M_i = m, W_i = w, Z_i = z, P_i = c) P(M_i = m | Z_i = z', W_i = w, P_i = c) P(W_i = w | Z_i = z', P_i = c) dm dw.
\end{aligned} \tag{A1}$$

The first equality follows from the law of total probability, the second equality follows from Assumption 3, and the third follows from Assumption 2. The last equality follows from the fact that  $Y_i = Y_i(Z_i, M_i(Z_i), W_i(Z_i))$ ,  $M_i = M_i(Z_i, W(Z_i))$ , and  $W_i = W_i(Z_i)$ .

By the same logic, we get

$$\begin{aligned}
& E(Y_i(z, M_i(z, W(z)), W_i(z)) | P_i = c) \\
&= \iint E(Y_i | M_i = m, W_i = w, Z_i = T_i = t, P_i = c) P(M_i = m | W_i = w, Z_i = T_i = t, P_i = c) P(W_i = w | P_i = c, Z_i = T_i = t) dm dw
\end{aligned} \tag{A2}$$

If the compliers are observable, using the difference between (A1) and (A2), the LACME attributed to  $M$  and  $W$  jointly under  $z$  is

$$\begin{aligned}
\delta(z) &= \iint E(Y_i | M_i = m, W_i = w, Z_i = z, P_i = c) \\
&\quad \times \{P(M_i = m | W_i = w, Z_i = 1, P_i = c) P(W_i = w | Z_i = 1, P_i = c) \\
&\quad - P(M_i = m | W_i = w, Z_i = 0, P_i = c) P(W_i = w | Z_i = 0, P_i = c)\} dm dw
\end{aligned} \tag{A3}$$

for  $z \in \{0, 1\}$  and  $z' = 1 - z$ . Using the definitions in Table 1, we have  $\delta(z) = \iint \mu_{mwzc} \{ \varphi_{w1c} \omega_{1c} - \varphi_{w0c} \omega_{0c} \} dm dw$ .

Under Assumption 4 (no defiers), compliance status is fully observable for those who are assigned to the treatment. Those who are both assigned to the treatment and received the treatment ( $Z_i = T_i = 1$ ) are compliers and those who are assigned to the treatment but did not receive the treatment ( $Z_i = 1$  and  $T_i = 0$ ) are never takers (see Figure 2). This implies that  $\mu_{mw1c}$ ,  $\varphi_{w1c}$ ,  $\omega_{1c}$ , and  $\pi_c$  are identified from those who are both assigned to the treatment and received the treatment and  $\mu_{mw1n}$ ,  $\varphi_{w1n}$ ,  $\omega_{1n}$ , and  $\pi_n$  are identified from those who are assigned to the treatment but did not receive the treatment. However, compliance status is only partially known for those who are assigned to the control condition and did not receive treatment ( $Z_i = T_i = 0$ ) since they are either compliers or never takers, which implies that  $\mu_{mw0c}$ ,  $\varphi_{w0c}$ , and  $\omega_{0c}$  are not identified.

Instead, the overall potential value of  $W$  for those who are assigned to the control condition ( $\omega_0$ ) is identified and can be seen as the combination of the respect distributions of compliers and never takers. Likewise, the overall potential value of  $M$  for those who are assigned to the control condition ( $\varphi_{w0} \omega_0$  for every value of  $W = w$ ) and the overall potential outcome for those who are assigned to the control condition ( $\mu_{mw0} \varphi_{w0} \omega_0$  for every values of  $M = m$  and  $W = w$ ) are also seen as the combination of the respect distributions of compliers and never takers. The potential value of mediators and the potential outcome for those who are assigned to the control condition can be expressed as

$$\begin{aligned}
\omega_0 &= \pi_c \omega_{0c} + \pi_n \omega_{0n}, \\
\varphi_{w0} \omega_0 &= \pi_c \varphi_{w0c} \omega_{0c} + \pi_n \varphi_{w0n} \omega_{0n}, \text{ and} \\
\mu_{mw0} \varphi_{w0} \omega_0 &= \pi_c \mu_{mw0c} \varphi_{w0c} \omega_{0c} + \pi_n \mu_{mw0n} \varphi_{w0n} \omega_{0n}
\end{aligned} \tag{A4}$$

