Methods for Policy Analysis

Burt S. Barnow, Editor

UNPACKING COMPLEX MEDIATION MECHANISMS AND THEIR HETEROGENEITY BETWEEN SITES IN A JOB CORPS EVALUATION

Xu Qin, Jonah Deutsch, and Guanglei Hong

Abstract

This study aims to test the theory underlying Job Corps, one of the largest education and training programs in the U.S. serving disadvantaged youth. Central to the program are vocational training and general education that serve as two concurrent mediators transmitting the program impact on earnings. To distinguish the relative contribution of each, we develop methods for decomposing the Job Corps impact on earnings into an indirect effect transmitted through vocational training, an indirect effect transmitted through general education, and a direct effect attributable to supplementary services. We further ask whether general education and vocational training reinforce each other and produce a joint impact greater than the sum of the two separate pathways. Moreover, we examine the heterogeneity of each causal effect across all the Job Corps centers. This article presents concepts and methods for defining, identifying, and estimating not only the population averages but also the between-site variance of these causal effects. Our analytic procedure incorporates a series of weighting strategies to enhance the internal and external validity of the results and assesses the sensitivity to potential violations of the identification assumptions. © *2020 by the Association for Public Policy Analysis and Management*

INTRODUCTION

Social programs often have multiple components since the targeted change they wish to bring about may involve more than one mechanism. As such, the theory underlying a program may suggest multiple pathways operating jointly to produce a desired outcome. Each pathway may represent the impact of the program on the outcome that may operate through a specific *mediator*. Different pathways may correspond to different components of the program. An intent-to-treat (ITT) analysis,

Journal of Policy Analysis and Management, Vol. 40, No. 1, 158–190 (2021)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.22268

albeit important, does not explicitly test the theory underlying a multifaceted program. Only a mediation analysis does.

Here we characterize three possible complex mediation mechanisms involving two parallel pathways. By design, the pathways transmitting the impacts of different program components are intended to be at least *complementary* to one another. For example, program developers may expect that pathway B adds onto pathway A. An absence of pathway B, in this case, does not necessarily undermine pathway A. If, however, pathway A is strengthened in the presence of pathway B and is weakened in the absence of the latter, and if the same is true vice versa, then the two pathways are *mutually reinforcing*. Sometimes, unforeseen by the program developers, an unintended pathway may produce a side effect that offsets the program impact transmitted through the intended pathway. When this is true, a null ITT effect may disguise a mechanism featuring two *counteracting* mediators.

In organizations or communities serving as sites, an intervention program may have the capacity for accommodating some but not all of the eligible individuals. A random assignment of eligible applicants to the program makes it possible to identify the program impact site by site. Researchers have anticipated that the ITT effect of a program may not be constant due to natural variations in local contexts, in participant composition, and in treatment implementation, among other factors (Weiss, Bloom, & Brock, 2014). Decomposing the total effect of a treatment into one or more indirect effects and a direct effect constitutes an essential step in testing the theoretical mechanisms. Each indirect effect is transmitted through a focal mediator while the direct effect is attributable to all other unspecified pathways. Unpacking between-site heterogeneity in the complex mechanisms further informs the generalizability of the program theory and may suggest specific site-level modifications for improving the intervention practice.

Research Questions

This study is motivated by past discussions, debates, and evaluations of Job Corps, one of the largest education and training programs in the U.S. for disadvantaged youths who are unemployed or underemployed. Analyzing data from the National Job Corps Study (NJCS), a rigorous and comprehensive evaluation of the program, Schochet, Burghardt, and McConnell (2008) found that Job Corps increased earnings for several years, but that these earnings gains did not persist in later years according to tax data, except for participants already in their 20s when entering the program. To date, the program's effectiveness continues to be a subject of contentious public debate (*New York Times*, August 26, 2018).

Most Job Corps applicants have dropped out of high school and hence lack human capital that is believed to be key to economic productivity. Becker (1964) made a further distinction between *general human capital* and *job-specific human capital*. The former includes education credentials as a proxy for literacy skills and work ethic, while the latter refers to technical knowledge or skills applicable in a certain vocational trade that may not transfer easily to other trades. Most job training programs tend to focus solely on vocational training. In contrast, Job Corps places both vocational training and general education at the center of the program. Its general education curriculum prepares those without a high school diploma to qualify for a GED certificate. Similar to other education and training programs, the Job Corps program design is consistent with human capital theory. It is unclear, however, whether the education pathway and the vocational training pathway are *complementary* or *mutually reinforcing*. Two complementary pathways are expected to transmit the program impact in an additive manner, while two mutually reinforcing pathways are expected furthermore to have a positive interaction effect. Past

research has suggested that general education and vocational training are at least complementary (Blundell et al., 1999; Zimmermann et al., 2013). Yet one may argue that, for vocational training to be effective in an economy with rapid technological change, a student may need basic academic preparation as a prerequisite (Hanushek et al., 2017). Following this reasoning, high school dropouts may benefit more from vocational training when they work toward a general education credential at the same time rather than receiving vocational training alone. It is of important theoretical and practical interest, therefore, to distinguish the relative contribution of each pathway and determine whether these two types of human capital investments reinforce each other and generate a joint impact greater than the sum of the two separate pathways. The latter is also known as an interaction effect. In our conceptual framework, general education and vocational training are focal mediators that constitute indirect pathways for Job Corp impacts.

Job Corps may also impact participants' earnings through pathways other than traditional human capital accumulation. In recognition of the multiple barriers that disconnected youths typically encounter in their transition to adulthood, education and training at Job Corps are supplemented by a wide array of services: individual and group counseling, behavioral management, social skills training, job search assistance, and drug and alcohol treatment. One of the most unique services provided by Job Corps is the residential component, in which most participants are required to live at the centers during the week, thus creating a protective environment for them as they engage in training and other services. This set of services constitutes a direct pathway in our framework.

Hence, our first set of research questions are:

1. What is the average program impact on earnings mediated by vocational training?
2. What is the average program impact on earnings mediated by general education?
3. Is the program impact mediated by vocational training reinforced by general education?
4. What is the average direct effect of the program transmitted through other mechanisms?

The program is delivered at more than 100 Job Corps centers around the country. Although Job Corps has shown on average a positive impact on earnings four years following random assignment (Schochet, Burghardt, & McConnell, 2008), the impact has been found to be uneven across sites (Weiss et al., 2017). The impact, however, did not vary systematically by the site characteristics examined in a previous study: operator type (private contractor vs. government agency), site size, region, or the rating of the site received in Job Corps' performance measurement system (Burghardt & Schochet, 2001).

According to past research, Job Corps' GED curriculum and vocational training curriculum have been standardized and strictly regulated either by the national Job Corps office and its regional offices or by national trade unions; in contrast, the management of other services has been left largely to the discretion of each local center (Johnson et al., 1999). Given this observation, one may hypothesize that Job Corps' centralized organization of education and training may have contributed to a relatively consistent impact of the program on human capital accrual, and that the variation in the ITT effect between the sites may be attributed primarily to the varying quantity and quality of other services. To empirically test these hypotheses, we ask a second set of research questions:

5. Does the program impact mediated by vocational training vary across sites?
6. Does the program impact mediated by general education vary across sites?

7. Does the direct effect of the program vary across sites?

These questions have not been addressed empirically in past evaluations of Job Corps. Researchers have typically combined education and training into a single measure (e.g., Flores & Flores-Lagunes, 2013; Qin & Hong, 2017; Qin et al., 2019). Hence it has been unclear whether Job Corps' investment in GED is worthwhile especially in light of the recent debate concerning the economic value of obtaining a GED certificate (Heckman & LaFontaine, 2006; Tyler, 2003). The possible direct effect of Job Corps attributable to the wide range of supplementary services deserves attention as well because these services are costly and are generally unavailable to disadvantaged youth in the absence of Job Corps. While policy discussions are sometimes dominated by the media that tend to focus on the underperformance of one or two Job Corps centers (e.g., *Miami Herald*, August 19, 2015), major decisions about the program should be made in light of the entire distribution of impacts across all the centers. This study contributes to substantive knowledge by not only making a fine-grained decomposition of the *average* program impact into pathways transmitted through general education, vocational training, and supplementary services, but also depicting the *distribution* of each for the population of Job Corps centers.

Methodological Contributions

In social science research, structural equation modeling (SEM) (Bollen, 1987; Jo, 2008; Jöreskog, 1970; MacKinnon, 2008; MacKinnon & Dwyer, 1993) has been a primary strategy for investigating mediation mechanisms involving multiple mediators. However, it is expected to generate biased estimates of the direct and indirect effects if the mediator and outcome models are misspecified (Bullock, Green, & Ha, 2010; Holland, 1988). Typically, an analyst may overlook nonlinear or non-additive relationships. In particular, he or she may ignore the fact that the treatment effect on the outcome may be generated not only through changing the mediator but also through changing the mediator-outcome relationship (Judd & Kenny, 1981).

Causal inference methods for investigating complex mediation mechanisms have begun to emerge recently. These include studies of multiple mediators that are either consecutive (e.g., one mediator affecting another) or concurrent (e.g., two mediators being parallel). Mediated effects defined in terms of potential outcomes are unconstrained by any specific structural models (Pearl, 2001; Robins & Greenland, 1992). Yet in practice, researchers have developed a series of analytic strategies within the system of linear structural question models or generalized linear structural equation models (Albert & Nelson, 2011; Daniel et al., 2015; Imai & Yamamoto, 2013; VanderWeele, 2015). When the treatment is randomized, these strategies all assume that the assignment of mediator values is as if randomized within levels of observed pre-treatment covariates. The identification assumptions, however, are typically accompanied by model-based assumptions. The analyst must correctly specify each mediator model and the outcome model and often further invoke distributional assumptions about the mediator and outcome measures.

A different set of studies in the context of multisite randomized trials has employed variants of a multiple instrumental variables (IV) strategy that uses the interactions between site indicators and the random treatment assignment within each site as instruments (Bloom et al., 2020; Duncan, Morris, & Rodrigues, 2011; Gennetian et al., 2005; Kling, Liebman, & Katz, 2007; Raudenbush, Reardon, & Nomi, 2012; Reardon & Raudenbush, 2013; Reardon et al., 2014). In investigations of multiple mediators, the identification required that all mediators be conditionally independent and that the treatment effect on the mediator should not co-vary with the mediator effect on the outcome at the site level. Moreover, some applications

required the assumption that the direct effect be zero, an assumption that often conflicted with theory; while others required that the effect of one mediator should not depend on any other mediators in the model—i.e., there is no moderated mediation, which would rule out the possibility of empirically examining whether two concurrent mediators are mutually reinforcing.

As stated earlier, it is well known that causal analytic results are sensitive to violations of parametric modeling assumptions (Drake, 1993; Goldberger, 1983; Schafer & Kang, 2008). In the presence of interactions between the treatment, mediators, and covariates, even if the structural models are correctly specified, the indirect effect estimator will take a rather complex form, which will add considerable complexity to estimation and statistical inference. More challenges will arise in the estimation of the between-site variance of the indirect effect in multisite trials (Bauer, Preacher, & Gil, 2006).

A weighting-based approach to mediation analysis (Hong, 2010a, 2015; Hong, Deutsch, & Hill, 2011, 2015; Hong & Nomi, 2012; Huber, 2014; Lange, Rasmussen, & Thygesen, 2014; Lange, Vansteelandt, & Bekaert, 2012; Tchetgen Tchetgen, 2013; Tchetgen Tchetgen & Shpitser, 2012), which Hong (2010a) named ratio-of-mediator-probability weighting (RMPW), relaxes the modeling constraints as it does not need to specify, nor does it simply assume away, treatment-by-mediator, treatment-by-covariate, mediator-by-covariate, or treatment-by-mediator-by-covariate interactions in the outcome model. These methodological advances are important for the current study: We expect that the impacts of education and training on earnings may differ between the Job Corps setting and a control setting if participants receive education and training of a higher quality in Job Corps than in other alternative programs. We also expect, as indicated by past research (Qin et al., 2019), that these impacts may differ between individuals and between sites. To reduce reliance on model specifications, Lange, Rasmussen, and Thygesen (2014) extended RMPW to the case of multiple concurrent mediators. This method estimates each of the direct and indirect effects as a contrast between weighted mean outcomes, which greatly reduces the required model-based assumptions and simplifies the estimation.

However, to our knowledge, there have been no formal scholarly discussions about the unique research opportunities and methodological challenges that arise in investigations of complex mediation mechanisms that may vary across sites. Extending the RMPW approach, recent work by Qin and Hong (2017) and by others (Bein et al., 2018; Hong, Deutsch, & Hill, 2015; Hong, Qin, & Yang, 2018; Qin et al., 2019) has advanced methods for causal mediation analysis accompanied by sensitivity analysis in single-level and multi-level settings. Yet these new advances have focused on a single mediator. Building directly on this line of work, the current study makes several methodological contributions. First and most importantly, we develop concepts and methods for defining, identifying, and estimating not only the population average but also the between-site variance of the program impact transmitted through each of the two concurrent mediators of focal interest, in addition to the population average and the between-site variance of the direct effect. We further examine whether the two concurrent mediators are complementary or mutually reinforcing. Second, unlike SEM and most existing causal mediation analysis strategies that require the analyst to correctly specify both the mediator and the outcome models, we adopt a weighting strategy that does not require outcome model specification. Third, in quantifying the variance of our estimators, we take into account the uncertainty of the estimated weights by capitalizing on the generalized method of moments (GMM) framework (Hansen, 1982; Newey, 1984). Fourth, to assess the potential consequences of violations of identification assumptions due to omitted confounders including a measure of compliance, we extend a novel weighting-based sensitivity analysis strategy. Fifth, the proposed analytic procedure promises to enhance both

the external validity and internal validity of the conclusions about the population average and the between-site variance of the causal direct and indirect effects. Sixth, to enable applied researchers to easily implement the proposed procedure for unpacking complex mediation mechanisms, we provide an R package "MultisiteMediation" that is user-friendly as it involves running only one line of code.

After introducing the National Job Corps Study data in the next section, we define the causal parameters under the potential outcomes framework. We then clarify the identification assumptions and present our identification strategy. The subsequent section summarizes our estimation and statistical inference procedure. The presentation of analytic results is accompanied by a sensitivity analysis. The last section concludes and discusses future topics.

NATIONAL JOB CORPS STUDY DATA

Research Design and Target Population

The National Job Corps Study (NJCS) was funded by the U.S. Department of Labor and conducted by Mathematica Policy Research. Through a stratified sampling procedure, 15,386 eligible applicants were selected into a nationally representative sample drawn from the 80,883 first-time applicants eligible for Job Corps nationwide between November 1994 and February 1996. Among them, 9,409 youths were randomly assigned to the program group, and 5,977 youths were randomly assigned to the control group. Sample members in the program group could enroll in Job Corps soon after random assignment, while those in the control group were barred from enrolling in Job Corps for three years although they could enroll in other programs. For each eligible applicant, the Job Corps center he or she would potentially be assigned to was determined prior to the random treatment assignment (Schochet, Burghardt, & Glazerman, 2001). Hence, the study has a multisite randomized design with Job Corps centers serving as experimental sites.

In a multisite study, there are two potential target populations: One is the population of individuals over all the sites; and the other is the population of sites (Raudenbush & Bloom, 2015; Raudenbush & Schwartz, 2020). The former would be the target of inference if researchers were primarily interested in the overall performance of a program among all the individuals in the population. In such a case, the population average program impact is defined as the average over all eligible individuals who would apply for the program. In this study, we choose the population of sites as our target population, given our primary interest in revealing the underlying causal mediation mechanism that contributes to the between-site variance of the Job Corps program impact. As such, we define the population average impact as the average of the site-specific impacts taken over the sites. We are also interested in the generalizability of the hypothesized mechanisms across sites, and thus focus attention on the between-site variance of the site-specific impact.

Variables of Interest and Study Sample

NJCS researchers conducted surveys with sampled participants shortly after randomization and at the 12-, 30-, and 48-month follow-ups (Schochet, Burghardt, & Glazerman, 2001). The baseline survey data contain a rich set of measures on demographic characteristics, educational attainment, labor market experience, criminal behavior, parental education and employment, and mental and physical health prior to the randomization. Corresponding to our theoretical questions, one of the two mediators indicates whether or not an individual gained a vocational train-

Table 1. Patterns of missing key variables of interest among non-responders.

Missing Center ID[1]	Missing Mediator (s)	Missing Outcome	*Treatment N (percent)*	*Control N (percent)*
		√	700 (11%)	453 (10%)
	√	√	578 (9%)	456 (10%)
	√		603 (9%)	352 (8%)
√			359 (6%)	241 (5%)
√	√	√	130 (2%)	188 (4%)
√		√	56 (1%)	41 (1%)
√	√		44 (1%)	38 (1%)
Total number of non-responders			2,470 (38%)	1,769 (40%)
Total number of responders			3,968 (62%)	2,646 (60%)

Notes: [1] As described in Schochet, Burghardt, and Glazerman (2001), the center ID variable was obtained by asking Job Corps outreach and admissions counselors which center a study applicant would likely be assigned to if he or she were enrolled in the program. This information was gathered prior to random assignment and is thus available for both the treatment and control group. In some cases, the counselors did not respond or were not able to predict a specific center.

ing certificate, and the other indicates whether the youth obtained a high school diploma or GED within the 30 months after the random assignment. The outcome is self-reported average weekly earnings during the year preceding the 48-month interview (in 1995 dollars). Schochet, Burghardt, and McConnell (2006) reported that Job Corps did not generate significant impacts on earnings as measured by tax data, except for the older participants, 5 to 10 years after the randomization. Schochet (2020) examined even longer-term outcomes, using tax data, and found that the program increased employment and tax filing rates and reduced disability benefit receipt for older Job Corps students 20 years following random assignment. As we do not have access to these long-term, administrative data records, we focus only on the shorter-term, self-reported earnings.

We begin with a sample of 14,125 youths who were targeted for the 48-month survey. Of theoretical interest to this study are youths who had neither a high school diploma or equivalent nor a vocational certification at the baseline, and who constitute the majority of the Job Corps applicants. After excluding youths who had a high school diploma, GED, or a vocational certificate at baseline, we are left with a sample of 10,853 individuals. Among them, 6,614 individuals have non-missing values for key variables: (1) the mediators, (2) the outcome, and (3) the centers they were assigned to prior to random assignment. The remaining 4,239 individuals are non-responders in this study. Table 1 displays the patterns of non-response.

Because the sample and survey weights by design are not a function of baseline educational and vocational attainment, the same weights apply to the subsample of Job Corps applicants who lacked an education or training credential at baseline. Through estimating the conditional probability of responding for each respondent, we construct non-response weights to transform the distributions of the observed pre-treatment covariates of the respondents such that they represent the sample of the respondents and the non-respondents combined. We clarify key assumptions and analytic details of the non-response weighting strategy in the subsequent sections. Combining the sample and survey design weights with non-response weights, we aim to generalize our analytic results based on respondents to a theoretical population of Job Corps centers serving disadvantaged youths who lacked an education or training credential at baseline.

DEFINITION OF THE CAUSAL EFFECTS

Individual-Specific Causal Effects

We define the causal effects of interest in terms of potential outcomes (Neyman & Iwaszkiewicz, 1935; Rubin, 1978). Let T_{ij} denote the treatment assignment of individual i at site j. It takes values $t = 1$ or 0 indicating the individual was assigned to the Job Corps program or the control group, respectively. The two concurrent mediators, each being a potential intermediate outcome of the treatment assignment, are denoted by $M_{Vij}(t)$ for vocational training attainment and $M_{Eij}(t)$ for general education attainment under treatment t. Hence, we denote the individual's potential mediators under the Job Corps condition as $M_{Vij}(1)$ and $M_{Eij}(1)$ and those under the control condition as $M_{Vij}(0)$ and $M_{Eij}(0)$. For each individual, we observe only the potential mediators associated with the treatment condition that the individual was actually assigned to. Each of these potential mediators is binary, taking value one if the individual obtained a credential in the corresponding domain, and zero otherwise.

Similarly, we use $Y_{ij}(t)$ to represent the potential outcome associated with treatment condition t for individual i at site j. Because the potential outcome depends on both the treatment assignment and the corresponding potential mediators, it can be equivalently written as $Y_{ij}(t, M_{Vij}(t), M_{Eij}(t))$. When $M_{Vij}(t) = m_V$ and $M_{Eij}(t) = m_E$, the potential outcome can be written as $Y_{ij}(t, m_V, m_E)$. Again, only one potential outcome is observed for each individual, depending on which treatment condition the individual was actually assigned to.

The above definitions are based on the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980, 1986, 1990), which requires no interference between sites and no interference between individuals within each site (Hong & Raudenbush, 2006). The former seems reasonable because Job Corps centers are located far apart from each other and applicants were assigned to centers relatively close to their original residences. However, the latter might be violated if Job Corps participants at the same center or individuals sharing a social network within a site affected each other's behaviors.