for every value of  $M = m$  and  $W = w$ , where  $\pi_c$  and  $\pi_n$  denote the probability of compliers and never takers, respectively. The term  $\omega_{tn}$  represents the population conditional probability of  $W = w$  among never takers when assigned to  $Z = t$ ;  $\varphi_{wtn}$  represents the population conditional probability of  $M = m$  among never takers for the respective values of  $W = w$

and  $Z = t$ ; and  $\mu_{mwt}$  represents the population average of the outcome among never takers for the respective values of  $M = m$ ,  $W = w$ , and  $Z = t$ . Using the equations in (A4),  $\omega_{0c}$ ,  $\omega_{0c}\varphi_{w0c}$ , and  $\mu_{mw0c}$  can be, respectively, expressed as

$$\begin{aligned} \omega_{0c} &= \frac{\omega_0 - \pi_n \omega_{0n}}{\pi_c}, \\ \varphi_{w0c}\omega_{0c} &= \frac{\varphi_{w0}\omega_0 - \pi_n \varphi_{w0n}\omega_{0n}}{\pi_c}, \text{ and} \\ \mu_{mw0c} &= \frac{\mu_{mw0}\varphi_{w0}\omega_0 - \pi_n \mu_{mw0n}\varphi_{w0n}\omega_{0n}}{\pi_c \varphi_{w0c}\omega_{0c}}. \end{aligned} \tag{A5}$$

Under Assumption 5 (the exclusion restriction),  $\mu_{mw0c}$ ,  $\varphi_{w0c}$ , and  $\omega_{0c}$  can be expressed using existing quantities as

$$\begin{aligned} \omega_{0c} &= \frac{\omega_0 - \pi_n \omega_{1n}}{\pi_c}, \\ \varphi_{w0c}\omega_{0c} &= \frac{\varphi_{w0}\omega_0 - \pi_n \varphi_{w1n}\omega_{1n}}{\pi_c}, \text{ and} \\ \mu_{mw0c} &= \frac{\mu_{mw0}\varphi_{w0}\omega_0 - \pi_n \mu_{mw1n}\varphi_{w1n}\omega_{1n}}{\pi_c \varphi_{w0c}\omega_{0c}} = \frac{\mu_{mw0}\varphi_{w0}\omega_0 - \pi_n \mu_{mw1n}\varphi_{1n}\omega_{1n}}{\varphi_{w0}\omega_0 - \pi_n \varphi_{w1n}\omega_{1n}}. \end{aligned} \tag{A6}$$

Using the first two equations of (A6), we can calculate the compliers-difference in the potential value of  $W$  between the treatment and control conditions ( $\omega_{1c} - \omega_{0c}$ ). Likewise, we can calculate the compliers-difference in the potential value of  $M$  between the treatment and control conditions for the value of mediator  $W = w$  ( $\varphi_{w1c}\omega_{1c} - \varphi_{w0c}\omega_{0c}$ ).

$$\begin{aligned} \omega_{1c} - \omega_{0c} &= \omega_{1c} - \frac{\omega_0 - \pi_n \omega_{1n}}{\pi_c} = \frac{\omega_1 - \omega_0}{\pi_c} \text{ and} \\ \varphi_{w1c}\omega_{1c} - \varphi_{w0c}\omega_{0c} &= \varphi_{w1c}\omega_{1c} - \frac{\varphi_{w0}\omega_0 - \pi_n \varphi_{w1n}\omega_{1n}}{\pi_c} = \frac{\varphi_{w1}\omega_1 - \varphi_{w0}\omega_0}{\pi_c}. \end{aligned} \tag{A7}$$

Now, plugging Equations A6 and A7 into Equation 4, the LACME and LANDE are nonparametrically identified as Equation 5. This completes the proof.

### APPENDIX B: PROOF OF IDENTIFYING $\delta^W(Z)$

First, note that the expected potential outcome of  $Y_i$  can be expressed as observable quantities under Assumption 1. We have

$$\begin{aligned} &E(Y_i(z, M_i(z, W(z')), W_i(z'))|P_i = c) \\ &= \int E(Y_i(z, M_i(z, w), w)|W_i(z') = w, P_i = c) \times P(W_i(z') = w|P_i = c)dw \\ &= \int E(Y_i(z, M_i(z, w), w)|P_i = c) \times P(W_i(z') = w|P_i = c)dw \\ &= \int E(Y_i(z, M_i(z, w), w)|W_i = w, Z_i = T_i = z, P_i = c)P(W_i(z') = w|Z_i = T_i = z', P_i = c)dw \\ &= \int E(Y_i|W_i = w, Z_i = T_i = z, P_i = c) \times P(W_i = w|Z_i = T_i = z', P_i = c)dw. \end{aligned} \tag{B1}$$

The first equality follows from the law of total probability and the second equality follows from Assumption (7). The third equality follows from Assumption 2. The last equality follows from the fact that  $Y_i = Y_i(Z_i, M_i(Z_i), W_i(Z_i))$ ,  $M_i = M_i(Z_i, W(Z_i))$ , and  $W_i = W_i(Z_i)$ .

By the same logic, we get

$$\begin{aligned} &E(Y_i(z, M_i(z, W(z)), W_i(z))|P_i = c) \\ &= \int E(Y_i|M_i = m, Z_i = T_i = z, P_i = c) \times P(W_i = w|P_i = c, Z_i = T_i = z)dw \end{aligned} \tag{B2}$$

Using the difference between (B1) and (B2), the LACME that goes through  $W$  under  $z$  is identified as follows if the compliers are observable.

$$\delta^W(z) = \int E(Y_i|W_i = w, Z_i = z, P_i = c) \{P(W_i = w|Z_i = 1, P_i = c) - P(W_i = w|Z_i = 0, P_i = c)\} dw. \quad (B3)$$

This completes the proof.

### APPENDIX C: EXTENSION TO THE TWO-SIDED COMPLIANCE CASE

Under Assumption 4 (the no defiers assumption), those who are assigned to the treatment but did not receive the treatment ( $Z_i = 1$  and  $T_i = 0$ ) are identified as never takers and those who are assigned to the control but received the treatment ( $Z_i = 0$  and  $T_i = 1$ ) are identified as always takers. This implies that  $\mu_{mw1n}$ ,  $\varphi_{w1n}$ ,  $\omega_{1n}$ , and  $\pi_n$  are identified from the treatment arm and  $\mu_{mw0a}$ ,  $\varphi_{w0a}$ ,  $\omega_{0a}$ , and  $\pi_a$  are identified from the control arm. However, compliance status is not known for those who are assigned to the treatment condition and received the treatment ( $Z_i = T_i = 1$ ) since they are either compliers or always takers. Likewise, compliance status is not known for those who are assigned to the control condition and did not receive the treatment ( $Z_i = T_i = 0$ ) since they are either compliers or never takers.

The potential outcomes of  $W$ ,  $M$ , and  $Y$  under  $Z_i = z$  are expressed as

$$\begin{aligned} \omega_z &= \pi_c \omega_{zc} + \pi_a \omega_{za} + \pi_n \omega_{zn}, \\ \varphi_{wz} \omega_z &= \pi_c \varphi_{wzc} \omega_{zc} + \pi_a \varphi_{wza} \omega_{za} + \pi_n \varphi_{wzn} \omega_{zn}, \text{ and} \\ \mu_{mwz} \varphi_{wz} \omega_z &= \pi_c \mu_{mwzc} \varphi_{wzc} \omega_{zc} + \pi_a \mu_{mwza} \varphi_{wza} \omega_{za} + \pi_n \mu_{mwzn} \varphi_{wzn} \omega_{zn} \end{aligned} \quad (C1)$$

for every value of  $M = m$  and  $W = w$ , where  $\pi_c$ ,  $\pi_a$ , and  $\pi_n$  denote the probability of compliers, always takers, and never takers, respectively. The term  $\omega_{za}$  represents the population conditional probability of  $W = w$  among always takers when assigned to  $Z = z$ ;  $\varphi_{wza}$  represents the population conditional probability of  $M = m$  among always takers for the respective values of  $W = w$  and  $Z = z$ ; and  $\mu_{mwza}$  represents the population average of the outcome among always takers for the respective values of  $M = m$ ,  $W = w$ , and  $Z = z$ .