For individual i at site j, by contrasting the Job Corps condition with the control condition, we define the ITT effect on vocational attainment denoted by $\beta_{ij}^{(T.V)}$, the ITT effect on educational attainment denoted by $\beta_{ij}^{(T.E)}$, and the ITT effect on the outcome denoted by $\beta_{ij}^{(T)}$:

$$\beta_{ij}^{(T.V)} = M_{Vij}(1) - M_{Vij}(0);$$

$$\beta_{ij}^{(T.E)} = M_{Eij}(1) - M_{Eij}(0).$$

$$\beta_{ij}^{(T)} = Y_{ij}\left(1, M_{Vij}(1), M_{Eij}(1)\right) - Y_{ij}\left(0, M_{Vij}(0), M_{Eij}(0)\right).$$

As illustrated in Figure 1, the ITT effect of Job Corps on the outcome can be decomposed into two indirect effects, one transmitted through vocational training and the other through general education, in addition to a direct effect. Following past research (e.g., Hong, 2015; Pearl, 2001; Robins & Greenland, 1992; VanderWeele, 2015), we define the direct and indirect effects in terms of potential outcomes. Our definitions involve two other potential outcomes: $Y_{ij}(1, M_{Vij}(1), M_{Eij}(0))$ denotes the individual's potential outcome under the Job Corps condition when the treatment counterfactually does not change general education attainment; $Y_{ij}(1, M_{Vij}(0), M_{Eij}(0))$ denotes the individual's potential outcome under the

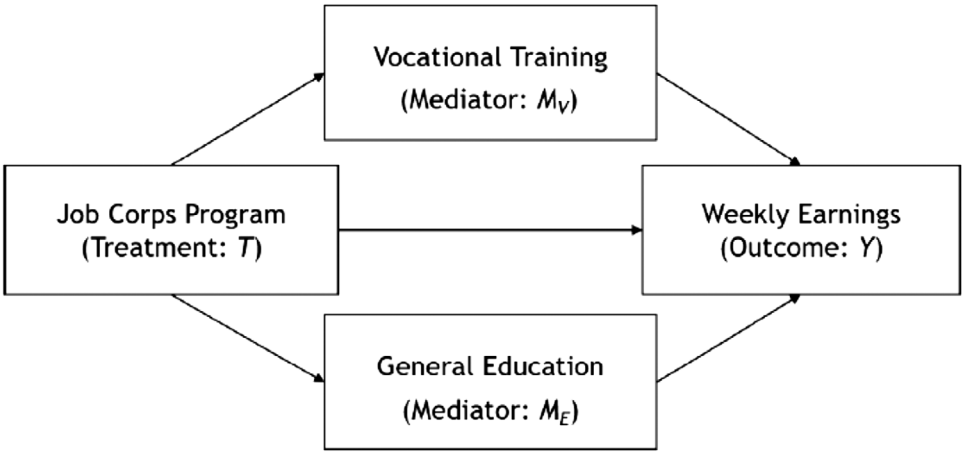


Figure 1. Causal Mediation in the Presence of Two Concurrent Mediators.

Job Corps condition when the treatment counterfactually affects neither general education nor vocational training. We define the indirect effect that operates through vocational training without affecting general education as

$$\beta_{ij}^{(I.V)}(0) = Y_{ij}\left(1, M_{Vij}(1), M_{Eij}(0)\right) - Y_{ij}\left(1, M_{Vij}(0), M_{Eij}(0)\right).$$

This represents the Job Corps impact on earnings to be attributed to the program-induced change in the person's vocational training attainment from $M_{Vij}(0)$ to $M_{Vij}(1)$, while general educational attainment is counterfactually maintained at its level under the counterfactual control condition. The indirect effect transmitted through general education beyond affecting vocational training can be defined as

$$\beta_{ij}^{(I.E)}(1) = Y_{ij}\left(1, M_{Vij}(1), M_{Eij}(1)\right) - Y_{ij}\left(1, M_{Vij}(1), M_{Eij}(0)\right).$$

This is the Job Corps impact attributable to the program-induced change in the person's general education attainment from $M_{Eij}(0)$ to $M_{Eij}(1)$, above and beyond the improvement in vocational training attainment. We further define the direct effect as

$$\beta_{ij}^{(D)}(0) = Y_{ij}\left(1, M_{Vij}(0), M_{Eij}(0)\right) - Y_{ij}\left(0, M_{Vij}(0), M_{Eij}(0)\right),$$

which is the effect that operates through all other unspecified mechanisms.

The ITT effect on the outcome is equal to the sum of the direct effect $\beta_{ij}^{(D)}(0)$ and the total indirect effect, the latter being the sum of the two indirect effects, $\beta_{ij}^{(I.V)}(0)$ and $\beta_{ij}^{(I.E)}(1)$. The total indirect effect can be alternatively decomposed into

$$\beta_{ij}^{(I.V)}(1) = Y_{ij}\left(1, M_{Vij}(1), M_{Eij}(1)\right) - Y_{ij}\left(1, M_{Vij}(0), M_{Eij}(1)\right),$$

$$\beta_{ij}^{(I.E)}(0) = Y_{ij}\left(1, M_{Vij}(0), M_{Eij}(1)\right) - Y_{ij}\left(1, M_{Vij}(0), M_{Eij}(0)\right).$$

The two decompositions are not equivalent in the presence of an interaction between the two mediators. For example, although both $\beta_{ij}^{(I.V)}(1)$ and $\beta_{ij}^{(I.V)}(0)$ repre-

sent the indirect effects transmitted through vocational training, general education attainment is kept at the level under the Job Corps condition in $\beta_{ij}^{(I.V)}(1)$ but kept at the level under the control condition in $\beta_{ij}^{(I.V)}(0)$. If the program impact mediated by vocational training depended on the improvement in general education, $\beta_{ij}^{(I.V)}(1)$ would not be equal to $\beta_{ij}^{(I.V)}(0)$; and similarly, $\beta_{ij}^{(I.E)}(1)$ would not be equal to $\beta_{ij}^{(I.E)}(0)$. When the ITT effects on educational attainment and vocational attainment are both positive, and when $\beta_{ij}^{(I.V)}(1)$ is positive and greater than $\beta_{ij}^{(I.V)}(0)$, it will suggest that the indirect effect transmitted through vocational training may be reinforced by general education attainment. The degree of this potential reinforcement, known as the interaction effect between the two mediators, is defined as the difference between $\beta_{ij}^{(I.V)}(1)$ and $\beta_{ij}^{(I.V)}(0)$, which is numerically equivalent to the difference between $\beta_{ij}^{(I.E)}(1)$ and $\beta_{ij}^{(I.E)}(0)$:

$$\beta_{ij}^{(I.V\times E)}(1) = \left[Y_{ij}\left(1, M_{Vij}(1), M_{Eij}(1)\right) - Y_{ij}\left(1, M_{Vij}(1), M_{Eij}(0)\right)\right]$$

$$- \left[Y_{ij}\left(1, M_{Vij}(0), M_{Eij}(1)\right) - Y_{ij}\left(1, M_{Vij}(0), M_{Eij}(0)\right)\right].$$

The above decompositions are not unique. Daniel et al. (2015) showed as many as six different ways of decomposing the total treatment effect in the case of two concurrent mediators. We have chosen the decomposition that corresponds directly to the first set of research questions in the application study.

Causal Parameters

As explained in the previous section, we consider a theoretical population of sites, given our central interest in not only the prevalent causal mechanisms but also how the mechanisms may vary across local settings. Within each site, there is a population of eligible Job Corps applicants who did not have a high school diploma, GED, or vocational certificate at baseline. Taking the expectation of each individual-specific causal effect over the population of individuals at each site, we define the corresponding site-specific causal effect. Using S_{ij} to indicate the site membership of individual i at site j, we list the site-specific causal effects in the first column of Table 2. Taking the expectation and the variance of each site-specific causal effect over the population of sites, we define the corresponding population average and between-site variance of the causal effect, as listed in the second and third columns of Table 2. These causal parameters correspond to the research questions presented in the introduction section.

IDENTIFICATION OF THE CAUSAL PARAMETERS

As introduced in the second section, groups of youths in the study population had different probabilities of being selected into the research sample. Sampled individuals were then randomly assigned to the program or control group with different probabilities, depending on personal and site-level characteristics. In the longitudinal follow-ups, non-random attrition and selective non-response to questions measuring the mediators and the outcome may cause some groups to become over- or underrepresented and may lead to a systematic difference between the treatment group and the control group in the sample of respondents. Furthermore, in a multisite randomized trial, even if the treatment is randomized, mediators are usually generated in a natural process, which gives rise to mediator selection bias.

Table 2. The population average and between-site variance of site-specific effects.

	Site-specific effect	Population average	Between-site variance
ITT effect on the mediator M_V	$\beta_j^{(T.V)} = E[\beta_{ij}^{(T.V)} \mid S_{ij} = j]$	$\gamma^{(T.V)} = E[\beta_j^{(T.V)}]$	$\sigma^2_{T.V} = var(\beta_j^{(T.V)})$
ITT effect on the mediator M_E	$\beta_j^{(T.E)} = E[\beta_{ij}^{(T.E)} \mid S_{ij} = j]$	$\gamma^{(T.E)} = E[\beta_j^{(T.E)}]$	$\sigma^2_{T.E} = var(\beta_j^{(T.E)})$
ITT effect on the outcome Y	$\beta_j^{(T)} = E[\beta_{ij}^{(T)} \mid S_{ij} = j]$	$\gamma^{(T)} = E[\beta_j^{(T)}]$	$\sigma^2_{T} = var(\beta_j^{(T)})$
Indirect effect via M_V given $M_E(0)$	$\beta_j^{(I.V)}(0) = E[\beta_{ij}^{(I.V)}(0) \mid S_{ij} = j]$	$\gamma^{(I.V)}(0) = E[\beta_j^{(I.V)}(0)]$	$\sigma^2_{I.V(0)} = var(\beta_j^{(I.V)}(0))$
Indirect effect via M_E given $M_V(1)$	$\beta_j^{(I.E)}(1) = E[\beta_{ij}^{(I.E)}(1) \mid S_{ij} = j]$	$\gamma^{(I.E)}(1) = E[\beta_j^{(I.E)}(1)]$	$\sigma^2_{I.E(1)} = var(\beta_j^{(I.E)}(1))$
Direct effect	$\beta_j^{(D)}(0) = E[\beta_{ij}^{(D)}(0) \mid S_{ij} = j]$	$\gamma^{(D)}(0) = E[\beta_j^{(D)}(0)]$	$\sigma^2_{D(0)} = var(\beta_j^{(D)}(0))$
Indirect effect via M_V given $M_E(1)$	$\beta_j^{(I.V)}(1) = E[\beta_{ij}^{(I.V)}(1) \mid S_{ij} = j]$	$\gamma^{(I.V)}(1) = E[\beta_j^{(I.V)}(1)]$	$\sigma^2_{I.V(1)} = var(\beta_j^{(I.V)}(1))$
Indirect effect via M_E given $M_V(0)$	$\beta_j^{(I.E)}(0) = E[\beta_{ij}^{(I.E)}(0) \mid S_{ij} = j]$	$\gamma^{(I.E)}(0) = E[\beta_j^{(I.E)}(0)]$	$\sigma^2_{I.E(0)} = var(\beta_j^{(I.E)}(0))$
Interaction effect between M_V and M_E	$\beta_j^{(I.V \times E)} = E[\beta_{ij}^{(I.V \times E)} \mid S_{ij} = j]$	$\gamma^{(I.V \times E)} = E[\beta_j^{(I.V \times E)}]$	

To relate the counterfactual quantities to the observed data, we invoke a series of identification assumptions about the sampling mechanism, the treatment assignment mechanism, the response mechanism, and the mediator selection mechanism. All these assumptions share the notion of "selection on observables." Let $\mathbf{X} = \mathbf{X}_D \cup \mathbf{X}_T \cup \boldsymbol{X}_R \cup \mathbf{X}_V \cup \mathbf{X}_E$ be a set of pre-treatment covariates that predict the outcome and cannot be affected by the treatment, where $\mathbf{X}_D$ predicts sample selection, $\mathbf{X}_T$ predicts treatment selection, $\boldsymbol{X}_R$ predicts non-response selection, $\mathbf{X}_V$ predicts selection into vocational attainment, and $\mathbf{X}_E$ predicts selection into educational attainment. $\mathbf{X}_D, \mathbf{X}_T, \boldsymbol{X}_R, \mathbf{X}_V$, and $\mathbf{X}_E$ may overlap with one another. For example, gender is arguably an important element in each subset of covariates. Under the assumptions presented below, we are able to remove sampling bias, treatment selection bias, non-response bias, and mediator selection bias by applying a series of propensity score-based weights to the observed data, thereby identifying the causal parameters defined in Table 2.

Identification of the ITT Effects

To identify the population average and between-site variance of the ITT effects on the mediators and the outcome, we propose a series of assumptions about the sample selection, the treatment selection, and the non-response selection.

Assumption 1. *Strongly Ignorable Sampling Mechanism*

Given the observed pre-treatment covariates $\mathbf{x}_D$, sample selection is independent of the potential mediators and potential outcomes at each site.

$$\left\{Y_{ij}\left(t, m_V, m_E\right), M_{Vij}(t), M_{Eij}(t)\right\} \perp\!\!\!\perp D_{ij} | \mathbf{X}_{Dij} = \mathbf{x}_D, \quad S_{ij} = j,$$

for $t = 0, 1$, $m_V, m_E \in \mathcal{M}$ where $\mathcal{M}$ is the support for all possible mediator values, and $j = 1, \ldots, J$, where J is the total number of sites. D_{ij} is a binary indicator for whether individual i at site j was selected into the sample by NJCS researchers. Additionally, it is assumed that $0 < Pr(\, D_{ij} = 1 |\, \mathbf{X}_{Dij} = \mathbf{x}_D, \;\; S_{ij} = j) < 1$, also known as the positivity assumption or assumption about the "common support" or "overlap." The assumption states that each individual has a nonzero probability of being selected into the sample at a given site. Researchers may empirically detect a violation of this theoretical assumption when the range of the covariate distribution clearly differs between the treatment group and the control group. Assumption 1 holds in NJCS because the sampling probability is a known function of a set of observed pre-treatment covariates $\mathbf{X}_D$ including an applicant's date of random assignment, gender, and whether the applicant was likely to be assigned to a residential or non-residential program slot.

Assumption 2. *Strongly Ignorable Treatment Assignment*

Given the observed pre-treatment covariates $\mathbf{x}_T$, the treatment assignment is independent of the potential mediators and potential outcomes among the sampled individuals at each site.

$$\left\{Y_{ij}\left(t, m_V, m_E\right), M_{Vij}(t), M_{Eij}(t)\right\} \perp\!\!\!\perp T_{ij} | D_{ij} = 1, \mathbf{X}_{Tij} = \mathbf{x}_T, S_{ij} = j.$$

It is also assumed that $0 < Pr(T_{ij} = t \,|\, D_{ij} = 1, \mathbf{X}_{Tij} = \mathbf{x}_T, S_{ij} = j) < 1$. That is, each sampled individual has a nonzero treatment assignment probability at a given site. Assumption 2 also holds in NJCS: the treatment assignment probability depended on an applicant's date of random assignment and residential status, among other factors, which constitute $\mathbf{X}_T$.

Assumption 3. *Strongly Ignorable Non-Response*

Given the observed baseline covariates $\mathbf{x}_R$, the response status of a sampled individual in a treatment group is independent of the potential mediators and potential outcomes associated with the corresponding treatment condition at each site.

$$\left\{Y_{ij}\left(t, m_V, m_E\right), M_{Vij}(t), M_{Eij}(t)\right\} \perp\!\!\!\perp R_{ij} | T_{ij} = t, D_{ij} = 1, \qquad \mathbf{X}_{Rij} = \mathbf{x}_R, \qquad S_{ij} = j.$$

R_{ij} is a binary indicator for whether individual i at site j responded. We also assume that $0 < \Pr(\ R_{ij} = 1 |\ T_{ij} = t, D_{ij} = 1, \quad \mathbf{X}_{Rij} = \mathbf{x}_R, \quad S_{ij} = j) < 1$. That is, each sampled individual has a nonzero response probability under either treatment condition at a given site. Under the strongly ignorable non-response assumption, respondents and non-respondents in each treatment group at each site who share the same observed pre-treatment characteristics are expected to be comparable in their distributions of potential mediators and potential outcomes. This assumption holds if individuals with the same values of X_R were as if randomized to respond or not to respond—such as missing a letter or a phone call from the NJCS interviewers by chance. This assumption, however, is not guaranteed by the NJCS design. For example, if, among individuals with the same values of X_R, those whose residential address remained unchanged over the next four years were more likely to respond than those who changed address, and if, in comparison with the movers, the stayers in each treatment group had a higher predisposition of obtaining a credential and receiving relatively high earnings, omitting residential mobility as a potential confounder would likely result in a potential violation of Assumption 3.

Under Assumptions 1, 2, and 3, the site-specific mean of each potential mediator and that of the potential outcome under treatment t can be respectively identified by the weighted mean of each observed mediator and that of the observed outcome among the respondents assigned to treatment group t at site j, as follows:

$$E\left[M_{Vij}(t) | S_{ij} = j\right] = E\left[W_{ITTij} M_{Vij} | R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j\right],$$

$$E\left[M_{Eij}(t) | S_{ij} = j\right] = E\left[W_{ITTij} M_{Eij} | R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j\right],$$

$$E\left[Y_{ij}\left(t, M_{Vij}(t), M_{Eij}(t)\right) | S_{ij} = j\right] = E\left[W_{ITTij} Y_{ij} | R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j\right].$$

Here $W_{ITTij} = W_{Dij} W_{Tij} W_{Rij}$, where

$$W_{Dij} = \frac{Pr\left(D_{ij} = 1 | S_{ij} = j\right)}{Pr\left(D_{ij} = 1 | \mathbf{X}_{Dij} = \mathbf{x}_D, S_{ij} = j\right)},$$

$$W_{Tij} = \frac{Pr\left(T_{ij} = t | D_{ij} = 1, S_{ij} = j\right)}{Pr\left(T_{ij} = t | \mathbf{X}_{Tij} = \mathbf{x}_T, D_{ij} = 1, S_{ij} = j\right)},$$

$$W_{Rij} = \frac{Pr\left(R_{ij} = 1 | T_{ij} = t, D_{ij} = 1, S_{ij} = j\right)}{Pr\left(R_{ij} = 1 | \mathbf{X}_{Rij} = \mathbf{x}_R, T_{ij} = t, D_{ij} = 1, S_{ij} = j\right)}. \tag{1}$$

W_{Dij}, W_{Tij}, and W_{Rij} are used to adjust for sample selection, treatment selection, and non-response selection, respectively. To be specific, W_{Dij} is the sample and survey weight that will equalize the sampling probability of all sampled individuals at a given site; W_{Tij} is the Inverse Probability of Treatment Weight (IPTW) (Horvitz & Thompson, 1952; Robins, Hernán, & Brumback, 2000) that transforms a sampled individual's treatment assignment probability to be equal to the average treatment assignment probability in the sample at his or her site; the non-response weight,

W_{Rij}, transforms a respondent's response probability to be equal to the average response rate in the same treatment group sample at his or her site.

By applying W_{ITTij}, we expect that, when the identification assumptions hold, the respondents in each treatment group will have the same composition of pre-treatment characteristics as that in the entire population at each site. This is because weighting will balance the joint distribution of the observed baseline covariates between the sampled and the non-sampled, between the treated and the untreated, and between the responders and the non-responders in each treatment group. The weighted mean difference in each mediator (or the outcome) between the program group and the control group at each site identifies the site-specific ITT effect of the treatment on the mediator (or the outcome). The population average and the between-site variance of each of these ITT effects can be identified correspondingly.

Identification of the Mediation-Related Effects

Identifying the mediation-related effects is more challenging for several reasons. First of all, these effects involve counterfactual outcomes that are not directly observed for any individual. Second, the mediator-outcome relationships are likely confounded by pre-treatment and post-treatment differences between individuals in different mediator categories. And third, the two concurrent mediators may act as confounders for each other. To identify these effects, we need two additional assumptions.

Assumption 4. *Strongly Ignorable Mediator Selection Mechanism*

Among sample respondents at each site, given the observed pre-treatment covariates $\mathbf{x}_V$, whether one obtains a vocational credential under either treatment condition is independent of the potential outcomes; similarly, given the observed pre-treatment covariates $\mathbf{x}_E$, whether one obtains an education credential under either treatment condition is independent of the potential outcomes.

$$Y_{ij}\left(t, m_V, m_E\right) \perp\!\!\!\perp \left\{M_{Vij}(t), M_{Vij}\left(t'\right)\right\} | R_{ij} = 1, \quad T_{ij} = t, \quad D_{ij} = 1, \mathbf{X}_{Vij} = \mathbf{x}_V, \quad S_{ij} = j,$$

$$Y_{ij}\left(t, m_V, m_E\right) \perp\!\!\!\perp \left\{M_{Eij}(t), M_{Eij}\left(t'\right)\right\} | R_{ij} = 1, T_{ij} = t, \quad D_{ij} = 1, \mathbf{X}_{Eij} = \mathbf{x}_E, \quad S_{ij} = j,$$

for all possible values of t, t', m_V, and m_E, where $t \neq t'$. It is also assumed that $0 < Pr(\, M_{Vij} = m_V | R_{ij} = 1, T_{ij} = t, D_{ij} = 1, \mathbf{X}_{Vij} = \mathbf{x}_V, S_{ij} = j) < 1$ and $0 < Pr(\, M_{Eij} = m_E | R_{ij} = 1, T_{ij} = t, D_{ij} = 1, \mathbf{X}_{Eij} = \mathbf{x}_E, S_{ij} = j) < 1$. That is, within levels of the observed pre-treatment covariates, every sample respondent at site j had a nonzero probability of obtaining (or not obtaining) a vocational (or an education) credential. The assumption states that, among respondents with the same pre-treatment characteristics, each mediator can be viewed as if randomized within the same treatment group or across treatment groups at each site. In other words, this may be viewed as a cross-randomized design or a factorial randomized design for the two mediators within levels of the pre-treatment covariates. This assumption is seemingly reasonable because, despite a Job Corps applicant's initial plan of pursuing education and training, with everything else being equal, unanticipated events such as receiving an attractive job offer, being caught and expelled from the program for a behavioral offense, or having to leave the program due to a family member's health problem could throw one off track; these or other random events could also occur to control group members who might have potential access to alternative education or training programs. Assumption 4 would be violated if a mediator-outcome relationship was confounded by omitted pre-treatment characteristics such as peer network prior to randomization or by post-treatment covariates such as a student's compliance status.

Assumption 5. *Conditional Independence Between the Potential Mediators*

Among sample respondents at each site and given the observed pre-treatment covariates $\mathbf{x}_V$ and $\mathbf{x}_E$, whether one would obtain a vocational credential under one treatment condition is independent of whether one would obtain an education credential under the same or the alternative treatment condition. For example, conditional on the observed covariates, one's vocational attainment when assigned to Job Corps is independent of his or her educational attainment when assigned to Job Corps; it is also independent of his or her educational attainment when assigned to the control condition.

$$M_{Vij}(t) \perp\!\!\!\perp M_{Eij}\left(t'\right) | R_{ij} = 1, T_{ij} = t, \quad D_{ij} = 1, \mathbf{X}_{Vij} = \mathbf{x}_V, \quad \mathbf{X}_{Eij} = \mathbf{x}_E, S_{ij} = j$$

for $t,\ t' = 0, 1$. This assumption is necessary for distinguishing the indirect effect transmitted through M_V from that transmitted through M_E. The assumption might seem plausible given that obtaining an education credential does not guarantee or preclude one from obtaining a vocational credential or vice versa. The assumption would be violated if covariates that contribute to a correlation between the two mediators, such as motivation to acquire new skills, were omitted.

We will evaluate each of these identification assumptions and will assess the consequence of a potential violation of Assumptions 3, 4, or 5 through a sensitivity analysis, which we discuss in a later section.

Theorem

Under Assumptions 1 through 5, the site-specific mean potential outcome associated with treatment condition t while one mediator or both mediators take values associated with the counterfactual condition can be identified. It will be equal to the average of the observed outcome among the sample respondents assigned to treatment group t at the site weighted by the product of the ITT weight and two additional weights constructed for the two mediators:

$$\begin{aligned} &E\left[\left[Y_{ij}\left(t, M_{Vij}\left(t'\right), M_{Eij}\left(t''\right)\right) | S_{ij} = j\right]\right] \\ &\quad = E\left[W_{ITTij} W_{Vt'ij} W_{Et''ij} Y_{ij} | R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j\right], \end{aligned}$$

where $t' \neq t$ or $t'' \neq t$, and for $m_V, m_E \in \mathcal{M}$,

$$W_{Vt'ij} = \frac{Pr\left(M_{Vij} = m_V | \mathbf{X}_{Vij} = \mathbf{x}_V, R_{ij} = 1, T_{ij} = t', D_{ij} = 1, S_{ij} = j\right)}{Pr\left(M_{Vij} = m_V | \mathbf{X}_{Vij} = \mathbf{x}_V, R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j\right)}, \tag{2}$$

$$W_{Et''ij} = \frac{Pr\left(M_{Eij} = m_E | \mathbf{X}_{Eij} = \mathbf{x}_E, R_{ij} = 1, T_{ij} = t'', D_{ij} = 1, S_{ij} = j\right)}{Pr\left(M_{Eij} = m_E | \mathbf{X}_{Eij} = \mathbf{x}_E, R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j\right)}. \tag{3}$$

If $t' = t$, then $W_{Vt'ij} = 1$; if $t'' = t$, then $W_{Et''ij} = 1$. Otherwise, $W_{Vt'ij}$ and $W_{Et''ij}$ are extensions of the RMPW weight originally developed by Hong (2010, 2015) for a single mediator and for two concurrent mediators. Each weight is a ratio of a sample respondent's conditional probability of displaying a given mediator value under the counterfactual treatment condition to that under the actual treatment condition. For respondents who were assigned to treatment group t and displayed pre-treatment characteristics $\mathbf{x}_V$ at site j, W_{Vij} transforms their distribution of vocational attainment to resemble that of their counterparts in the counterfactual group t' if $t' \neq t$; W_{Eij} transforms their distribution of educational attainment to resemble that of their counterparts in the counterfactual group t'' if $t'' \neq t$. Hence, by applying

Table 3. Identification of the site-specific causal effects.

Site-specific effect	Identification result	Assumptions
ITT effect on the mediator M_V, $\beta_j^{(T.V)}$	$\upsilon_{V1j} - \upsilon_{V0j}$	Assumptions 1–3
ITT effect on the mediator M_E, $\beta_j^{(T.E)}$	$\upsilon_{E1j} - \upsilon_{E0j}$	
ITT effect on the outcome Y, $\beta_j^{(T)}$	$\mu_{1j} - \mu_{0j}$	
Indirect effect via M_V given $M_E(0)$, $\beta_j^{(I.V)}(0)$	$\mu_{1.1.0j} - \mu_{1.0.0j}$	Assumptions 1–5
Indirect effect via M_E given $M_V(1)$, $\beta_j^{(I.E)}(1)$	$\mu_{1j} - \mu_{1.1.0j}$	
Direct effect, $\beta_j^{(D)}(0)$	$\mu_{1.0.0j} - \mu_{0j}$	
Indirect effect via M_V given $M_E(1)$, $\beta_j^{(I.V)}(1)$	$\mu_{1j} - \mu_{1.0.1j}$	
Indirect effect via M_E given $M_V(0)$, $\beta_j^{(I.E)}(0)$	$\mu_{1.0.1j} - \mu_{1.0.0j}$	
Interaction effect between M_V and M_E, $\beta_j^{(I.V \times E)}$	$(\mu_{1j} - \mu_{1.1.0j}) - (\mu_{1.0.1j} - \mu_{1.0.0j})$	

the product of the ITT weight and the RMPW weights to the sample respondents in treatment group t at each site, we are able to identify the site-specific population average counterfactual outcomes, including $E[Y_{ij}(1, M_{Vij}(1), M_{Eij}(0))|S_{ij} = j]$, $E[Y_{ij}(1, M_{Vij}(0), M_{Eij}(0))|S_{ij} = j]$, and $E[Y_{ij}(1, M_{Vij}(0), M_{Eij}(1))|S_{ij} = j]$. Appendix A[1] presents a proof of the Theorem.

To simplify the notation, we let $\upsilon_{Vtj} = E[W_{ITTij}M_{Vij}|\, R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j]$ denote the weighted average of the observed mediator M_V in treatment group t at site j, which identifies $E[M_{Vij}(t)|S_{ij} = j]$; let $\upsilon_{Etj} = E[W_{ITTij}M_{Eij}|\, R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j]$ denote the weighted average of the observed mediator M_E in treatment group t at site j, which identifies $E[M_{Eij}(t)|S_{ij} = j]$; let $\mu_{tj} = E[W_{ITTij}Y_{ij}|\ R_{ij} = 1, T_{ij} = t,\ D_{ij} = 1,\ S_{ij} = j]$ denote the weighted average of the observed outcome in treatment group t at site j, which identifies $E[Y_{ij}(t, M_{Vij}(t), M_{Eij}(t))|S_{ij} = j]$; and let $\mu_{t.t'.t''j} = E[W_{ITTij}W_{Vt'ij}W_{Et''ij}Y_{ij}|R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j]$ denote the average of the observed outcome further weighted by the RMPW weights in treatment group t at site j, which identifies $E[Y_{ij}(t, M_{Vij}(t'), M_{Eij}(t''))|S_{ij} = j]$ for $t, t', t'' = 0, 1$. Table 3 summarizes the identification results. The site-specific causal effects listed in the first column are then identified by the weighted mean contrasts of the observed outcome at each site as listed in the second column, under the assumptions that are listed in the third column. It is then straightforward to identify the population parameters characterizing the joint distribution of the site-specific effects as defined in Table 2.

ESTIMATION AND INFERENCE OF THE CAUSAL EFFECTS

As shown in the previous section, the identification of the population average and between-site variance of the causal effects partly relies on the product of the sample weight W_{Dij} and the IPTW weight W_{Tij}. This product is given by the NJCS design. In addition, the identification also relies on the non-response weight, W_{Rij}, and the RMPW weights, $W_{Vt'ij}$ and $W_{Et''ij}$, all of which need to be estimated. Hence, the estimation of the causal parameters involves two major steps. Step 1 estimates the non-response weight and the RMPW weights through a propensity scoring approach. Step 2 estimates the site-specific mean of each potential outcome

[1] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

(or potential mediator) through obtaining the weighted mean observed outcome (or observed mediator) of sample respondents in a given treatment group at each site. However, the uncertainty of the estimated weights is usually ignored in causal inference studies that employ propensity score-based weighting strategies in multilevel settings (Leite et al., 2015). Consequently, the estimated standard errors of the causal effect estimates are biased and conclusions of statistical inference can even be reversed. A bootstrap procedure has been recommended as a typical solution (Goldstein, 2011). Yet bootstrapping is computationally intensive. In addition, standard errors of the population average causal effect estimates tend to be overestimated by bootstrap when the site size is relatively small (Qin & Hong, 2017).

To incorporate the sampling variability of the estimated weights, we extend an estimation procedure proposed by Qin et al. (2019) based on the GMM framework. While the two estimation steps are presented sequentially for expository purposes, in practice, we estimate them jointly through GMM. This joint estimation allows us to consistently estimate standard errors that account for the estimation error in the weights (Newey, 1984). When all the identification assumptions hold, the estimates are consistent and generalizable to the study population of Job Corps centers. The proposed method may also be appropriate in other settings where propensity score-based weighting strategies are used in multilevel settings. Appendix B[2] provides the mathematical details.

DATA ANALYSIS

To account for selective non-response and for selection into different levels of education and vocational training attainment while ensuring efficiency, we select theoretically important baseline predictors of earnings that are generally also important predictors of the mediators and in many cases important predictors of the response indicator as well. They include demographic characteristics such as gender, age, race/ethnicity, native language, fertility and living arrangements, education and training experiences, reasons for leaving school, employment and earnings, physical and mental health, drug use and treatment, public assistance receipt, and motivation and support for joining Job Corps prior to the random assignment. We generate a missing indicator for the missing cases in each covariate and discretize the continuous covariates to reduce the risk of model misspecification. The analysis makes adjustment for up to 52 variables. See Appendix C for the full list. Our selection of covariates is based on the rationale that adjusting for variables related to the mediator or the response indicator but not to the outcome may reduce efficiency without further reducing bias (De Luna et al., 2011; Patrick et al., 2011), while adjusting for variables associated with the outcome but not with the mediator or the response indicator may improve estimation efficiency without increasing bias (Brookhart et al., 2006). Variable selection is necessary also for the purpose of avoiding model overfitting.

To estimate a sampled respondent's propensity to respond to the survey, the propensity to obtain an education credential, and the propensity to obtain a vocational credential under each treatment condition as a function of the observed covariates, we specify a random-effects logistic regression model in each case. The site effects are conceptualized as random rather than fixed in the interest of generalizing to a super-population of sites (Raudenbush & Schwartz, 2020). The estimated non-response weights and RMPW weights range from 0.02 to 4.12. The product

[2] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Table 4. Attainment of education and vocational credentials (sample weights and non-response weights applied).

	Obtained neither	Obtained education credential only	Obtained vocational credential only	Obtained both
Assigned to Control (n = 2,646)	76%	19%	3%	2%
Assigned to Job Corps (n = 3,968)	57%	25%	7%	11%
Enrolled in Job Corps (n = 2,916, 74%)[1]	50%	26%	9%	15%
Did not enroll (n = 993, 25%)[1]	67%	26%	3%	4%

Notes: [1] 59 people assigned to Job Corps had missing information on whether or not they enrolled.

of the sample weight, non-response weight, and RMPW does not contain extreme values that could have resulted in unstable estimation results (Lee, Lessler, & Stuart, 2011).

Our analytic strategy does not require specifying an outcome model. Yet possible misspecifications of the propensity score models due to omissions of confounding covariates or due to misspecified functional forms would undermine the effort at reducing selection bias (Rosenbaum & Rubin, 1984). This can be detected at least partly when there is a lack of balance in the distribution of the observed covariates between the treatment groups after we have made the weighting-based adjustment for the selected set of covariates. Hence, we adopt a weighting-based balance checking procedure to assess if any observed covariate remains imbalanced after weighting. According to the results of balance checking as shown in Appendix D, within each treatment group and across all the sites, the non-response weighting has greatly improved the balance in the distribution of the observed covariates between the respondents and the non-respondents; and the RMPW weighting has improved the balance in the distribution of the observed covariates between different mediator categories. Additionally, to assess potential hidden bias due to omitted confounders, we conduct a series of weighting-based sensitivity analyses and report the results in a later section.

ITT Effect of Job Corps on Each Mediator

As shown in Table 4, Job Corps improved educational attainment and vocational attainment among disadvantaged youth. During the 30 months after randomization, 36 percent of the individuals assigned to Job Corps obtained an education credential; in contrast, only 21 percent of those assigned to the control group had the same level of attainment. This is a 70 percent increase. The proportion difference is 0.15 on average and is statistically significant (SE = 0.017, $p < 0.001$); the between-site standard deviation of this proportion difference is estimated to be 0.108 ($p = 0.01$). Meanwhile, 18 percent of those assigned to Job Corps obtained a vocational credential, and only 5 percent of those assigned to the control group did the same. The result indicates an increase by almost a factor of three. The proportion difference on average is 0.13 and is also statistically significant (SE = 0.010, $p < 0.001$); the standard deviation of this proportion difference across sites is 0.058 ($p < 0.001$).

ITT Effect of Job Corps on Earnings

Forty-eight months after randomization, the Job Corps program generated an average impact on earnings that is statistically significant and substantively

important. The population average weekly earnings among individuals assigned to Job Corps is estimated to be \$21.74 (SE = 5.94, $p < 0.001$) higher than their counterparts assigned to the control group. Given that a typical person in the control group earned \$185.83 (in 1995 dollars) weekly, this result indicates a nearly 12 percent increase in earnings and amounts to about 13 percent of a standard deviation of the outcome.[3] Yet the impact on earnings varied significantly across sites. A site-by-site analysis shows that, for each causal effect, the distribution of the estimated site-specific effects is approximately normal. Thus, we assume a normal distribution of the site-specific effects in the population of sites. Under this assumption, the impact on earnings would range from -\$29.96 to \$73.44 in 95 percent of the sites. This result indicates that even though Job Corps significantly improved earnings on average, the impact was negligible or even negative in some of the sites. There are several reasons for the substantial between-site heterogeneity of the Job Corps impact. First, the Job Corps program might generate different impacts for different subgroups of participants while the relative proportions of these subgroups might vary across the sites. Second, some of the Job Corps centers might be dysfunctional or simply did not have adequate resources to serve the needs of a concentration of highly vulnerable youths, who might otherwise receive better services from alternative programs under the control condition. This would likely be reflected in between-site variation in the participation rate among those assigned to Job Corps relative to the participation rate of their control counterparts in alternative programs. Third, the vocational training provided by some of the Job Corps centers might not match the labor market demand in their local areas.

Population Average Mediation Mechanism

We further decompose the total Job Corps impact on earnings into the direct and indirect effects. The analytic results, as summarized in Table 5 , reveal the mediation mechanism characteristic of the educational process central to Job Corps' theoretical rationale.

The population average indirect effect operating through vocational training without affecting education is estimated to be \$3.15 (SE = 1.63, $p = 0.052$), which amounts to about 2 percent of a standard deviation of the outcome and about 14 percent of the total ITT effect. With vocational training already improved, the population average indirect effect via an improvement in education is estimated to be \$7.52 (SE = 1.63, $p < 0.001$), about 4 percent of a standard deviation of the outcome and about a third of the total ITT effect. Vocational training and general education together mediated about half of the total program impact on earnings. The mediating role of general education is seemingly greater than that for vocational training.

The indirect effect of the program on earnings operating through vocational training can instead be defined with education already improved. This indirect effect is estimated to be \$4.05 rather than \$3.15 and is statistically significant (SE = 1.72, $p = 0.019$). The estimated interaction effect between the two mediators, however, is not statistically significant (estimated interaction effect = 0.90, SE = 0.70, $p = 0.20$). Hence we infer that vocational training and general education were complementary

[3] Schochet et al. (2006) originally reported an ITT effect estimate close to \$16. The discrepancy is mainly due to the difference in the target of inference. Schochet et al. (2006) chose the population of individuals, while we are interested in the generalizability across sites, as explained in the second section. In addition, we defined non-respondents (those missing a site identification number, the mediator, or the outcome) differently than Schochet et al. (2006) (those missing the outcome only).

Table 5. Decomposition of the Job Corps impact on earnings (1995 dollars).

	Population average effect				
	Estimate (dollars)[1]	Effect size[2]	*p*-value	Between-site standard deviation (dollars)	95% plausible value range of site-specific effects (dollars)[3]
ITT effect on the outcome	21.743 (5.944)	0.125	<0.001	26.379	[-29.960, 73.446]
Indirect effect via M_V given $M_E(0)$	3.152 (1.625)	0.018	0.052	8.690	[-13.880, 20.184]
Indirect effect via M_E given $M_V(1)$	7.518 (1.630)	0.043	<0.001	2.666	[2.293, 12.743]
Direct effect	11.073 (6.322)	0.064	0.080	29.288	[-46.331, 68.474]
Indirect effect via M_V given $M_E(1)$	4.050 (1.723)	0.023	0.019	9.175	[-13.933, 22.033]
Indirect effect via M_E given $M_V(0)$	6.620 (1.761)	0.038	<0.001	6.498	[-6.116, 19.356]
Interaction effect between M_V and M_E	0.898 (0.700)	0.005	0.200	0.940	[-0.944, 2.740]

Notes: [1]The standard error of the point estimate of each population average effect is provided in parentheses.
[2]The effect size of each population average effect estimate is calculated by dividing the point estimate by the weighted average within-site standard deviation of the outcome in the control group.
[3]The bounds for the 95 percent plausible value range of the site-specific effects are 1.96 times the between-site standard deviation estimate, away from the population average effect estimate, under the assumption that the site-specific effects are normally distributed.

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

to each other but not necessarily mutually reinforcing when transmitting the Job Corps impact on earnings.

The population average direct effect, operating through other mechanisms without affecting vocational training and education, is estimated to be \$11.07 (SE = 6.32, p = 0.08). This effect amounts to about 6 percent of a standard deviation of the outcome and accounts for the other half of the total ITT effect. This evidence indicates that the supplementary services and other unspecified intermediate process that distinguished the Job Corps experiences from the control group experiences played a role at least as important as education and training in promoting economic well-being.

Between-Site Variance of the Mediation Mechanism

To explain the between-site heterogeneity in the total Job Corps impact on earnings, we further investigate how the causal mediation mechanism varies across sites. The estimated between-site standard deviation of the Job Corps impact mediated by vocational training without affecting education is \$8.69, and that of the Job Corps impact mediated by general education beyond affecting vocational training is \$2.67. In contrast, the remaining bulk of the between-site heterogeneity in the Job Corps impact, accounted for by the direct effect, is estimated to be as large as \$29.29. The 95 percent plausible range of the site-specific direct effect is from -\$46.34 to \$68.48. According to these results, the variation in the Job Corps impact across the sites is mainly explained by the heterogeneity in the direct effect. As explained earlier, the direct effect captures unspecified pathways including differences between the treatment groups in experiences with supplementary services and other intermediate process. The finding may reflect local discretion in the provision of supplementary services and uneven quality and quantity of such services across the Job Corps centers, in contrast with education and vocational training curricula that were standardized by the national Job Corps office and regional offices (Johnson, Dugan, & Gritz, 2000; Schochet, Burghardt, & McConnell, 2008).

SENSITIVITY ANALYSIS

The causal effects are identified under the assumptions that, within levels of the observed pre-treatment covariates at each site, the sampling mechanism and treatment assignment mechanism as well as the response mechanism and mediator value assignment under each treatment are all strongly ignorable (Assumptions 1 through 4), and that the two mediators are conditionally independent within and across the treatment conditions (Assumption 5). While Assumptions 1 and 2 are guaranteed by the research design, omitting pre-treatment or post-treatment confounders or overlooking between-site variation in the selection mechanisms would violate Assumption 3 (strongly ignorable non-response) or Assumption 4 (strongly ignorable mediator value assignment); potential violations of Assumption 5 are also likely. For example, we find that the respondents and non-respondents in the Job Corps group differ in the distributions of M_E and Y even after we have controlled for the set of observed covariates. This indicates possible omissions of confounders in our current adjustment for non-response-related bias. It is thus imperative to evaluate the sensitivity of the analytic results to these potential violations.

A sensitivity analysis helps determine whether the removal of a hidden bias would lead to a qualitative change in a causal conclusion (Ichino et al., 2008; Imai & Yamamoto, 2013; Rosenbaum, 2017). Following this literature, we consider primarily a change in effect size. Hong and colleagues (Hong, Qin, & Yang, 2018; Qin et al., 2019) extended from single-site to multisite mediation analysis

a weighting-based sensitivity analysis approach for assessing the potential consequences of omitted pre-treatment confounders. This strategy quantifies the amount of bias due to the omission by comparing an initial weight with a new weight that adjusts for the omissions without reliance on outcome model specifications. Its extension to multisite mediation analysis involving two concurrent mediators is mostly straightforward. Here we give special consideration to the potential consequences of omitting a post-treatment measure of compliance, an omission that would likely violate Assumption 4. We additionally assess the potential consequences of violating Assumption 5 when the two concurrent mediators are not conditionally independent. Appendix E[4] provides mathematical details for assessing the bias associated with each type of potential violation as well as details of the results.

Sensitivity to Potential Omissions of Pre-Treatment Confounders

Under the supposition that the confounding impact of an unmeasured pre-treatment covariate is likely comparable to that of some of the observed covariates on the basis of scientific reasoning or past empirical evidence, we obtain a distribution of plausible values of the effect size of bias for each estimated causal effect. These distributions are shown in Figures E1 through E7 in Appendix E.

For the ITT effect, the initial estimate of the effect size is 0.13. Figure E1 shows the distribution of plausible bias values if any one of the observed pre-treatment covariates had been omitted. The plausible bias values are predominantly negative; and the 1st quartile, the mean, and the 3rd quartile of the distribution are -0.12, -0.09, and 0.01, respectively. Removing a plausible hidden bias at the median level would lead to an increase of the estimated ITT effect by 75 percent. One may argue that an individual's past trajectory of deviant behaviors is likely an omitted confounder even after adjustment has been made for all the observed pre-treatment covariates. Should such an omission contribute a positive bias as large as 0.13 in effect size, removing the bias would lead to a point estimate of the ITT effect equal to zero. According to the empirical distribution in Figure E1, however, a positive bias of such a magnitude is plausible but not very probable. Hence, we may conclude that the initial estimate of the ITT effect is likely an underestimate rather than an overestimate of the benefit of Job Corps on earnings. This tentative conclusion could be refuted if further evidence becomes available and suggests a large positive bias associated with the omission of past deviant behaviors.