By assuming Assumption 5 (the exclusion restriction),  $\omega_{zc}$ ,  $\omega_{zc} \varphi_{wzc}$ , and  $\mu_{mwzc}$  can respectively be expressed as

$$\begin{aligned} \omega_{zc} &= \frac{\omega_z - \pi_a \omega_{1a} - \pi_n \omega_{0n}}{\pi_c}, \\ \varphi_{wzc} \omega_{zc} &= \frac{\varphi_{wz} \omega_z - \pi_a \varphi_{w1a} \omega_{1a} - \pi_n \varphi_{w0n} \omega_{0n}}{\pi_c}, \text{ and} \\ \mu_{mwzc} &= \frac{\mu_{mwz} \varphi_{wz} \omega_z - \pi_a \mu_{mw1a} \varphi_{w1a} \omega_{1a} - \pi_n \mu_{mw0n} \varphi_{w0n} \omega_{0n}}{\varphi_{wz} \omega_z - \pi_a \varphi_{w1a} \omega_{1a} - \pi_n \varphi_{w0n} \omega_{0n}} \end{aligned} \quad (C2)$$

where  $\pi_c$  is identified as  $1 - \pi_n - \pi_a$ .

The LACME and LANDE can be expressed as below by plugging Equation C2 into Equation 4:

$$\begin{aligned} \delta^W(z) &= \int \mu_{wzc} \times \left\{ \frac{\omega_1 - \omega_0}{\pi_c} \right\} dw, \\ \delta(z) &= \iint \mu_{mwzc} \times \left\{ \frac{\varphi_{w1} \omega_1 - \varphi_{w0} \omega_0}{\pi_c} \right\} dmdw, \text{ and} \\ \zeta(z) &= \iint \left\{ \frac{\mu_{mw1} \varphi_{w1} \omega_1 - \pi_a \mu_{mw1a} \varphi_{w1a} \omega_{1a} - \pi_n \mu_{mw0n} \varphi_{w0n} \omega_{0n}}{\varphi_{w1} \omega_1 - \pi_a \varphi_{w1a} \omega_{1a} - \pi_n \varphi_{w0n} \omega_{0n}} - \frac{\mu_{mw0} \varphi_{w0} \omega_0 - \pi_a \mu_{mw1a} \varphi_{w1a} \omega_{1a} - \pi_n \mu_{mw0n} \varphi_{w0n} \omega_{0n}}{\varphi_{w0} \omega_0 - \pi_a \varphi_{w1a} \omega_{1a} - \pi_n \varphi_{w0n} \omega_{0n}} \right\} \\ &\quad \times \varphi_{wzc} \omega_{zc} dmdw. \end{aligned} \quad (C3)$$

The term  $\mu_{wzc}$ ,  $\mu_{mwzc}$ , and  $\varphi_{wzc} \omega_{zc}$  are identified as in Equation C2.

### APPENDIX D: PARAMETERS FOR COMPLIERS AND NEVER TAKERS

The parameter means for compliers and never takers that were used in Equation 12 are



$$\begin{aligned}\alpha_1 &\sim N(0.5\{P_i = c\} + 1.5\{P_i = n\}, \sigma), \beta_1 \sim N(1\{P_i = c\} + 1.6\{P_i = n\}, \sigma), \\ \alpha_2 &\sim N(-1\{P_i = c\} - 0.5\{P_i = n\}, \sigma), \beta_2 \sim N(1\{P_i = c\} + 0.6\{P_i = n\}, \sigma), \\ \beta_3 &\sim N(1\{P_i = c\} + 0.2\{P_i = n\}, \sigma), \gamma_1 \sim N(1\{P_i = c\} + 0.3\{P_i = n\}, \sigma), \\ \alpha_3 &\sim N(1\{P_i = c\} + 0.5\{P_i = n\}, \sigma), \beta_4 \sim N(1\{P_i = c\} + 0.2\{P_i = n\}, \sigma), \\ \beta_5 &\sim N(0.3\{P_i = c\} + 0.2\{P_i = n\}, \sigma), \beta_6 \sim N(1\{P_i = c\} + 0.8\{P_i = n\}, \sigma), \\ \gamma_2 &\sim N(1\{P_i = c\} + 0.2\{P_i = n\}, \sigma), \gamma_3 \sim N(1\{P_i = c\} + 0.2\{P_i = n\}, \sigma), \text{ and} \\ \gamma_4 &\sim N(-0.5\{P_i = c\} + 1\{P_i = n\}, \sigma)\end{aligned}\tag{D1}$$

for  $\sigma$  ranges from 0 to 2 SD.