The plausible bias values in the estimation of the indirect effects and the direct effect are mostly close to zero; positive bias values and negative bias values tend to be evenly distributed. However, the initial estimates of the indirect effects via $\boldsymbol{M_V}$ and the interaction effect are small in effect size. A positive hidden bias equal to the 3rd quartile of the respective distributions, once removed, would turn the indirect effects via $\boldsymbol{M_V}$ to zero and would result in a negative point estimate of the interaction effect. Therefore, it seems that these initial estimates are likely sensitive to the omission of confounders that would contribute a positive bias. In contrast, the indirect effects via $\boldsymbol{M_E}$ and the direct effect are not as sensitive to positive plausible values of hidden bias; while it is noteworthy that removing a negative bias equal to the 1st quartile of the distribution would increase the point estimates of the indirect effects via $\boldsymbol{M_E}$ by 75 percent and would increase the point estimate of the direct effect by almost 80 percent.

[4] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Table 6. Sensitivity to the violation of Assumption 5 (effect size).

	Original Estimate	Bias if the Undifferentiated Part is via M_V	Bias if the Undifferentiated Part is via M_E
ITT effect on the outcome	0.125	–	–
Indirect effect via M_V given $M_E(0)$	0.018	0.005	−0.019
Indirect effect via M_E given $M_V(1)$	0.043	0.001	0.025
Direct effect	0.064	−0.006	−0.006
Indirect effect via M_V given $M_E(1)$	0.023	0.013	−0.028
Indirect effect via M_E given $M_V(0)$	0.038	−0.007	0.034
Interaction effect between M_V and M_E	0.005	0.008	−0.009

Sensitivity to the Omission of a Post-Treatment Confounder

During the three years after the randomized treatment assignment, individuals in the control group were barred from attending Job Corps; however, 30 percent of those assigned to Job Corps did not ever attend the program. After we have adjusted for the observed pre-treatment covariates, if the compliers and the non-compliers in Job Corps still differ in their predisposition for obtaining an education or training credential and for receiving relatively high earnings, then compliance behavior is a post-treatment covariate that may confound the mediator-outcome relationships.

Let $Z_{ij}(1)$ take value one if individual i assigned to Job Corps in site j actually attended Job Corps and zero otherwise. Should the individual have been assigned to the control condition instead, we would have that $Z_{ij}\ (0) = 0$ given the research design. As shown in Appendix E, additional adjustment for $Z(1)$ could be made in the propensity score model for M_V and that for M_E under the experimental condition. The new weights for individual i in site j are

$$W^*_{Vij} = \frac{pr\left(M_{Vij} = m_V | \mathbf{X}_{Vij} = \mathbf{x}_V, R_{ij} = 1, T_{ij} = 0, D_{ij} = 1, S_{ij} = j\right)}{pr\left(M_{Vij} = m_V | \mathbf{X}_{Vij} = \mathbf{x}_V, R_{ij} = 1, T_{ij} = 1, D_{ij} = 1, Z_{ij}(1) = z, S_{ij} = j\right)};$$

$$W^*_{Eij} = \frac{pr\left(M_{Eij} = m_E | \mathbf{X}_{Eij} = \mathbf{x}_E, R_{ij} = 1, T_{ij} = 0, D_{ij} = 1, S_{ij} = j\right)}{pr\left(M_{Eij} = m_E | \mathbf{X}_{Eij} = \mathbf{x}_E, R_{ij} = 1, T_{ij} = 1, D_{ij} = 1, Z_{ij}(1) = z, S_{ij} = j\right)}.$$

Due to the omission of the compliance measure in the initial analysis, the bias in identifying the indirect effect transmitted through vocational training is

$$E\left[W_{ITT}\left(W_E - W^*_E\right)Y | R = 1, T = 1, D\ = 1\right] - E\left[W_{ITT}\left(W_V W_E - W^*_V W^*_E\right)Y | R = 1, T = 1, D = 1\right];$$

the bias in identifying the indirect effect transmitted through education is

$$E\left[W_{ITT}\left(W^*_E - W_E\right)Y | R = 1, T = 1, D = 1\right];$$

and the bias in identifying the direct effect is

$$E\left[W_{ITT}\left(W_V W_E - W^*_V W^*_E\right)Y | R = 1, T = 1, D\ = 1\right].$$

A comparison of the estimated average causal effects and their between-site variances before and after adjusting for compliance reveals minimal discrepancies, as

shown in Table E1, which indicates that the initial results may not be sensitive to the omission of compliance behavior as a post-treatment confounder.

Sensitivity to Potential Violations of Assumption 5

Even when Assumptions 1 through 4 hold, omitting certain pre-treatment or post-treatment covariates—such as one's motivation to acquire new skills—would likely leave vocational training attainment and general education attainment correlated, which would violate Assumption 5. This can be tested by fitting a multilevel logistic model in each treatment group, regressing one mediator on the other while conditioning on the observed pre-treatment covariates. A large or statistically significant conditional association between the two mediators would indicate that the indirect effect via M_V and that via M_E are entangled. To address this problem, we decompose the total indirect effect transmitted via M_V or M_E into an indirect effect solely via M_V, an indirect effect solely via M_E, and the remaining indirect effect undifferentiated between M_V and M_E. This alternative decomposition allows us to assess the potential bias that may arise due to the violation of Assumption 5.

Indirect Effects Transmitted Solely via M_V

To disentangle the indirect effect via M_V and that via M_E, we represent $M_V(t)$ for $t = 0, 1$ as the sum of two orthogonal variables,

$$M_V\ (t) = M'_V\ (t) + M^*_V(t).$$

Here $M'_V\ (t) = f(M_E(t))$ is strictly determined by $M_E(t)$ while $M^*_V(t)$ is independent of $M_E(t)$ under treatment t. Regressing M_V on M_E in a multilevel linear probability model for respondents in treatment group t conditioning on the observed pre-treatment covariates, we obtain a residual that is orthogonal to M_E at each site. We may reason that covariates that are common causes of $M_V(t)$ and $M_E(t)$ may also predict both $M_V(t)$ and $M_E(t')$, where $t' \neq t$. Such covariates are expected to contribute to a conditional association between $M_V(t)$ and $M_E(t')$ as well as a conditional association between $M_V(t)$ and $M_E(t)$. Therefore, when $M^*_V(t)$ is conditionally independent of $M_E(t)$, it seems highly plausible that $M^*_V(t)$ and $M_E(t')$ are conditionally independent as well. Using M^*_V in replacement of M_V in the mediation analysis removes the confounding by M_E and captures the part of the indirect effect that is transmitted solely through M_V. Based on the Theorem stated earlier, we obtain the following identification result for $t, t',\ \ t'' = 0,\ 1$,

$$\begin{aligned} &E[Y_{ij}(t, M^*_{Vij}(t'), M_{Eij}(t''))|S_{ij} = j] \\ &\quad = E[W_{ITTij} W_{V^*t'ij} W_{Et''ij} Y_{ij} | R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j]. \end{aligned}$$

Because $M^*_V(t)$ is continuous, $W_{V^*t'ij}$ is a ratio of the conditional densities of $M^*_V(t)$, which can be estimated by replacing multilevel logistic regressions with multilevel linear regressions.

Indirect Effects Transmitted Solely via M_E

In parallel, to remove the confounding effect of M_V in identifying the indirect effects solely via M_E, we replace M_E with M^*_E, where M^*_E is the residual obtained from analyzing a multilevel linear probability model regressing M_E on M_V and the observed pre-treatment covariates for respondents under treatment t.

Total Indirect Effect Transmitted via M_V or M_E

The difference between the total effect and the direct effect is the total indirect effect transmitted via M_V or M_E. This causal effect can be identified without invoking Assumption 5. Following Appendix A,[5] it is straightforward to prove that, under Assumptions 1 through 4,

$$E\left[Y_{ij}\left(1, M_{Vij}(0), M_{Eij}(0)\right)|S_{ij}=j\right] = E\left[W_{ITTij}W_{VEij}Y_{ij}|R_{ij}=1, T_{ij}=1, D_{ij}=1, S_{ij}=j\right],$$

$$W_{VEij} = \frac{pr\left(M_{Vij}=m_V, M_{Eij}=m_E|\mathbf{X}_{Vij}=\mathbf{x}_V, \mathbf{X}_{Eij}=\mathbf{x}_E, R_{ij}=1, T_{ij}=0, D_{ij}=1, S_{ij}=j\right)}{pr\left(M_{Vij}=m_V, M_{Eij}=m_E|\mathbf{X}_{Vij}=\mathbf{x}_V, \mathbf{X}_{Eij}=\mathbf{x}_E, R_{ij}=1, T_{ij}=1, D_{ij}=1, S_{ij}=j\right)}. \quad (4)$$

We can estimate the weight by modeling the joint conditional probability of M_V and M_E under each treatment condition. Because M_V and M_E are both discrete, the numerator and the denominator of the weight can each be estimated through a multinomial logistic regression.

Indirect Effect Undifferentiated Between M_V and M_E

When M_V and M_E are not conditionally independent, we may subtract from the total indirect effect the sum of the indirect effect solely transmitted through M_V and the indirect effect solely transmitted through M_E; the remaining is the indirect effect for which the two pathways cannot be differentiated.

Bias in Identifying the Indirect Effects via M_V and Those via M_E

If the indirect effect undifferentiated between the two pathways was in fact transmitted via M_E, the true indirect effect via M_V would be equivalent to the indirect effect solely transmitted through M_V; while the true indirect effect via M_E would be equivalent to the sum of the indirect effect solely transmitted through M_E and the undifferentiated component. Similarly, if the indirect effect undifferentiated between the two pathways was in fact transmitted via M_V, the true indirect effect via M_E would be equivalent to the indirect effect solely transmitted through M_E; and the true indirect effect via M_V would be equivalent to the sum of the indirect effect solely transmitted through M_V and the undifferentiated component. The bias in each indirect effect is therefore the discrepancy between the initial identification result under Assumption 5 and the updated one as described here.

Bias in Identifying the Direct Effect

When Assumption 5 holds, the identification result for the direct effect based on the weighting scheme in the Theorem and that based on the weighting scheme in equation (4) are equivalent. If Assumption 5 is violated, the latter will still be valid and thus can be used to estimate the bias in the former identification result for the direct effect.

[5] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Results

We find evidence in the data that M_V and M_E are not conditionally independent given the observed covariates, which indicates a violation of Assumption 5. Therefore, a part of the total indirect effect is undifferentiated between M_V and M_E. Table 6 lists the effect sizes of the original estimates obtained under Assumption 5 in column 2; the effect sizes of bias values if the undifferentiated part was transmitted through M_V and those if the undifferentiated part was transmitted through M_E are listed in columns 3 and 4, respectively. According to the sensitivity analysis, the violation of Assumption 5 has led to a negative bias in the amount of -0.006 in the effect size of the direct effect, an amount too small to be consequential in this case. Next, we consider the hypothetical scenario that the portion of the total indirect effect undifferentiated between the two pathways was in fact transmitted through M_V. The effect size of the bias in the estimated indirect effect via M_E without affecting vocational training and that beyond affecting vocational training would be -0.007 and 0.001, respectively. Removing the bias in each case would not change the initial conclusion. However, the effect size of bias in the estimated indirect effect via M_V without affecting education would be 0.005 and that in the effect via M_V beyond affecting education would be 0.013. Removing these bias values would decrease the former effect by 28 percent and the latter by 57 percent. Finally, we consider the scenario that the undifferentiated portion was transmitted through M_E instead. Removing the bias in the estimated indirect effects via M_E would reduce the magnitude of the point estimates of these effects by at least 58 percent, while removing the bias in the estimated indirect effects via M_V would double the point estimates. Hence, the estimated effect sizes of some of the indirect effects appear to be sensitive to the violation of Assumption 5.

DISCUSSION

Substantive Contributions

Even though the National Job Corps Study was conducted more than two decades ago, the findings that we have generated are highly relevant in today's policy environment. This is because the population of youth served by Job Corps today is similar in composition and faces a similar array of challenges as the population of Job Corps applicants 20 years ago. Comparing the Job Corps Annual Reports from the mid-1990s to the present, we have found no major changes to the program theory, the program design, and its operation over these years (see U.S. Department of Labor Office of Job Corps "Policy and Requirements Handbook" and "Policy and Requirements Handbook Record of Changes," both dated April 2016). In the meantime, the operation of the program has received public scrutiny from time to time especially when the media pays attention to incidents at individual Job Corps centers without offering a comprehensive evaluation of program effectiveness.

The Job Corps program theory involves general education and vocational training as two concurrent mediators that are expected to play pivotal roles in transmitting the program impact on earnings for disadvantaged youth. Our re-analysis of the National Job Corps Study reveals that both vocational training and general education mediated the impact of Job Corps on earnings. Our evidence suggests that these two pathways are complementary rather than mutually reinforcing. Meanwhile, the importance of supplementary services and other intermediate process that distinguished the experiences between Job Corps and the control condition should not be dismissed. These services are generally unavailable to youths disconnected from school and work in the absence of Job Corps. As we noted earlier, about half of the ITT effect is transmitted through supplementary services on average. By further

examining the between-site variance of each causal effect, we have found that, relative to the control condition, most Job Corps centers successfully increased educational and vocational attainment that subsequently increased earnings on average. However, Job Corps centers did not equally succeed in promoting economic well-being among disadvantaged youths. This was likely determined to a considerable extent by the variation in the quantity and quality of supplementary services which were at the discretion of each local center.

These findings suggest useful directions for future research on program improvement. Many social programs designed for reducing economic inequality have produced minimal impacts because program participants tend to "have more trouble in their lives than the programs could correct" (Pouncy, 2000, p. 269). The holistic approach to human development, a unique trademark of Job Corps, is best manifest when supplementary services effectively reduce risks and improve the social-emotional well-being of disadvantaged youth while the education and training programs enhance their literacy and vocational skills. We reason that the overall effectiveness of the program would be greatly improved if the effectiveness of supplementary services at the centers that were in the bottom half of the distribution could be raised to the average level. In addition, even though Job Corps increased vocational training attainment by almost a factor of three and increased general education attainment by about two-thirds when compared with the control condition, a great majority of Job Corps participants left the program without a vocational certificate and more than 50 percent left with neither a vocational certificate nor an education certificate. These results indicate a clear need to improve the retention and engagement of Job Corps participants and a potential to further strengthen the Job Corps education and training programs. Analysts have reported in the past that the Job Corps performance evaluation system failed to distinguish the centers that operated effectively from the ineffective ones (Burghardt & Schochet, 2001; Schochet, Burghardt, & McConnell, 2006). Future research may investigate whether students' experiences with the program at different sites explain the variation in program impacts revealed in this study and may further identify exemplary program features that show promise for enhancing the quality of educational experiences. These may include the collective knowledge, skills, and commitment of program staff and the allocation of support resources to students according to individual needs.

Importantly, even though the success rate of education and training was limited under Job Corps, the average indirect effect of each, especially the indirect effect transmitted through education mainly in the form of obtaining a GED certificate, is noteworthy. Researchers in the past have reported that GED recipients earn less than high school graduates and that they even earn less than high school dropouts who do not have GED certificates yet are at the same level of cognitive ability (e.g., Heckman, Hsee, & Rubinstein, 2003; Heckman & Rubinstein, 2001). Heckman and colleagues attributed the discrepancy to a lack of non-cognitive skills, such as persistence and self-discipline, among GED recipients. In contrast, our results seem to indicate a clear value added by a GED certificate for Job Corps participants. This is likely because, unlike typical GED programs available to those in the control group, the Job Corps program aims to improve both cognitive and non-cognitive skills. In addition, Job Corps tailored the pace of GED instruction to participants' individual abilities and provided individualized tutorial assistance to participants who were not performing at the expected pace. Some centers further offered "enrichment" courses beyond the basic curriculum to relatively advanced participants. We suspect that the comprehensiveness and the flexibility of the educational opportunities offered by Job Corps may have effectively facilitated individual academic growth and thus enhanced the mediating role of GED attainment.

According to our sensitivity analysis results, additional adjustment for unmeasured pre-treatment confounders could possibly change the conclusions about the

positive direction of the indirect effects via vocational training and of the indirect effect via the education-by-training interaction. The conclusions about the positive direction of the indirect effects via education and of the direct effect are relatively insensitive. The initial estimate of the ITT effect is likely an underestimate of the benefit of Job Corps on earnings given that the plausible values of bias are mostly negative. Omitting the compliance measure in the initial analysis has introduced minimal bias to the estimated causal effects and the between-site variances of these effects. Finally, there is evidence that vocational training attainment and general education attainment are not conditionally independent given the observed covariates. We find that some of the estimated indirect effects are sensitive to this violation.

Methodological Contributions and Limitations

The proposed procedure for investigating complex mediation mechanisms in multisite trials fills an important gap in the literature. It enables researchers to ask a new set of empirical questions crucial for testing across a wide range of settings the generalizability of an intervention theory that involves two concurrent mediators. Relying on a series of weighting adjustments, the procedure reduces reliance on the outcome model specification and is expected to enhance the internal validity and the external validity of the analytic results under the identification assumptions. We emphasize the importance of assessing the sensitivity of causal conclusions to potential violations of the identification assumptions and present a novel sensitivity analysis procedure for disentangling the indirect effects via two potentially correlated mediators.

We acknowledge limitations of the proposed procedure. First, even though the weighting methods do not require correct outcome model specifications, the specifications of the parametric propensity score models are at best a close approximation of the true models. Misspecifying the functional form of the propensity score models may in fact introduce bias (Hong, 2010b). To address this issue, other researchers have proposed nonparametric modeling or machine learning as alternative ways to estimate propensity scores (Shortreed & Ertefaie, 2017). Second, even though the weighting-based estimators of the causal effects attain the semi-parametric efficiency bound when the propensity score models are correctly specified, efficiency may be improved through an inclusion of strong predictors of the outcome in a weighted outcome model or through other versions of the multiply robust estimation strategy such as the one proposed by Tchetgen Tchetgen and Shpitser (2012). Third, the proposed estimation procedure may not be optimal for data with small site sizes. In general, no matter what analytic method one employs, a reduction in site size reduces the statistical power, especially when the number of mediators increases. Fourth, this study has considered binary measures of educational and vocational attainment. Since human capital is clearly not dichotomous in nature, a more informative analysis could use continuous measures of academic and vocational skills, focusing on skill accrual as mediators. Fifth, to assess whether the between-site variance estimates are sensitive to omitted confounding would additionally require accounting for potential changes in the uncertainty of the estimated weights. We leave these topics for future research.

XU QIN is an Assistant Professor in the Department of Health and Human Development at the University of Pittsburgh, 5924 Wesley W. Posvar Hall, 230 South Bouquet Street, Pittsburgh, PA 15260 (e-mail: xuqin@pitt.edu).

JONAH DEUTSCH is a Senior Researcher at Mathematica Policy Research, 111 E. Wacker Drive, Suite 3000, Chicago, IL 60601 (e-mail: JDeutsch@mathematica-mpr.com).

GUANGLEI HONG is a Professor in the Department of Comparative Human Development at the University of Chicago, Rosenwald Hall, Room 325A, 1126 E. 59th Street, Chicago, IL 60637 (e-mail: ghong@uchicago.edu).

ACKNOWLEDGMENTS

The authors would like to thank Edward Bein, Donald Hedeker, Stephen Raudenbush, Peter Schochet, and Margaret Beale Spencer for their contribution of ideas and their comments on earlier versions of this article. We also acknowledge the invaluable comments of the editors and the three reviewers for the journal on earlier drafts. Alma Vigil provided invaluable research assistance to the analysis of the NJCS data. Sheena Flowers provided support in reformatting the appendices. This study was supported by a grant from the National Science Foundation (SES 1659935), a U.S. Department of Education Institute of Education Sciences, Statistical and Research Methodology Grant (R305D120020), and a subcontract from MDRC funded by the Spencer Foundation. In addition, the first author received a Quantitative Methods in Education and Human Development Research Predoctoral Fellowship from the University of Chicago and a National Academy of Education/Spencer Foundation Dissertation Fellowship.

REFERENCES

Albert, J. M., & Nelson, S. (2011). Generalized causal mediation analysis. Biometrics, 67, 1028–1038.

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, 142.

Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. Chicago, IL: University of Chicago Press.

Bein, E., Deutsch, J., Hong, G., Porter, K., Qin, X., & Yang, C. (2018). Two-step estimation in RMPW analysis. Statistics in Medicine, 37, 1304–1324.

Bloom, H. S., Unterman, R., Zhu, P., & Reardon, S. F. (2020). Lessons from New York City's small schools of choice about high school features that promote graduation for disadvantaged students. Journal of Policy Analysis and Management, 39, 740–771.

Blundell, R., Dearden, L., Meghir, C., & Sianesi, B. (1999). Human capital investment: The returns from education and training to the individual, the firm and the economy. Fiscal Studies, 20, 1–23.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. Sociological methodology, 17, 37-69.

Brookhart, M., Schneeweiss, S., Rothman, K., Glynn, R., Avorn, J., & Sturmer, T. (2006). Variable selection for propensity score models. American Journal of Epidemiology, 163, 1149–1156.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). Journal of Personality and Social Psychology, 98, 550–558.

Burghardt, J., & Schochet, P. Z. (2001). National Job Corps study: Impacts by center characteristics. Princeton, NJ: Mathematica Policy Research.

Daniel, R., De Stavola, B., Cousens, S., & Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. Biometrics, 71, 1–14.

De Luna, X., Waernbaum, I., & Richardson, T. (2011). Covariate selection for the nonparametric estimation of an average treatment effect. Biometrika, 98, 861–875.

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. Biometrics, 49, 1231–1236.

Duncan, G., Morris, P., & Rodrigues, C. (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. Developmental Psychology, 47, 1263–1279.

Flores, C. A., & Flores-Lagunes, A. (2013). Partial identification of local average treatment effects with an invalid instrument. Journal of Business & Economic Statistics, 31, 534–545.

Gennetian, L., Morris, P., Bos, J., & Bloom, H. S. (2005). Constructing instrumental variables from experimental data to explore how treatments produce effects. In H. S. Bloom (Ed.), Learning more from social experiments: Evolving analytic approaches (pp. 75–114). New York, NY: Russell Sage Foundation.

Goldberger, A. S. (1983). Abnormal selection bias. In Studies in econometrics, time series, and multivariate statistics (pp. 67–84). Cambridge, MA: Academic Press.

Goldstein, H. (2011). Multilevel statistical models (Vol. 922). Chichester, UK: John Wiley.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica: Journal of the Econometric Society, 50, 1029–1054.

Hanushek, E. A., Schwerdt, G., Woessmann, L., & Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the lifecycle. Journal of Human Resources, 52, 48–87.

Heckman, J. J., Hsee, J., & Rubinstein, Y. (2003). The GED is a mixed signal: The effect of cognitive skills and personality skills on human capital and labor market outcomes. Unpublished manuscript, University of Chicago.

Heckman, J. J., & LaFontaine, P. A. (2006). Bias-corrected estimates of GED returns. Journal of Labor Economics, 24, 661–700.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. American Economic Review, 91, 145–149.

Holland, P. (1988). Causal inference, path analysis, and recursive structural equations models. Sociological Methodology, 18, 449–484.

Hong, G. (2010a). Ratio of mediator probability weighting for estimating natural direct and indirect effects. In Proceedings of the American Statistical Association, biometrics section (pp. 2401–2415). American Statistical Association.

Hong, G. (2010b). Marginal mean weighting through stratification: Adjustment for selection bias in multilevel data. Journal of Educational and Behavioral Statistics, 35, 499–531.

Hong, G. (2015). Causality in a social world: Moderation, mediation and spill-over. West Sussex, England: John Wiley.

Hong, G., Deutsch, J., & Hill, H. D. (2015). Ratio-of-mediator-probability weighting for causal mediation analysis in the presence of treatment-by-mediator interaction. Journal of Educational and Behavioral Statistics, 40, 307–340.

Hong, G., & Nomi, T. (2012). Weighting methods for assessing policy effects mediated by peer change. Journal of Research on Educational Effectiveness, 5, 261–289.

Hong, G., Qin, X., & Yang, F. (2018). Weighting-based sensitivity analysis in causal mediation studies. Journal of Educational and Behavioral Statistics, 43, 32–56.

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. Journal of the American Statistical Association, 101, 901–910.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47, 663–685.

Huber, M. (2014). Identifying causal mechanisms (primarily) based on inverse probability weighting. Journal of Applied Econometrics, 29, 920–943.

Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? Journal of Applied Econometrics, 23, 305–327.

Imai, K., & Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. Political Analysis, 21, 141–171.

Jo, B. (2008). Causal inference in randomized experiments with mediational processes. Psychological Methods, 13, 314-336.

Johnson, T. R., Dugan, M. K., & Gritz, R. M. (2000). National Job Corps study: Job Corps applicants' programmatic experiences. Seattle, WA: Battelle Memorial Institute.

Johnson, T., Gritz, M., Jackson, R., Burghardt, J., Boussy, C., Leonard, J., & Orians, C. (1999). National Job Corps study: Report on the process analysis. Princeton, NJ: Mathematica Policy Research.

Jöreskog, K.G. (1970). A general method for analysis of covariance structures. Biometrika, 57, 239–251.

Judd, C. M., & Kenny, D. A. (1981). Process analysis estimating mediation in treatment evaluations. Evaluation Review, 5, 602–619.

Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. Statistical Science, 22, 523–539.

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

Lange, T., Rasmussen, M., & Thygesen, L. C. (2014). Assessing natural direct and indirect effects through multiple pathways. American Journal of Epidemiology, 179, 513–518.

Lange, T., Vansteelandt, S., & Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects. American Journal of Epidemiology, 176, 190–195.

Lee, B. K., Lessler, J., & Stuart, E. A. (2011). Weight trimming and propensity score weighting. PloS One, 6, e18174.

Leite, W. L., Jimenez, F., Kaya, Y., Stapleton, L. M., MacInnes, J. W., & Sandbach, R. (2015). An evaluation of weighting methods based on propensity scores to reduce selection bias in multilevel observational studies. Multivariate Behavioral Research, 50, 265–284.

MacKinnon, D.P. (2008). Introduction to Statistical Mediation Analysis. Mahwah, NJ: Lawrence Erlbaum Associates.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. Evaluation Review, 17, 144–158.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. Economics Letters, 14, 201–206.

Neyman, J., & Iwaszkiewicz, K. (1935). Statistical problems in agricultural experimentation. Supplement to the Journal of the Royal Statistical Society, 2, 107–180.

Ovalle, D. (2015). Gruesome details emerge in murder of Homestead Job Corps student. Miami Herald. Retrieved August 19, 2015 from https://www.miamiherald.com/news/local/crime/article31564502.html.

Patrick, A., Schneeweiss, S., Brookhart, M., Glynn, R., Rothman, K., Avorn, J., & Stürmer, T. (2011). The implications of propensity score variable selection strategies in pharmacoepidemiology: An empirical illustration. Pharmacoepidemiology and Drug Safety, 20, 551–599.

Pearl, J. (2001). Direct and indirect effects. In J. Breese & D. Koller (Eds.), Proceedings of the seventeenth conference on uncertainty in artificial intelligence (pp. 411–420). San Francisco, CA: Morgan Kaufmann.

Pouncy, H. (2000). New directions in job training strategies for the disadvantaged. In S. Danziger & J. Waldfogel (Eds.), Securing the future: Investing in children from birth to college (pp. 264–282). New York, NY: Russel Sage Foundation.

Qin, X., & Hong, G. (2017). A weighting method for assessing between-site heterogeneity in causal mediation mechanism. Journal of Educational and Behavioral Statistics, 42, 308–340.

Qin, X., Hong, G., Deutsch, J., & Bein, E. (2019). Multisite causal mediation analysis in the presence of complex sample and survey designs and non-random attrition. Journal of the Royal Statistical Society: Series A, 182, 1343–1370.

Raudenbush, S. W., & Bloom, H. S. (2015). Learning about and from a distribution of program impacts using multisite trials. American Journal of Evaluation, 36, 475–499.

Raudenbush, S. W., Reardon, S. F., & Nomi, T. (2012). Statistical analysis for multi-site trials using instrumental variables with random coefficients. Journal of Research on Educational Effectiveness, 5, 303–332.

Raudenbush, S. W., & Schwartz, D. (2020). Randomized experiments in education, with implications for multilevel causal inference. Annual Review of Statistics and Its Application, 7, 177–208.

Reardon, S. F., & Raudenbush, S. W. (2013). Under what assumptions do site-by-treatment instruments identify average causal effects? Sociological Methods & Research, 42, 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, 53–86.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—Application to control of the healthy worker survivor effect. Mathematical Modelling, 7, 1393–1512.

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

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11, 551.

Rosenbaum, P. R. (2017). Observation & experiment: An introduction to causal inference. Cambridge, MA: Harvard University Press.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association, 79, 516–524.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. The Annals of Statistics, 6, 34–58.

Rubin, D. B. (1980). Randomization analysis of experimental data: The Fisher Randomization Test comment. Journal of the American Statistical Association, 75, 591–593.

Rubin, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. Journal of the American Statistical Association, 81, 961–962.

Rubin, D. B. (1990). Formal mode of statistical inference for causal effects. Journal of Statistical Planning and Inference, 25, 279–292.

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. Psychological Methods, 13, 279.

Schochet, P. Z. (2020). Long-run labor market effects of the Job Corps program: Evidence from a nationally representative experiment. Journal of Policy Analysis and Management, 40(1).

Schochet, P. Z., Burghardt, J., & Glazerman, S. (2001). National Job Corps study: The impacts of Job Corps on participants' employment and related outcomes [and] methodological appendixes on the impact analysis. Princeton, NJ: Mathematica Policy Research.

Schochet, P. Z., Burghardt, J., & McConnell, S. (2006). National Job Corps study and longer-term follow-up study: Impact and benefit-cost findings using survey and summary earnings records data. Princeton, NJ: Mathematica Policy Research.

Schochet, P. Z., Burghardt, J., & McConnell, S. (2008). Does Job Corps work? Impact findings from the national Job Corps study. The American Economic Review, 98, 1864–1886.

Shortreed, S. M., & Ertefaie, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. Biometrics, 73, 1111–1122.

Tchetgen Tchetgen, E. J. (2013). Inverse odds ratio-weighted estimation for causal mediation analysis. Statistics in Medicine, 32, 4567–4580.

Tchetgen Tchetgen, E. J., & Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness, and sensitivity analysis. Annals of Statistics, 40, 1816.

Thrush, G. (2018). $1.7 billion federal job training program is "failing the students." New York Times. Retrieved August 26, 2018, from https://www.nytimes.com/2018/08/26/us/politics/job-corps-training-program.html.

Tyler, J. H. (2003). Economic benefits of the GED: Lessons from recent research. Review of Educational Research, 73, 369–403.

VanderWeele, T. (2015). Explanation in causal inference: Methods for mediation and interaction. Oxford, UK: Oxford University Press.

Weiss, M., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. Journal of Policy Analysis and Management, 33, 778–808.

Weiss, M. J., Bloom, H. S., Verbitsky-Savitz, N., Gupta, H., Vigil, A. E., & Cullinan, D. N. (2017). How much do the effects of education and training programs vary across sites? Evidence from past multisite randomized trials. Journal of Research on Educational Effectiveness, 10, 843–876.

Zimmermann, K. F., Biavaschi, C., Eichhorst, W., Giulietti, C., Kendzia, M. J., Muravyev, A., … Schmidl, R. (2013). Youth unemployment and vocational training. Foundations and Trends® in Microeconomics, 9, 1–157.

APPENDIX A: IDENTIFICATION OF THE CAUSAL EFFECTS

As a supplement to the section on identification of the causal parameters in the article, this appendix provides a proof of the identification of the site-specific mean of each potential outcome. The population average effects and the between-site variances can be identified accordingly. The derivation below proves that, under Assumptions 1 through 5, the expectation of each potential outcome at site j, $E[Y(t, M_V(t'), M_E(t''))|S = j]$, for $t, t', t'' = 0, 1$, can be identified with a weighted average of the observed outcome at that site. Let $X = \{\mathbf{X}_D \cup \mathbf{X}_T \cup \mathbf{X}_R \cup \mathbf{X}_V \cup \mathbf{X}_E\}$ be the union of all the observed pre-treatment confounders. To simplify notations, we drop the subscript ij of each variable.

$$
\begin{aligned}
&E\left[Y\left(t, M_V\left(t'\right), M_E\left(t''\right)\right)|S = j\right] \\
&\quad = E\left\{E\left[Y\left(t, M_V\left(t'\right), M_E\left(t''\right)\right)|\mathbf{X} = \mathbf{x}, S = j\right]\right\} \\
&\quad = \int_x \int_{m_V} \int_{m_E} \int_y y \times f\left(Y\left(t, m_V, m_E\right) = y|\, M_V\left(t'\right) = m_V, M_E\left(t''\right) = m_E, \mathbf{X} = \mathbf{x}, S = j\right) \\
&\qquad \times Pr(M_V(t') = m_V|M_E(t'') = m_E, \mathbf{X} = \mathbf{x}, S = j) \times Pr(M_E(t'') = m_E|\mathbf{X} = \mathbf{x}, S = j) \\
&\qquad \times g\left(\mathbf{X} = \mathbf{x}\,|S = j\right) dy dm_V dm_E dx
\end{aligned}
$$

Under Assumption 1, $\{Y(t, m_V, m_E), M_V(t), M_E(t)\}\perp\!\!\!\perp D|\mathbf{X}_D = x_D, S = j$. Because $\mathbf{X}_D \subset \mathbf{X}$, $\{Y(t, m_V, m_E), M_V(t), M_E(t)\}\perp\!\!\!\perp D|\mathbf{X} = \mathbf{x}, S = j$ also holds. Hence, the above equation is equal to

$$
\begin{aligned}
&\int_x \int_{m_V} \int_{m_E} \int_y y \times f(Y(t, m_V, m_E) = y|D = 1, M_V(t') \\
&\quad = m_V, M_E(t'') = m_E, \mathbf{X} = \mathbf{x}, S = j) \times Pr(M_V(t') \\
&\quad = m_V|D = 1, M_E(t'') = m_E, \mathbf{X} = \mathbf{x}, S = j) \times Pr(M_E(t'') = m_E|D = 1,.\mathbf{X} = \mathbf{x}, S = j) \\
&\qquad \times g(\mathbf{X} = \mathbf{x}|S = j) dy dm_V dm_E dx.
\end{aligned}
$$

By Bayes theorem,

$$
g\left(\mathbf{X} = \mathbf{x}\,|S = j\right) = g\left(\mathbf{X} = \mathbf{x}\,|D = 1, S = j\right) \times \frac{Pr\left(D = 1\,|S = j\right)}{Pr\left(D = 1\,|\mathbf{X} = \mathbf{x}, S = j\right)},
$$

where $0 < Pr(D = 1|\mathbf{X} = \mathbf{x}, S = j) < 1$. When the strongly ignorable sampling mechanism (Assumption 1) holds, controlling for $\mathbf{X}_D$ removes sampling selection. Because $\mathbf{X}_D \subset \mathbf{X}$, $Pr(D = 1|\mathbf{X} = \mathbf{x}, S = j) = Pr(D = 1|\mathbf{X}_D = \mathbf{x}_D, S = j)$. Hence, it is equivalent to assuming that $0 < Pr(D = 1|\mathbf{X}_D = \mathbf{x}_D, S = j) < 1$. This is known as positivity assumption. Let $W_D = \frac{Pr(D=1|S=j)}{Pr(D=1|\mathbf{X}=\mathbf{x}, S=j)} = \frac{Pr(D=1|S=j)}{Pr(D=1|\mathbf{X}_D=\mathbf{x}_D, S=j)}$, and thus

$$
\begin{aligned}
&E[Y(t, M_V(t'), M_E(t''))|S = j] \\
&\quad = \int_x \int_{m_V} \int_{m_E} \int_y^{W_D} \times y \times f(Y(t, m_V, m_E) = .y|D = 1, M_V(t') \\
&\quad = m_V, M_E(t'') = m_E, \mathbf{X} = \mathbf{x}, S = j) \\
&\qquad \times Pr(M_V(t') = m_V|D = 1, M_E(t'') \\
&\quad = m_E, \mathbf{X} = \mathbf{x}, S = j) \times Pr(M_E(t'') = m_E|D = 1,\mathbf{X} = \mathbf{x}, S = j) \\
&\qquad \times g(\mathbf{X} = \mathbf{x}|D = 1, S = j) dy dm_V dm_E dx.
\end{aligned}
$$

Similarly, under the assumption that $0 < Pr(T = 1|\mathbf{X}_T = \mathbf{x}_T, D = 1, S = j) < 1$, let $W_T = \frac{Pr(T=t|D=1,S=j)}{Pr(T=t|\mathbf{X}_T=\mathbf{x}_T,D=1,S=j)}$. When Assumption 2 holds, i.e., $\{Y(t, m_V, m_E), M_V(t), M_E(t)\}\perp\perp T|D = 1, \mathbf{X}_T = \mathbf{x}_T, S = j$, and by Bayes theorem, the above equation is equal to

$$\begin{aligned}\int_x \int_{m_V} \int_{m_E} \int_y & W_D W_T \times y \\ & \times f(Y(t, m_V, m_E) = y|T = t, D = 1, M_V(t') = m_V, M_E(t'') = m_E, \mathbf{X} = \mathbf{x}, S = j) \\ & \times Pr(M_V(t') = m_V|T = t', D = 1, M_E(t'') = m_E, \mathbf{X} = \mathbf{x}, S = j) \\ & \times Pr(M_E(t'') = m_E|T = t'', D = 1, \mathbf{X} = \mathbf{x}, S = j) \\ & \times g(\mathbf{X} = \mathbf{x}|T = t, D = 1, S = j) dy dm_V dm_E dx\end{aligned}$$

Under the assumption that $0 < Pr(R = 1|\mathbf{X}_R = \mathbf{x}_R, T = t, D = 1, S = j) < 1$, let $W_R = \frac{Pr(R=1|T=t,D=1,S=j)}{Pr(R=1|\mathbf{X}_R=\mathbf{x}_R,T=t,D=1,S=j)}$. When Assumption 3 holds, i.e., $\{Y(t, m_V, m_E), M_V(t), M_E(t)\}\perp\perp R|T = t, D = 1, \mathbf{X}_R = \mathbf{x}_R, S = j$, and by Bayes theorem, the above equation is equal to

$$\begin{aligned}\int_x \int_{m_V} \int_{m_E} \int_y & W_D W_T W_R \times y \times f(Y(t, m_V, m_E) \\ & = y|R = 1, T = t, D = 1, M_V(t') = m_V, M_E(t'') = m_E, \mathbf{X} = \mathbf{x}, S = j) \\ & \times Pr\left(M_V(t') = m_V \middle| R = 1, T = t', D = 1, M_E(t'') = m_E, \mathbf{X} = \mathbf{x}, S = j\right) \\ & \times Pr\left(M_E(t'') = m_E \middle| R = 1, T = t'', D = 1, \mathbf{X} = \mathbf{x}, S = j\right) \\ & \times g(\mathbf{X} = x|R = 1, T = t, D = 1, S = j) dy dm_V dm_E dx.\end{aligned}$$

When $t = t' = t''$, it is easy to obtain the following identification result,

$$E\left[Y(t, M_V(t), M_E(t))|S = j\right] = E\left[W_D W_T W_R Y|R = 1, T = t, D = 1, S = j\right].$$

Similarly, it can be proved that

$$\begin{aligned}E[M_V(t)|S = j] &= E[W_D W_T W_R M_V|R = 1, T = t, D = 1, \quad S = j], \\ E[M_E(t)|S = j] &= E[W_D W_T W_R M_E|R = 1, T = t, D = 1, S = j].\end{aligned}$$

When $t' \neq t$ or $t'' \neq t$ and if Assumption 5 holds, i.e., $M_V(t)\perp\perp M_E(t')|R = 1, T = t, D = 1, \mathbf{X}_V = \mathbf{x}_V, \mathbf{X}_E = \mathbf{x}_E, S = j$, because $\{\mathbf{X}_V, \mathbf{X}_E\} \subset \mathbf{X}$, $M_V(t)\perp\perp M_E(t')|R = 1, T = t, D = 1, \mathbf{X} = \mathbf{x}, S = j$ also holds. Hence,

$$\begin{aligned}& E\left[Y(t, M_V(t'), M_E(t''))|S = j\right] \\ & = \int_x \int_{m_V} \int_{m_E} \int_y W_D W_T W_R \times y \times f(Y(t, m_V, m_E) \\ & = y|R = 1, T = t, D = 1, M_V(t') = m_V, M_E(t'') = m_E, \mathbf{X} = \mathbf{x}, S = j) \\ & \quad \times Pr\left(M_V(t') = m_V \middle| R = 1, T = t', D = 1, \mathbf{X} = \mathbf{x}, S = j\right) \\ & \quad \times Pr\left(M_E(t'') = m_E \middle| R = 1, T = t'', D = 1, \mathbf{X} = \mathbf{x}, S = j\right) \\ & \quad \times g(\mathbf{X} = \mathbf{x}|R = 1, T = t, D = 1, S = j) dy dm_V dm_E dx\end{aligned}$$

If Assumption 4 holds, because $\mathbf{X}_V \subset \mathbf{X}$ and $\mathbf{X}_E \subset \mathbf{X}$, $Y(t, m_V, m_E) \perp\!\!\!\perp \{M_V(t), M_V(t')\} | R = 1, T = t, D = 1, \mathbf{X} = \mathbf{x}, S = j$ and $Y(t, m_V, m_E) \perp\!\!\!\perp \{M_E(t), M_E(t')\} | R = 1, T = t, D = 1, \mathbf{X} = \mathbf{x}, S = j$ also hold. Hence,

$$
\begin{aligned}
&E\left[Y\left(t, M_V\left(t'\right), M_E\left(t''\right)\right) | S = j\right] \\
&\quad = \int_x \int_{m_V} \int_{m_E} \int_y W_D W_T W_R \times y \times \; f(Y\left(t, m_V, m_E\right) = y | R = 1, T = t, D = 1, M_V(t) \\
&\quad = m_V, M_E(t) = m_E, \mathbf{X} = \mathbf{x}, S = j) \\
&\qquad \times Pr\left(M_V\left(t'\right) = m_V \,\middle|\, R = 1, T = t', D = 1, \mathbf{X} = \mathbf{x}, S = j\right) \\
&\qquad \times Pr\left(M_E\left(t''\right) = m_E \,\middle|\, R = 1, T = t'', D = 1, X = \mathbf{x}, S = j\right) \\
&\qquad \times g\left(\mathbf{X} = \mathbf{x} | R = 1, T = t, D = 1, S = j\right) dy dm_V dm_E dx
\end{aligned}
$$

Let

$$
\begin{aligned}
W_{Vt'} &= \frac{Pr(M_V(t') = m_V | \mathbf{X} = \mathbf{x}, R = 1, T = t', D = 1, S = j)}{Pr(M_V(t) = m_V | \mathbf{X} = \mathbf{x}, R = 1, T = t, D = 1, S = j)} \\
&= \frac{Pr(M_V = m_V | \mathbf{X} = \mathbf{x}, R = 1, T = t', D = 1, S = j)}{Pr(M_V = m_V | \mathbf{X} = \mathbf{x}, R = 1, T = t, D = 1, S = j)}
\end{aligned}
$$

and

$$
\begin{aligned}
W_{Et''} &= \frac{Pr(M_E(t'') = m_E | \mathbf{X} = \mathbf{x}, R = 1, T = t'', D = 1, S = j)}{Pr(M_E(t) = m_E | \mathbf{X} = \mathbf{x}, R = 1, T = t, D = 1, S = j)} \\
&= \frac{Pr(M_E = m_E | \mathbf{X} = \mathbf{x}, R = 1, T = t'', D = 1, S = j)}{Pr(M_E = m_E | \mathbf{X} = \mathbf{x}, R = 1, T = t, D = 1, S = j)}
\end{aligned}
$$

because $M_V(t) = M_V$ and $M_E(t) = M_E$ when $T = t$ and, similarly $M_V(t') = M_V$ when $T = t'$, and $M_E(t'') = M_E$ when $T = t''$. This is based on the assumption that $0 < Pr(M_V = m_V | \mathbf{X} = \mathbf{x}, R = 1, T = t, D = 1, S = j) < 1$ and $0 < Pr(M_E = m_E | \mathbf{X} = \mathbf{x}, R = 1, T = t, D = 1, S = j) < 1$. When the strongly ignorable mediator selection mechanism (Assumption 4) holds, controlling for $\mathbf{X}_V$ removes selection of mediator M_V and controlling for $\mathbf{X}_E$ removes selection of mediator M_E. Because $\mathbf{X}_V \subset \mathbf{X}$ and $\mathbf{X}_E \subset \mathbf{X}$, the positivity assumptions can be simplified as $0 < Pr(M_V = m_V | \mathbf{X}_V = \mathbf{x}_V, R = 1, T = t, D = 1, S = j) < 1$ and $0 < Pr(M_E = m_E | \mathbf{X}_E = \mathbf{x}_E, R = 1, T = t, D = 1, S = j) < 1$, and the weights are equal to

$$
W_{Vt'} = \frac{Pr(M_V = m_V | \mathbf{X}_V = \mathbf{x}_V, R = 1, T = t', D = 1, S = j)}{Pr(M_V = m_V | \mathbf{X}_V = \mathbf{x}_V, R = 1, T = t, D = 1, S = j)}
$$

and

$$
W_{Et''} = \frac{Pr(M_E = m_E | \mathbf{X}_E = \mathbf{x}_E, R = 1, T = t'', D = 1, S = j)}{Pr(M_E = m_E | \mathbf{X}_E = \mathbf{x}_E, R = 1, T = t, D = 1, S = j)}.
$$

Then

$$
\begin{aligned}
&E\left[Y\left(t, M_V\left(t'\right), M_E\left(t''\right)\right) | S=j\right] \\
&\quad=\int_x \int_{m_V} \int_{m_E} \int_y W_D W_T W_R W_{Vt'} W_{Et''} \times y \times f(Y(t, m_V, m_E) \\
&=y | R=1, T=t, D=1, M_V(t)=m_V, M_E(t)=m_E, \mathbf{X}=\mathbf{x}, S=j) \\
&\quad \times Pr\left(M_V(t)=m_V | R=1, T=t, D=1, \mathbf{X}=\mathbf{x}, S=j\right) \\
&\quad \times Pr\left(M_E(t)=m_E | R=1, T=t, D=1, \mathbf{X}=\mathbf{x}, S=j\right) \\
&\quad \times g\left(\mathbf{X}=\mathbf{x} | R=1, T=t, D=1, S=j\right) dy dm_V dm_E dx \\
&=E\left[W_D W_T W_R W_{Vt'} W_{Et''} Y | R=1, T=t, D=1, S=j\right]
\end{aligned}
$$

If $t' = t$, then $W_{Vt'} = 1$; if $t'' = t$, then $W_{Et''} = 1$.

APPENDIX B: ASYMPTOTIC VARIANCE OF THE TWO-STEP ESTIMATORS

As described in the section on estimation and inference of the causal effects in the article, the estimation of both the RMPW and non-response weight poses a challenge to statistical inference. Multilevel logistic regressions are employed in step 1 to estimate the non-response weight and the RMPW weight while step 2 involves site-by-site method-of-moments analysis to estimate the causal parameters. To represent the sampling variability of the estimated weights in the standard errors of the causal effect estimates, we extend the procedure proposed by Qin and Hong (2017). The procedure stacks the moment functions from the two steps and solves them simultaneously. This appendix presents the detailed estimation procedure and provides derivations of the asymptotic variance of the two-step estimators, as a supplement to the section on estimation and inference of the causal effects in the main paper. Due to the space limit, we focus on estimating the population average and between-site variance of the mediation-related effects in step 2. The estimation of the population average and between-site variance of the ITT effects follows the same logic.

Non-Response and RMPW Weight Estimation in Step 1

Non-Response Weight Estimation

We first estimate the non-response weight in equation (1) based on the 48-month sample including both responders and non-responders by fitting multilevel logistic regression models of the response indicator.

To estimate the numerator of the weight, $p^{(N)}_{Rtij} = Pr(R_{ij} = 1|T_{ij} = t, D_{ij} = 1, S_{ij} = j)$, which is constant among all the sampled individuals in the same treatment group t at site j, we fit the following models, one to the program group sample and the other to the control group sample:

$$\log\left[\frac{p^{(N)}_{Rtij}}{1-p^{(N)}_{Rtij}}\right] = \pi^{(N)}_{Rt} + r^{(N)}_{Rtj}, r^{(N)}_{Rtj} \sim N\left(0, \sigma^{(N)2}_{Rt}\right), \tag{B.1}$$

in which $\pi^{(N)}_{Rt}$ indicates the overall mean of the log-odds of response of the sampled individuals assigned to group t, $r^{(N)}_{Rtj}$ is a site-specific random intercept, indicating the deviance of the log-odds of response in treatment group t at site j from its overall mean. The superscript N serves as shorthand for the numerator.

The denominator, $p^{(D)}_{Rtij} = Pr(R_{ij} = 1|\mathbf{X}_{Rij} = \mathbf{x}_R, T_{ij} = t, D_{ij} = 1, S_{ij} = j)$, is the conditional probability of response for each sampled individual in each treatment group as a function of the pre-treatment covariates $\mathbf{X}_R$. We fit the model

$$log\left[\frac{p^{(D)}_{Rtij}}{1-p^{(D)}_{Rtij}}\right] = \mathbf{X}'_{Rij}\boldsymbol{\pi}^{(D)}_{Rt} + r^{(D)}_{Rtj}, r^{(D)}_{Rtj} \sim N\left(0, \sigma^{(D)2}_{Rt}\right), \tag{B.2}$$

in which $\mathbf{X}_{Rij}$ is a vector of the intercept and all the observed pre-treatment covariates, $\boldsymbol{\pi}^{(D)}_{Rt}$ is the corresponding vector of coefficients, and $r^{(D)}_{Rtj}$ is a site-specific random intercept following a normal distribution with variance $\sigma^{(D)2}_{Rt}$. The superscript D is shorthand for the denominator. As the sample size increases, more flexible forms of the models could be considered.

We fit the response models (1) and (2) through maximum likelihood estimation. After obtaining the maximum likelihood estimates of the coefficients and the Empirical Bayes estimates of the random intercepts (Raudenbush & Bryk, 2002), we predict the response probability for each sampled individual. As the number of sites and the sample size at each site increase, the predicted probabilities converge in probability to the corresponding true probabilities. Based on equation (1), we then compute the non-response weight for each respondent $\widehat{W}_{Rij} = \widehat{p}^{(N)}_{Rtij}/\widehat{p}^{(D)}_{Rtij}$.

RMPW Weight Estimation

Following a similar procedure, we estimate for each respondent the RMPW weights, $W_{Vt'ij}$ and $W_{Et''ij}$, as defined in equations (2) and (3). Below, we focus on the estimation of $W_{Vt'ij}$, while the same procedure applies to $W_{Et''ij}$. The following multilevel mixed-effects logistic regression models are fitted to the respondents in each treatment group through maximum likelihood estimation:

$$\log\left[\frac{p_{Vtij}}{1-p_{Vtij}}\right] = \mathbf{X}'_{M_{Vij}}\boldsymbol{\pi}_{Vt} + r_{Vtj}, r_{Vtj} \sim N\left(0, \sigma^2_{Vt}\right), \tag{B.3}$$

for $t = 0, 1$ in which $p_{Vtij} = Pr(M_{Vij} = 1|\mathbf{X}_{Vij} = \mathbf{x}_V, R_{ij} = 1, T_{ij} = t, D_{ij} = 1, S_{ij} = j)$. As the sample size increases, more flexible forms of the models could be considered.

Based on the estimates of the parameters in the above mediator model fitted to treatment group t, we can directly predict the mediator probability for a respondent in treatment group t. To predict the individual's mediator probability under the alternative treatment condition t', we apply to the individual the coefficients and the random intercept estimated from the model fitted to the alternative treatment group t'. We will then obtain the estimated RMPW weight for a respondent who was assigned to treatment group t. The weight is $\widehat{W}_{Vt'ij} = \widehat{p}_{Vt'ij}/\widehat{p}_{Vtij}$ if he or she successfully attained a vocational credential, and is $\widehat{W}_{Vt'ij} = (1-\widehat{p}_{Vt'ij})/(1-\widehat{p}_{Vtij})$ if he or she failed to do so. The estimated weights converge to the true weight with the increase in the number of sites and the site size.

Moment Functions in Step 1

Take model (B.2) for the estimation of the denominator of the non-response weight as an example. For computational simplicity, following Hedeker and Gibbons (2006), we standardize the random effects $r^{(D)}_{Rtj}$ by representing it as $\sigma^{(D)}_{Rt}\theta^{(D)}_{Rtj}$, where $\theta^{(D)}_{Rtj}$ follows a standardized normal distribution. The estimators $\widehat{\eta}^{(D)}_{Rt} = \left(\widehat{\boldsymbol{\pi}}^{(D)'}_{Rt}, \widehat{\sigma}^{(D)}_{Rt}\right)'$, for $t = 0, 1$, solve the following estimating equations,

$$\frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{n_j} h^{(1.D)}_{Rtij}\left(R_{ij}, T_{ij}, D_{ij}, \mathbf{X}_{ij}, \theta^{(D)}_{Rtj}, \eta^{(D)}_{Rt}\right) = 0,$$

where $N = \sum_{j=1}^{J} n_j$ and $h_{Rtij}^{(1.D)}$ has the same dimension as $\eta_{Rt}^{(D)} = (\boldsymbol{\pi}_{Rt}^{(D)'}, \sigma_{Rt}^{(D)})'$. The above equation is essentially first-order conditions for maximum-likelihood estimators in the multilevel logistic regression.

$$\sum_{j=1}^{J}\sum_{i=1}^{n_j} h_{Rtij}^{(1.D)} = \frac{\partial \log \mathcal{L}_{Rt}^{(D)}}{\partial \eta_{Rt}^{(D)}} = \frac{\partial \sum_{j=1}^{J} \log l_{Rt}^{(D)}\left(R_j\right)}{\partial \eta_{Rt}^{(D)}} = \sum_{j=1}^{J} \frac{1}{l_{Rt}^{(D)}\left(R_j\right)} \cdot \frac{\partial l_{Rt}^{(D)}\left(R_j\right)}{\partial \eta_{Rt}^{(D)}},$$

in which

$$l_{Rt}^{(D)}\left(R_j\right) = \int_{\theta_{Rtj}^{(D)}} f_{Rt}^{(D)}\left(R_j \left| \theta_{Rtj}^{(D)}\right.\right) g_{Rt}^{(D)}\left(\theta_{Rtj}^{(D)}\right) d\theta_{Rtj}^{(D)}.$$

To approximate the above integral, we use Gauss-Hermite quadrature (Stroud & Secrest, 1966) by summing over a specified number of quadrature points Q for the integration, given that $\theta_{Rtj}^{(D)}$ is assumed to follow a normal distribution (Hedeker & Gibbons, 2006). Let the optimal points be $B_{Rtq}^{(D)}$ and the weights be $A_{Rt}^{(D)}(B_{Rtq}^{(D)})$, for $q = 1, \ldots, Q$, under treatment condition $t = 0, 1$. Finally, it can be proved that

$$h_{Rtij}^{(1.D)} \approx \frac{1}{l_{Rt}^{(D)}\left(R_j\right)} \cdot \sum_{q=1}^{Q} I\left(T_{ij} = t\right)\left[\frac{R_{ij} - p_{Rtijq}^{(D)}}{p_{Rtijq}^{(D)}\left(1 - p_{Rtijq}^{(D)}\right)} \cdot \frac{\partial p_{Rtijq}^{(D)}}{\partial \eta_{Rt}^{(D)}}\right]$$
$$\times\, f_{Rt}^{(D)}\left(R_j \left| B_{Rtq}^{(D)}\right.\right) A_{Rt}^{(D)}\left(B_{Rtq}^{(D)}\right),$$

in which $I(T_{ij} = t)$ is an indicator taking value 1 if individual i at site j is from treatment group t and 0 otherwise, and $l_{Rt}^{(D)}(R_j) \approx \sum_{q=1}^{Q} f_{Rt}^{(D)}(R_j|B_{Rtq}^{(D)})\ A_{Rt}^{(D)}(B_{Rtq}^{(D)})$, $f_{Rt}^{(D)}(R_j|\theta_{Rtj}^{(D)}) = \prod_{i=1}^{n_j}[(p_{Rtij}^{(D)})^{R_{ij}}(1 - p_{Rtij}^{(D)})^{1-R_{ij}}]^{I(T_{ij}=t)}$, $\frac{\partial p_{Rtijq}^{(D)}}{\partial \eta_{Rt}^{(D)}} = (p_{Rtijq}^{(D)}(1 - p_{Rtijq}^{(D)})$ $\mathbf{X}'_{ij}, p_{Rtijq}^{(D)}(1 - p_{Rtijq}^{(D)})B_{Rtq}^{(D)})'$.

Following the same procedure, we could obtain maximum likelihood estimates of the parameters in the response model (B.1) for the estimation of the numerator of the non-response weight and the parameters in the mediator models, based on moment functions $h_{Rtij}^{(1.N)}$, $h_{Vtij}^{(1)}$and $h_{Etij}^{(1)}$.

Let $h_{Rij}^{(1)'} = (h_{R0ij}^{(1.D)'}, h_{R1ij}^{(1.D)'}, h_{R0ij}^{(1.N)'}, h_{R1ij}^{(1.N)'})'$ indicate moment functions for $\widehat{\eta}'_R = (\widehat{\eta}_{R0}^{(D)'}, \widehat{\eta}_{R1}^{(D)'}, \widehat{\eta}_{R0}^{(N)'}, \widehat{\eta}_{R1}^{(N)'})'$, and let $h_{Mij}^{(1)'} = (h_{V0ij}^{(1)'}, h_{V1ij}^{(1)'}, h_{E0ij}^{(1)'}, h_{E1ij}^{(1)'})'$ be moment functions of $\widehat{\eta}'_M = (\widehat{\eta}'_{V0}, \widehat{\eta}'_{V1}, \widehat{\eta}'_{E0}, \widehat{\eta}'_{E1})'$. We use $h_{ij}^{(1)} = (h_{Rij}^{(1)'}, h_{Mij}^{(1)'})'$ to indicate the moment functions for the step-1 estimators $\widehat{\eta} = (\widehat{\eta}'_R, \widehat{\eta}'_M)'$.

Site-Specific Mean Potential Outcome Estimation in Step 2

Estimators in Step 2

In the next step, we estimate the site-specific means of the five potential outcomes, identified by $\mu = (\mu'_1 \ldots \mu'_J)'$ where $\mu_j = (\mu_{0j}, \mu_{1j}, \mu_{1.0.1j}, \mu_{1.1.0j}, \mu_{1.0.0j})'$ for

$j = 1, \ldots, J$, through weighted mean outcomes of the sample respondents across all the sites. The estimators at site s are as follows.

$$\widehat{\mu}_{t.t'.t''s} = \frac{\frac{1}{N}\sum_{i=1}^{N} W_{Dij} W_{Tij} \widehat{W}_{Rij} \widehat{W}_{Vt'ij} \widehat{W}_{Et''ij} \left(Y_{ij} - \mu_{t.t'.t''s}\right) R_{ij} I\left(T_{ij} = t\right) D_{ij} I\left(S_{ij} = s\right) Y_{ij}}{\frac{1}{N}\sum_{i=1}^{N} W_{Dij} W_{Tij} \widehat{W}_{Rij} \widehat{W}_{Vt'ij} \widehat{W}_{Et''ij} \left(Y_{ij} - \mu_{t.t'.t''s}\right) R_{ij} I\left(T_{ij} = t\right) D_{ij} I\left(S_{ij} = s\right)},$$

in which the non-response weight $\widehat{W}_{Rij}$ and the RMPW weights $\widehat{W}_{Vt'ij}$ and $\widehat{W}_{Et''ij}$ are estimated based on the first-step estimators, $\widehat{\eta}$ in step 1. $I(T_{ij} = t)$ takes value 1 if the individual was assigned to treatment group t and 0 if not. $I(S_{ij} = s)$ is an indicator taking value 1 if individual i is from site s and 0 otherwise. In other words, each individual only contributes to the estimation of the mean potential outcomes at the site that the individual is from.

Moment Functions in Step 2

The step-2 estimators $\widehat{\mu}_s = (\widehat{\mu}_{0.0.0s}, \widehat{\mu}_{1.1.1s}, \widehat{\mu}_{1.0.1s}, \widehat{\mu}_{1.1.0s}, \widehat{\mu}_{1.0.0s})'$ are obtained by solving the moment conditions

$$\begin{aligned} h_{ij}^{(2)} = \Big(&h_{ij,0.0.01}^{(2)}, h_{ij,1.1.11}^{(2)}, h_{ij,1.0.11}^{(2)}, h_{ij,1.1.01}^{(2)}, h_{ij,1.0.01}^{(2)}, \\ &\ldots, h_{ij,0.0.0J}^{(2)}, h_{ij,1.1.1J}^{(2)}, h_{ij,1.0.1J}^{(2)}, h_{ij,1.1.0J}^{(2)}, h_{ij,1.0.0J}^{(2)}\Big)', \end{aligned}$$

in which

$$\begin{aligned} \frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{n_j} h_{ij,t.t'.t''s}^{(2)} = \frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{n_j} &W_{Dij} W_{Tij} W_{Rij} W_{Vt'ij} W_{Et''ij} \\ &\times \left(Y_{ij} - \mu_{t.t'.t''s}\right) R_{ij} I\left(T_{ij} = t\right) D_{ij} I\left(S_{ij} = s\right) = 0. \end{aligned}$$

Asymptotic Sampling Variance of the Two-Step Estimators

Stacking the moment functions from both steps, we have that $h_{ij} = (h_{ij}^{(1)'}, h_{ij}^{(2)'})'$. The estimators in the two steps can be rewritten as a one-step estimator $\widehat{\vartheta} = (\widehat{\eta}', \widehat{\mu}')'$, which jointly solves $\frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{n_j} h_{ij} = \mathbf{0}$. Under the standard regularity conditions, $\widehat{\vartheta}$ is a consistent estimator of $\vartheta = (\eta', \mu')'$ with the asymptotic sampling distribution (Hansen, 1982):

$\sqrt{N}(\widehat{\vartheta} - \vartheta) \xrightarrow{d} \mathcal{N}(0, \widetilde{\text{var}}(\widehat{\vartheta} - \vartheta))$.

The asymptotic covariance matrix of $\widehat{\vartheta} - \vartheta$ is $\widetilde{\text{var}}(\widehat{\vartheta} - \vartheta)/N$, in which

$$\widetilde{\text{var}}\left(\widehat{\vartheta} - \vartheta\right) = \begin{pmatrix} \widetilde{\text{var}}\left(\widehat{\eta} - \eta\right) & \widetilde{\text{cov}}\left(\widehat{\eta} - \eta, \widehat{\mu} - \mu\right) \\ \widetilde{\text{cov}}\left(\widehat{\mu} - \mu, \widehat{\eta} - \eta\right) & \widetilde{\text{var}}\left(\widehat{\mu} - \mu\right) \end{pmatrix} = G^{-1} H \left(G^{-1}\right)',$$

where

$$H = E\left[h_{ij}h'_{ij}\right] = E\begin{bmatrix} h_{ij}^{(1)}h_{ij}^{(1)'} & h_{ij}^{(1)}h_{ij}^{(2)'} \\ h_{ij}^{(2)}h_{ij}^{(1)'} & h_{ij}^{(2)}h_{ij}^{(2)'} \end{bmatrix};$$

$$G = E\left[\frac{\partial h_{ij}}{\partial \vartheta}\right] = E\begin{bmatrix} \frac{\partial h_{ij}^{(1)}}{\partial \eta'} & \mathbf{0} \\ \frac{\partial h_{ij}^{(2)}}{\partial \eta'} & \frac{\partial h_{ij}^{(2)}}{\partial \mu'} \end{bmatrix} = \begin{bmatrix} G_{11} & \mathbf{0} \\ G_{21} & G_{22} \end{bmatrix},$$

in which

$$G_{11} = E\left[\frac{\partial h_{ij}^{(1)}}{\partial \eta}\right] = E\begin{bmatrix} \frac{\partial h_{Rij}^{(1)}}{\partial \eta'_R} & \frac{\partial h_{Rij}^{(1)}}{\partial \eta'_M} \\ \frac{\partial h_{Mij}^{(1)}}{\partial \eta'_R} & \frac{\partial h_{Mij}^{(1)}}{\partial \eta'_M} \end{bmatrix};$$

$$= \begin{bmatrix} -E\begin{bmatrix} h_{Rij}^{(1)} & h_{Rij}^{(1)'} \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & -E\begin{bmatrix} h_{Mij}^{(1)} & h_{Mij}^{(1)'} \end{bmatrix} \end{bmatrix}$$

$$G_{22} = E\left[\frac{\partial h_{ij}^{(2)}}{\partial \mu}\right] = E\begin{bmatrix} G_{221} & & & & \\ & \ddots & & & \\ & & G_{22j} & & \\ & & & \ddots & \\ & & & & G_{22J} \end{bmatrix},$$

in which

$$G_{22j} = E\begin{bmatrix} \frac{\partial h_{ij,0j}^{(2)}}{\partial \mu_{0j}} & 0 & 0 & 0 & 0 \\ 0 & \frac{\partial h_{ij,1j}^{(2)}}{\partial \mu_{1j}} & 0 & 0 & 0 \\ 0 & 0 & \frac{\partial h_{ij,1.0.1j}^{(2)}}{\partial \mu_{1.0.1j}} & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial h_{ij,1.1.0j}^{(2)}}{\partial \mu_{1.1.0j}} & 0 \\ 0 & 0 & 0 & 0 & \frac{\partial h_{ij,1.0.0j}^{(2)}}{\partial \mu_{1.0.0j}} \end{bmatrix},$$

where for $t = 0, 1$,

$$\frac{\partial h^{(2)}_{ij,tj}}{\partial \mu_{tj}} = -W_{Dij} W_{Rij} R_{ij} I\left(T_{ij} = t\right) D_{ij} I(S_{ij} = j),$$

$$\frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \mu_{1.0.1j}} = -W_{Dij} W_{Rij} W_{V0ij} R_{ij} T_{ij} D_{ij} I(S_{ij} = j),$$

$$\frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \mu_{1.1.0j}} = -W_{Dij} W_{Rij} W_{E0ij} R_{ij} T_{ij} D_{ij} I(S_{ij} = j),$$

$$\frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \mu_{1.0.0j}} = -W_{Dij} W_{Rij} W_{V0ij} W_{E0ij} R_{ij} T_{ij} D_{ij} I(S_{ij} = j);$$

$$G_{21} = E\left[\frac{\partial h^{(2)}_{ij}}{\partial \eta}\right] = E\begin{bmatrix} G_{211} \\ \vdots \\ G_{21j} \\ \vdots \\ G_{21J} \end{bmatrix},$$

where

$$G_{21j} = E\left[\frac{\partial h^{(2)}_{ij}}{\partial \eta}\right] = E\begin{bmatrix} \frac{\partial h^{(2)}_{ij,0j}}{\partial \eta'_R} & \mathbf{0} \\ \frac{\partial h^{(2)}_{ij,1j}}{\partial \eta'_R} & \mathbf{0} \\ \frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \eta'_R} & \frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \eta'_M} \\ \frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \eta'_R} & \frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \eta'_M} \\ \frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta'_R} & \frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta'_M} \end{bmatrix}$$

$$= E\begin{bmatrix} \frac{\partial h^{(2)}_{ij,0j}}{\partial \eta^{(D)'}_{R0}} & \mathbf{0} & \frac{\partial h^{(2)}_{ij,0j}}{\partial \eta^{(N)'}_{R0}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\partial h^{(2)}_{ij,1j}}{\partial \eta^{(D)'}_{R1}} & \mathbf{0} & \frac{\partial h^{(2)}_{ij,1j}}{\partial \eta^{(N)'}_{R1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \eta^{(D)'}_{R1}} & \mathbf{0} & \frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \eta^{(N)'}_{R1}} & \frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \eta'_{V0}} & \frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \eta'_{V1}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \eta^{(D)'}_{R1}} & \mathbf{0} & \frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \eta^{(N)'}_{R1}} & \mathbf{0} & \mathbf{0} & \frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \eta'_{E0}} & \frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \eta'_{E1}} \\ \mathbf{0} & \frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta^{(D)'}_{R1}} & \mathbf{0} & \frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta^{(N)'}_{R1}} & \frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta'_{V0}} & \frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta'_{V1}} & \frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta'_{E0}} & \frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta'_{E1}} \end{bmatrix},$$

in which

$$\frac{\partial h^{(2)}_{ij,tj}}{\partial \eta^{(D)'}_{Rt}} = W_{Dij}\left(Y_{ij} - \mu_{tj}\right) R_{ij} I\left(T_{ij} = t\right) D_{ij} I(S_{ij} = j) \frac{\partial W_{Rij}}{\partial \eta^{(D)'}_{Rt}}, \frac{\partial W_{Rij}}{\partial \eta^{(D)'}_{Rt}} = \left(\frac{\partial W_{Rij}}{\partial \boldsymbol{\pi}^{(D)'}_{Rt}}, \frac{\partial W_{Rij}}{\partial \sigma^{(D)}_{Rt}}\right)$$

$$\frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \eta^{(D)'}_{R1}} = W_{Dij} W_{V0ij}\left(Y_{ij} - \mu_{1.0.1j}\right) R_{ij} T_{ij} D_{ij} I(S_{ij} = j) \frac{\partial W_{Rij}}{\partial \eta^{(D)'}_{R1}},$$

$$\frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \eta^{(D)'}_{R1}} = W_{Dij} W_{E0ij}\left(Y_{ij} - \mu_{1.1.0j}\right) R_{ij} T_{ij} D_{ij} I(S_{ij} = j) \frac{\partial W_{Rij}}{\partial \eta^{(D)'}_{R1}},$$

$$\frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta^{(D)'}_{R1}} = W_{Dij} W_{V0ij} W_{E0ij}\left(Y_{ij} - \mu_{1.0.0j}\right) R_{ij} T_{ij} D_{ij} I(S_{ij} = j) \frac{\partial W_{Rij}}{\partial \eta^{(D)'}_{R1}},$$

$$\frac{\partial h^{(2)}_{ij,tj}}{\partial \eta^{(N)'}_{Rt}} = W_{Dij}\left(Y_{ij} - \mu_{tj}\right) R_{ij} I\left(T_{ij} = t\right) D_{ij} I(S_{ij} = j) \frac{\partial W_{Rij}}{\partial \eta^{(N)'}_{Rt}}, \frac{\partial W_{Rij}}{\partial \eta^{(N)'}_{Rt}} = \left(\frac{\partial W_{Rij}}{\partial \pi^{(N)'}_{Rt}}, \frac{\partial W_{Rij}}{\partial \sigma^{(N)}_{Rt}}\right)$$

$$\frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \eta^{(N)'}_{R1}} = W_{Dij} W_{V0ij}\left(Y_{ij} - \mu_{1.0.1j}\right) R_{ij} T_{ij} D_{ij} I(S_{ij} = j) \frac{\partial W_{Rij}}{\partial \eta^{(N)'}_{R1}},$$

$$\frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \eta^{(N)'}_{R1}} = W_{Dij} W_{E0ij}\left(Y_{ij} - \mu_{1.1.0j}\right) R_{ij} T_{ij} D_{ij} I(S_{ij} = j) \frac{\partial W_{Rij}}{\partial \eta^{(N)'}_{R1}},$$

$$\frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta^{(N)'}_{R1}} = W_{Dij} W_{V0ij} W_{E0ij}\left(Y_{ij} - \mu_{1.0.0j}\right) R_{ij} T_{ij} D_{ij} I(S_{ij} = j) \frac{\partial W_{Rij}}{\partial \eta^{(N)'}_{R1}},$$

$$\frac{\partial h^{(2)}_{ij,1.0.1j}}{\partial \eta'_{Vt}} = W_{Dij}W_{Rij}\left(Y_{ij} - \mu_{1.0.1j}\right)R_{ij}T_{ij}D_{ij}I\left(S_{ij}=j\right)\frac{\partial W_{V0ij}}{\partial \eta'_{Vt}}, \frac{\partial W_{V0ij}}{\partial \eta'_{Vt}}$$
$$= \left(\frac{\partial W_{V0ij}}{\partial \boldsymbol{\pi}'_{Vt}}, \frac{\partial W_{V0ij}}{\partial \sigma_{Vt}}\right)$$

$$\frac{\partial h^{(2)}_{ij,1.1.0j}}{\partial \eta'_{Et}} = W_{Dij}W_{Rij}\left(Y_{ij} - \mu_{1.1.0j}\right)R_{ij}T_{ij}D_{ij}I\left(S_{ij}=j\right)\frac{\partial W_{E0ij}}{\partial \eta'_{Et}}, \frac{\partial W_{E0ij}}{\partial \eta'_{Et}}$$
$$= \left(\frac{\partial W_{E0ij}}{\partial \boldsymbol{\pi}'_{Et}}, \frac{\partial W_{E0ij}}{\partial \sigma_{Et}}\right)$$

$$\frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta'_{Vt}} = W_{Dij}W_{Rij}W_{E0ij}\left(Y_{ij} - \mu_{1.0.0j}\right)R_{ij}T_{ij}D_{ij}I\left(S_{ij}=j\right)\frac{\partial W_{V0ij}}{\partial \eta'_{Vt}}$$

$$\frac{\partial h^{(2)}_{ij,1.0.0j}}{\partial \eta'_{Et}} = W_{Dij}W_{Rij}W_{V0ij}\left(Y_{ij} - \mu_{1.0.0j}\right)R_{ij}T_{ij}D_{ij}I\left(S_{ij}=j\right)\frac{\partial W_{E0ij}}{\partial \eta'_{Et}}$$

where

$$W_{Rij} = \frac{p^{(N)}_{Rtij}}{p^{(D)}_{Rtij}},$$

$$W_{M0ij} = M_{ij}\frac{p_{M0ij}}{p_{M1ij}} + \left(1 - M_{ij}\right)\frac{1 - p_{M0ij}}{1 - p_{M1ij}}, M \text{ is shorthand for } V \text{ or } E.$$

Hence,

$$\frac{\partial W_{Rij}}{\partial \boldsymbol{\pi}^{(D)'}_{Rt}} = -\frac{p^{(N)}_{Rtij}}{p^{(D)2}_{Rtij}}\frac{\partial p^{(D)}_{Rtij}}{\partial \boldsymbol{\pi}^{(D)'}_{Rt}} = -\frac{p^{(N)}_{Rtij}\left(1 - p^{(D)}_{Rtij}\right)}{p^{(D)}_{Rtij}}\mathbf{X}'_{ij};$$

$$\frac{\partial W_{Rij}}{\partial \sigma^{(D)}_{Rt}} = -\frac{p^{(N)}_{Rtij}}{p^{(D)2}_{Rtij}}\frac{\partial p^{(D)}_{Rtij}}{\partial \sigma^{(D)}_{Rt}} = -\frac{p^{(N)}_{Rtij}\left(1 - p^{(D)}_{Rtij}\right)}{p^{(D)}_{Rtij}}\theta^{(D)}_{Rtj};$$

$$\frac{\partial W_{Rij}}{\partial \pi^{(N)'}_{Rt}} = \frac{1}{p^{(D)}_{Rtij}}\frac{\partial p^{(N)}_{Rtij}}{\partial \pi^{(N)'}_{Rt}} = \frac{p^{(N)}_{Rtij}\left(1 - p^{(N)}_{Rtij}\right)}{p^{(D)}_{Rtij}};$$

$$\frac{\partial W_{Rij}}{\partial \sigma^{(N)}_{Rt}} = \frac{1}{p^{(D)}_{Rtij}}\frac{\partial p^{(N)}_{Rtij}}{\partial \sigma^{(N)}_{Rt}} = \frac{p^{(N)}_{Rtij}\left(1 - p^{(N)}_{Rtij}\right)}{p^{(D)}_{Rtij}}\theta^{(N)}_{Rtj};$$

$$\frac{\partial W_{M0ij}}{\partial \boldsymbol{\pi}'_{M0}} = \left[\frac{M_{ij}}{p_{M1ij}} - \frac{1 - M_{ij}}{1 - p_{M1ij}}\right]\frac{\partial p_{M0ij}}{\partial \boldsymbol{\pi}'_{M0}} = \left[\frac{M_{ij}}{p_{M1ij}} - \frac{1 - M_{ij}}{1 - p_{M1ij}}\right]p_{M0ij}\left(1 - p_{M0ij}\right)\mathbf{X}'_{ij};$$

$$\frac{\partial W_{M0ij}}{\partial \sigma_{M0}} = \left[\frac{M_{ij}}{p_{M1ij}} - \frac{1-M_{ij}}{1-p_{M1ij}}\right]\frac{\partial p_{M0ij}}{\partial \sigma_{MO}} = \left[\frac{M_{ij}}{p_{M1ij}} - \frac{1-M_{ij}}{1-p_{M1ij}}\right] p_{M0ij}\left(1-p_{M0ij}\right)\theta_{M0j};$$

$$\begin{aligned}\frac{\partial W_{M0ij}}{\partial \boldsymbol{\pi}'_{M1}} &= \left[-M_{ij}\frac{p_{M0ij}}{\left(p_{M1ij}\right)^2} + \left(1-M_{ij}\right)\frac{1-p_{M0ij}}{\left(1-p_{M1ij}\right)^2}\right]\frac{\partial p_{M1ij}}{\partial \boldsymbol{\pi}'_{M1}} \\ &= \left[-M_{ij}\frac{p_{M0ij}}{\left(p_{M1ij}\right)^2} + \left(1-M_{ij}\right)\frac{1-p_{M0ij}}{\left(1-p_{M1ij}\right)^2}\right] p_{M1ij}\left(1-p_{M1ij}\right)\mathbf{X}'_{ij} \\ &= \left[-M_{ij}\frac{p_{M0ij}}{p_{M1ij}}\left(1-p_{M1ij}\right) + \left(1-M_{ij}\right)\frac{1-p_{M0ij}}{1-p_{M1ij}}p_{M1ij}\right]\mathbf{X}'_{ij} \\ \frac{\partial W_{M0ij}}{\partial \sigma_{M1}} &= \left[-M_{ij}\frac{p_{M0ij}}{p_{M1ij}}\left(1-p_{M1ij}\right) + \left(1-M_{ij}\right)\frac{1-p_{M0ij}}{1-p_{M1ij}}p_{M1ij}\right]\theta M_{1j}.\end{aligned}$$

We estimate H with $\widehat{H} = \frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{n_j}\widehat{h}_{ij}\widehat{h}'_{ij}$, and estimate G with $\widehat{G} = \frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{n_j}\frac{\partial h_{ij}}{\partial \vartheta}|\widehat{\vartheta}$. According to Lemma 3.3 of Hansen (1982), $\text{plim}\widehat{G}^{-1}\widehat{H}(\widehat{G}^{-1})' = G^{-1}H(G^{-1})$. We thus obtain the consistent estimator of the asymptotic sampling variance of the estimators in the two steps.

Population Average Effect Estimation

Having estimated the site-specific mean potential outcomes, we now estimate each site-specific causal effect through mean contrasts of potential outcomes, as represented in Table 2. We then estimate each population average causal effect for the population of sites by averaging the corresponding site-specific effect estimates. If the sampling probabilities for the sites are unequal, a site-level sample weight will be needed to adjust for the site-level sample selection. The site-level sample weight is 1 for all sites in the current study given its sample design. Therefore, our estimator of each population average causal effect is a simple average of the corresponding site-specific effect estimates over all the J sites in the sample. Denoting the vector of the population average causal effect estimates and that of the causal effect estimates at site j respectively as $\widehat{\gamma}$ and $\widehat{\beta}_j$, we have

$$\widehat{\gamma} = \frac{1}{J}\sum_{j=1}^{J}\widehat{\beta}_j. \tag{1}$$

We then estimate the sampling variance of the population average effect estimates by integrating the sampling variance of the site-specific mean potential outcomes as derived above. Let the vector of the site-specific causal effect estimates across all the sites be

$$\widehat{\beta} = \left(\widehat{\beta}'_1, \ldots, \widehat{\beta}'_J\right)'.$$

Based on Table 2, it can be written as

$$\widehat{\beta} = [I_J \otimes \Phi]\,\widehat{\mu}, \text{where } \Phi = \begin{pmatrix} 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & -1 & -1 & 1 \end{pmatrix},$$

in which $\beta_j = (\beta_j^{(I.V)}(0), \beta_j^{(I.E)}(1), \beta_j^{(D)}(0), \beta_j^{(I.V)}(1), \beta_j^{(I.E)}(0), \beta_j^{(I.V\times E)})'$, μ is defined above, and I_J is a $J \times J$ identity matrix. Similarly, $\beta = [I_J \otimes \Phi]\mu$. We could then obtain the variance matrix of the site-specific effects,

$$\text{var}\left(\widehat{\beta} - \beta\right) = \left[I_J \otimes \Phi\right] \text{var}\left(\widehat{\mu} - \mu\right) \left[I_J \otimes \Phi'\right],$$

in which $\text{var}(\widehat{\beta} - \beta)$ is a $6J \times 6J$ matrix, with $\text{var}(\widehat{\beta}_j - \beta_j)$ as the jth 6×6 submatrix along the diagonal. The off-diagonal elements $\text{cov}(\widehat{\beta}_j - \beta_j, \widehat{\beta}_{j'} - \beta_{j'})$, where $j \neq j'$, are non-zero due to the use of pooled data from all the sites in estimating the weights. Relying on the consistent estimator of $\text{var}(\widehat{\mu} - \mu)$ from the last section, we obtain the consistent estimator of var $\text{var}(\widehat{\beta} - \beta)$.

The population average effect estimator, $\widehat{\gamma}$, can be equivalently written as

$$\widehat{\gamma} = (\Psi'\Psi)^{-1}\Psi'\widehat{\beta}$$

where $\Psi = 1_J \otimes I_6$, in which 1_J is a $J \times 1$ vector of $1's$, and I_6 is a 6×6 identity matrix.

Correspondingly, the variance matrix of the population average effect estimates is

$$\text{var}\left(\widehat{\gamma}\right) = \left(\Psi'\Psi\right)^{-1}\Psi'\text{var}\left(\widehat{\beta}\right)\Psi\left(\Psi'\Psi\right)^{-1},$$

in which

$$\text{var}\left(\widehat{\beta}\right) = \text{var}\left(\widehat{\beta} - \beta + \beta\right) = \text{var}\left(\widehat{\beta} - \beta\right) + \text{var}\left(\beta\right),$$

where $\text{var}(\beta) = I_J \otimes \text{var}(\beta_j)$. The between-site variance $\text{var}(\beta_j)$ is of key scientific interest. We discuss its estimation in the next section.

Between-Site Variance Estimation

In this section, we adopt the between-site variance estimators derived by Qin and Hong (2017) as follows.

Let $G = \sum_{j=1}^{J}(\widehat{\beta}_j - \widehat{\gamma})(\widehat{\beta}_j - \widehat{\gamma})'$, we can show that

$$\begin{aligned} &E(G) = \textstyle\sum_{j=1}^{J} E[(\widehat{\beta}_j - \gamma) - (\widehat{\gamma} - \gamma)][(\widehat{\beta}_j - \gamma) - (\widehat{\gamma} - \gamma)]' \\ &= \textstyle\sum_{j=1}^{J} E[(\widehat{\beta}_j - \gamma)(\widehat{\beta}_j - \gamma)' - (\widehat{\gamma} - \gamma)(\widehat{\beta}_j - \gamma)' - (\widehat{\beta}_j - \gamma)(\widehat{\gamma} - \gamma)' + (\widehat{\gamma} - \gamma)(\widehat{\gamma} - \gamma)'] \end{aligned}$$

in which

$$\sum_j E\left(\widehat{\beta}_j-\gamma\right)\left(\widehat{\beta}_j-\gamma\right)'=\sum_j \mathrm{var}\left(\widehat{\beta}_j\right)=\sum_j \mathrm{var}\left(\widehat{\beta}_j-\beta_j+\beta_j\right)$$
$$=\sum_j\left(\mathrm{var}\left(\widehat{\beta}_j-\beta_j\right)+\mathrm{var}\left(\beta_j\right)\right);$$

$$\sum_j E\left(\widehat{\gamma}-\gamma\right)\left(\widehat{\beta}_j-\gamma\right)'=\sum_j E\left(\frac{1}{J}\sum_{j'}\widehat{\beta}_{j'}-\gamma\right)\left(\widehat{\beta}_j-\gamma\right)'=\frac{1}{J}\sum_j\sum_{j'}E\left(\widehat{\beta}_{j'}-\gamma\right)\left(\widehat{\beta}_j-\gamma\right)';$$

$$\sum_j E\left(\widehat{\beta}_j-\gamma\right)\left(\widehat{\gamma}-\gamma\right)'=\sum_j E\left(\widehat{\beta}_j-\gamma\right)\left(\frac{1}{J}\sum_{j'}\widehat{\beta}_{j'}-\gamma\right)'=\frac{1}{J}\sum_j\sum_{j'}E\left(\widehat{\beta}_j-\gamma\right)\left(\widehat{\beta}_{j'}-\gamma\right)';$$

$$\sum_j E\left(\widehat{\gamma}-\gamma\right)\left(\widehat{\gamma}-\gamma\right)'=\sum_j E\left(\frac{1}{J}\sum_j\widehat{\beta}_j-\gamma\right)\left(\frac{1}{J}\sum_{j'}\widehat{\beta}_{j'}-\gamma\right)'$$
$$=\frac{1}{J}\sum_j\sum_{j'}E\left(\widehat{\beta}_j-\gamma\right)\left(\widehat{\beta}_{j'}-\gamma\right)',$$

where

$$\sum_j\sum_{j'}E\left(\widehat{\beta}_{j'}-\gamma\right)\left(\widehat{\beta}_j-\gamma\right)'=\sum_j\sum_{j'}E\left[\left(\widehat{\beta}_{j'}-\beta_{j'}\right)+\left(\beta_{j'}-\gamma\right)\right]\left[\left(\widehat{\beta}_j-\beta_j\right)+\left(\beta_j-\gamma\right)\right]'$$
$$=\sum_j\sum_{j'}E\left[\left(\widehat{\beta}_{j'}-\beta_{j'}\right)+\left(\widehat{\beta}_j-\beta_j\right)'\right]+\sum_j\sum_{j'}E\left[\left(\beta_{j'}-\gamma\right)+\left(\beta_j-\gamma\right)'\right]$$
$$=\sum_j\sum_{j'\neq j}\mathrm{cov}\left(\widehat{\beta}_j-\beta_j,\widehat{\beta}_{j'}-\beta_{j'}\right)+\sum_j\mathrm{var}\left(\widehat{\beta}_j-\beta_j\right)+J\mathrm{var}\left(\beta_j\right).$$

Therefore,

$$E(G)=\sum_{j=1}^{J}\left(\mathrm{var}\left(\widehat{\beta}_j-\beta_j\right)+\mathrm{var}\left(\beta_j\right)\right)-\frac{1}{J}\sum_j\sum_{j'}E\left(\widehat{\beta}_{j'}-\gamma\right)\left(\widehat{\beta}_j-\gamma\right)'$$
$$=\sum_{j=1}^{J}\mathrm{var}\left(\widehat{\beta}_j-\beta_j\right)+J\mathrm{var}\left(\beta_j\right)-\frac{1}{J}\sum_j\sum_{j'\neq j}\mathrm{cov}\left(\widehat{\beta}_j-\beta_j,\widehat{\beta}_{j'}-\beta_{j'}\right)$$
$$-\frac{1}{J}\sum_j\mathrm{var}\left(\widehat{\beta}_j-\beta_j\right)-\mathrm{var}\left(\beta_j\right)$$
$$=\frac{J-1}{J}\sum_j\mathrm{var}\left(\widehat{\beta}_j-\beta_j\right)+(J-1)\,\mathrm{var}\left(\beta_j\right)-\frac{1}{J}\sum_j\sum_{j'\neq j}\mathrm{cov}\left(\widehat{\beta}_j-\beta_j,\widehat{\beta}_{j'}-\beta_{j'}\right)$$

Replacing $\mathrm{var}(\widehat{\beta}_j-\beta_j)$ and $\mathrm{cov}(\widehat{\beta}_j-\beta_j,\widehat{\beta}_{j'}-\beta_{j'})$ with the corresponding consistent estimators, as shown in the last section, we obtain the consistent estimator for the between-site variance:

$$\widehat{\mathrm{var}}\left(\beta_j\right)=\frac{1}{J-1}\sum_{j=1}^{J}\left(\widehat{\beta}_j-\widehat{\gamma}\right)\left(\widehat{\beta}_j-\widehat{\gamma}\right)'-\frac{1}{J}\sum_{j=1}^{J}\widehat{\mathrm{var}}\left(\widehat{\beta}_j-\beta_j\right)$$

$$+\frac{1}{J(J-1)}\sum_{j}\sum_{j'\neq j}\widehat{\mathrm{cov}}\left(\widehat{\beta}_j-\beta_j,\widehat{\beta}_{j'}-\beta_{j'}\right).$$

As shown above, subtracting the average sampling variance of the site-specific effect estimators from the total variance of the site-specific effect estimators, we obtain the between-site variance of the true effects. On the right side of the above equation, the first component estimates the total variance of $\widehat{\beta}_j$; the second component estimates the average sampling variance of $\widehat{\beta}_j$. Due to the pooling of data from all the sites in the weight estimation, the covariance among the sampling errors of $\widehat{\beta}_j$'s between sites is likely nonzero. Hence, the third component provides additional adjustment for the nonzero covariance. When a negative variance estimate is obtained, which is known as a Heywood case, we set both the variance estimate and the related covariance estimate to 0.

We further employ a permutation test to conduct hypothesis testing for the between-site variance of each causal effect (Qin & Hong, 2017).

APPENDIX C: PRE-TREATMENT COVARIATES IN THE PROPENSITY SCORE MODELS FOR THE DATA ANALYSIS

Table C1.

Variable Name	Description	Non-response models	Models for ME	Model for MV under the treatment condition	Model for MV under the control condition
Demographic characteristics					
Female	Female	√	√	√	√
age_1617	Age was 16-17 at application	√	√	√	√
age_1819	Age was 18-19 at application	√	√	√	
race_w	Race/ethnicity is White	√	√	√	
race_b	Race/ethnicity is Black	√	√	√	
race_h	Race/ethnicity is Hispanic	√	√	√	√
Otherling	If native language is not English	√	√	√	
Inpers	Lived in in-person areas	√	√	√	√
in57	In 57 areas from which many nonresidential students came from	√	√	√	√
Nonres	Designated for a nonresidential slot	√	√	√	√
area_d	From a dense area	√	√	√	
area_nd	From a nondense area	√	√	√	
jcmsa_1	Residence status is PMSA at baseline	√	√	√	
jcmsa_2	Residence status is MSA at baseline	√	√	√	
jcmsa_12	Residence status is PMSA or MSA at baseline				√
jcmsa_miss	Whether residence status is missing	√			
app_1	Applied to Job Corps in quarter 1	√	√	√	
app_2	Applied to Job Corps in quarter 2	√	√	√	
app_3	Applied to Job Corps in quarter 3	√	√	√	
phaseII_demo_nmiss	Phase II demographics not missing	√	√	√	
Fertility and living arrangements at the baseline interview					
Haschld	Had children at baseline	√	√	√	√
r_head	Sample member is head of household at baseline	√	√	√	
livespou	Lived with spouse or partner at baseline	√	√	√	
publich	Lived in public or rent-subsidized housing at baseline	√	√	√	
hh1_3	Family size is 1-3 at baseline	√	√	√	√
hh4_6	Family size is 4-6 at baseline	√	√	√	
hs_ged	Had high school diploma or GED certificate at baseline	√	√	√	

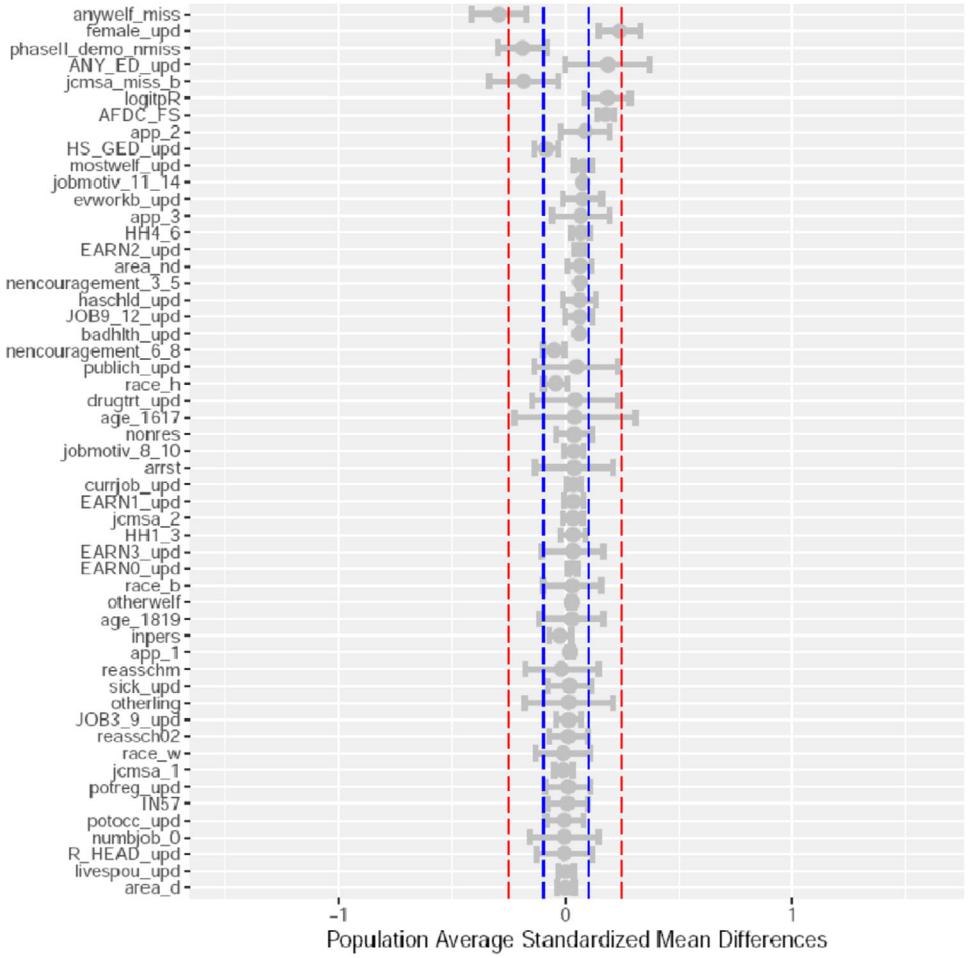
Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

Table C1. Continued

Variable Name	Description	Non-response models	Models for ME	Model for MV under the treatment condition	Model for MV under the control condition
any_ed	Joined any education program in the past year	√	√	√	
reassch02	Reason for leaving school	√	√	√	√
reasschm	Reason for leaving school is missing	√	√	√	
arrst	Ever arrested at baseline	√	√	√	
sick	Had physical or emotional problems that limited the amount of work that could be done at baseline	√	√	√	
badhlth	Bad health at baseline	√	√	√	
potreg	Used Marijuana regularly in the past year	√	√	√	
potocc	Used Marijuana occasionally in the past year	√	√	√	
drugtrt	Ever in drug treatment	√	√	√	
Employment and earnings prior to random assignment					
evworkb	Had full- or part-time job at baseline	√	√	√	
currjob	Employed at baseline	√	√	√	
numbjob_0	No jobs in the past year	√	√	√	
job3_9	Months employed in the past year is 3-9	√	√	√	
job9_12	Months employed in the past year is 9-12	√	√	√	
earn0	Earnings in the past year is 0	√	√	√	
earn1	Earnings in the past year is 0-1000	√	√	√	
earn2	Earnings in the past year is 1,000-5,000	√	√	√	
earn3	Earnings in the past year is 5,000-10,000	√	√	√	
earn4	Earnings in the past year is more than 10,000				√
Public assistance receipt prior to random assignment					
afdc_fs	Received AFDC or food stamps in the past year	√	√	√	√
otherwelf	Received other welfare in the past year	√	√	√	√
anywelf_miss	Whether receiving any welfare in the past year is missing	√	√	√	√
mostwelf	Family was on welfare most of the time when youth was growing up	√	√	√	
Motivation and support for joining Job Corps (JC)					
jobmotiv_8_10	Motivation for joining JC is moderate	√	√	√	
jobmotiv_11_14	Motivation for joining JC is strong	√	√	√	√
nencouragement_3_5	Received moderate support for JC participation	√	√	√	
nencouragement_6_8	Received strong support for JC participation	√	√	√	

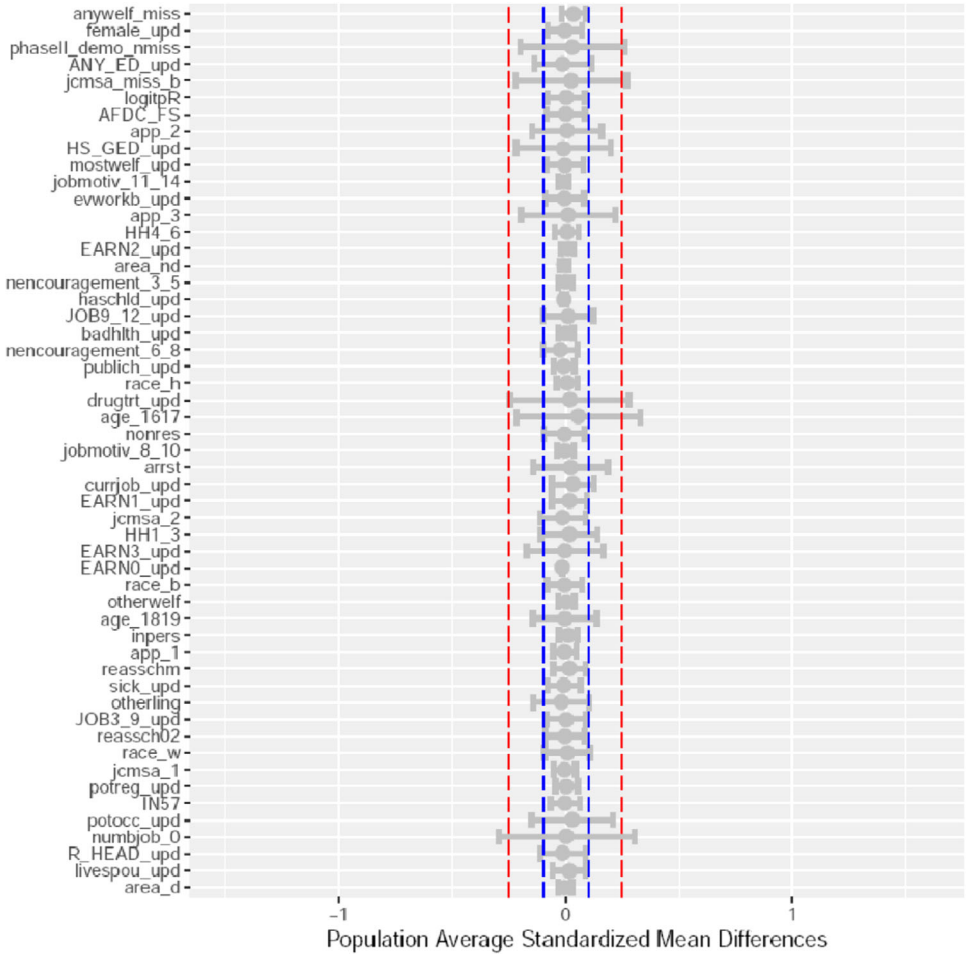
APPENDIX D: BALANCE CHECKING RESULTS

We adopt a weighting-based balance checking procedure to assess if there remains overt bias (bias associated with the observed covariates) after having made the weighting-based adjustment for the selected set of covariates. This appendix displays balance checking results before and after the adjustments.



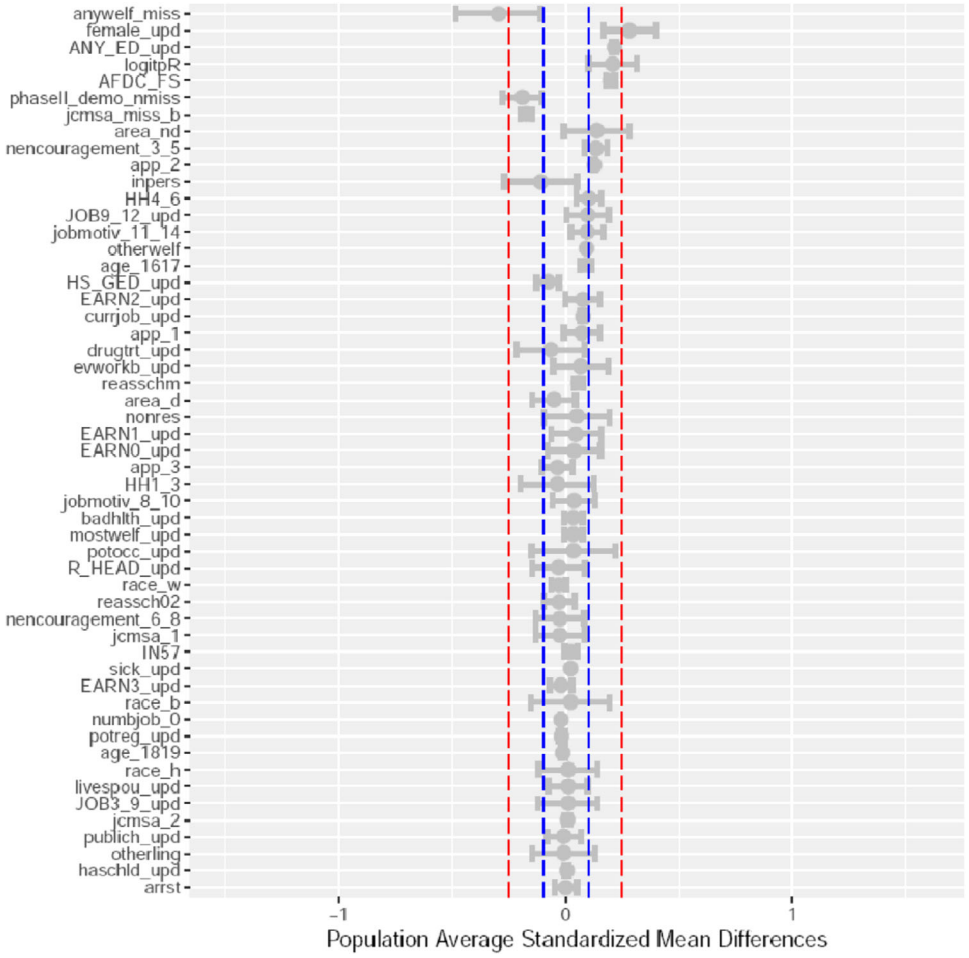
Notes: logitpR is the logit of the propensity score of the response status. Each grey dot indicates the overall mean difference in the corresponding covariate between response levels in the Job Corps group, divided by the pooled standard deviation of the covariate in the Job Corps group. Each grey interval indicates the 95 percent plausible value range of the site-specific mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D1. Imbalance Between Response Levels Before Weighting in the Job Corps Group.



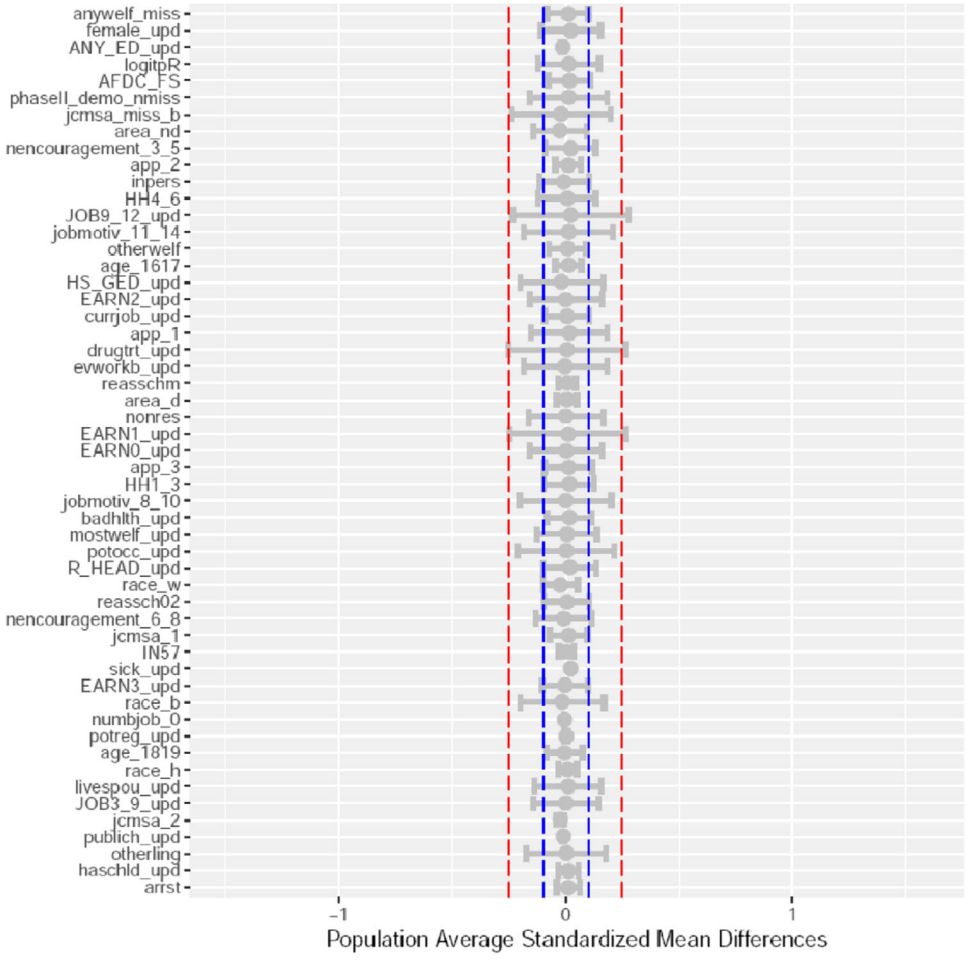
Notes: logitpR is the logit of the propensity score of the response status. Each grey dot indicates the overall weighted mean difference in the corresponding covariate between response levels in the Job Corps group, divided by the pooled standard deviation of the covariate in the Job Corps group. Each grey interval indicates the 95 percent plausible value range of the site-specific weighted mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D2. Imbalance Between Response Levels After Weighting in the Job Corps Group.



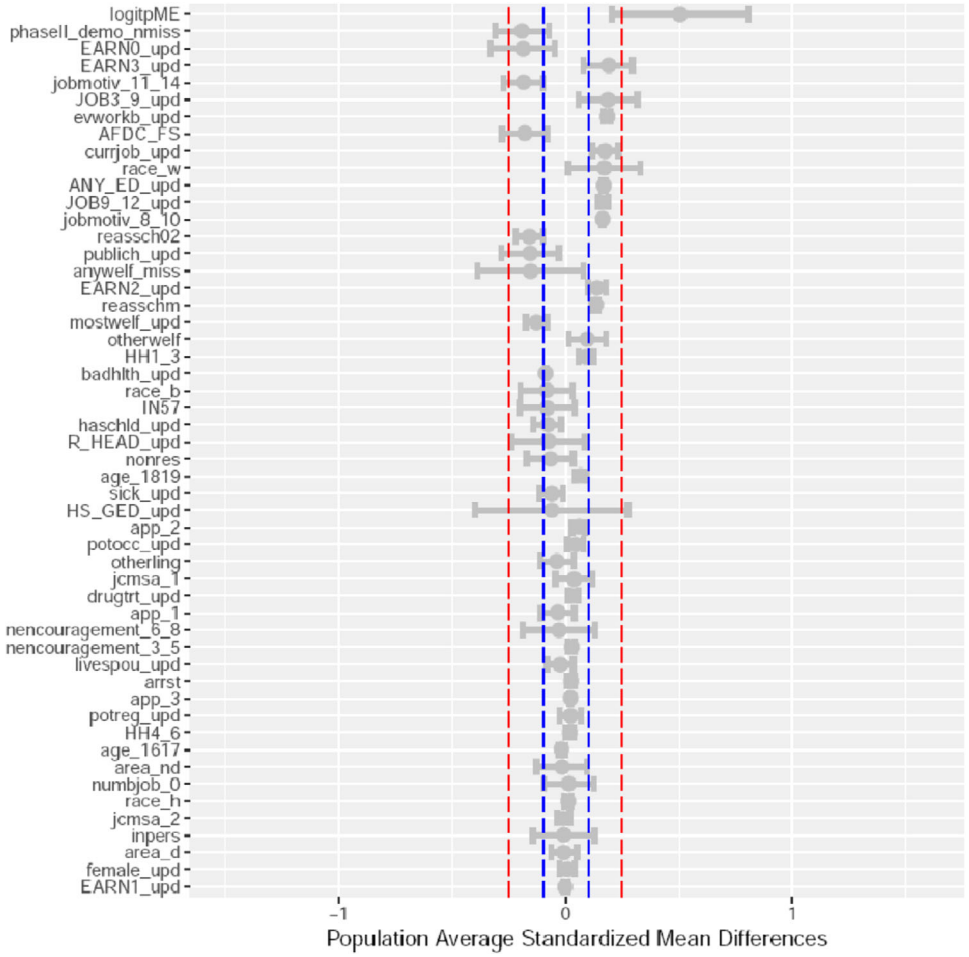
Notes: logitpR is the logit of the propensity score of the response status. Each grey dot indicates the overall mean difference in the corresponding covariate between response levels in the control group, divided by the pooled standard deviation of the covariate in the control group. Each grey interval indicates the 95 percent plausible value range of the site-specific mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D3. Imbalance Between Response Levels Before Weighting in the Control Group.



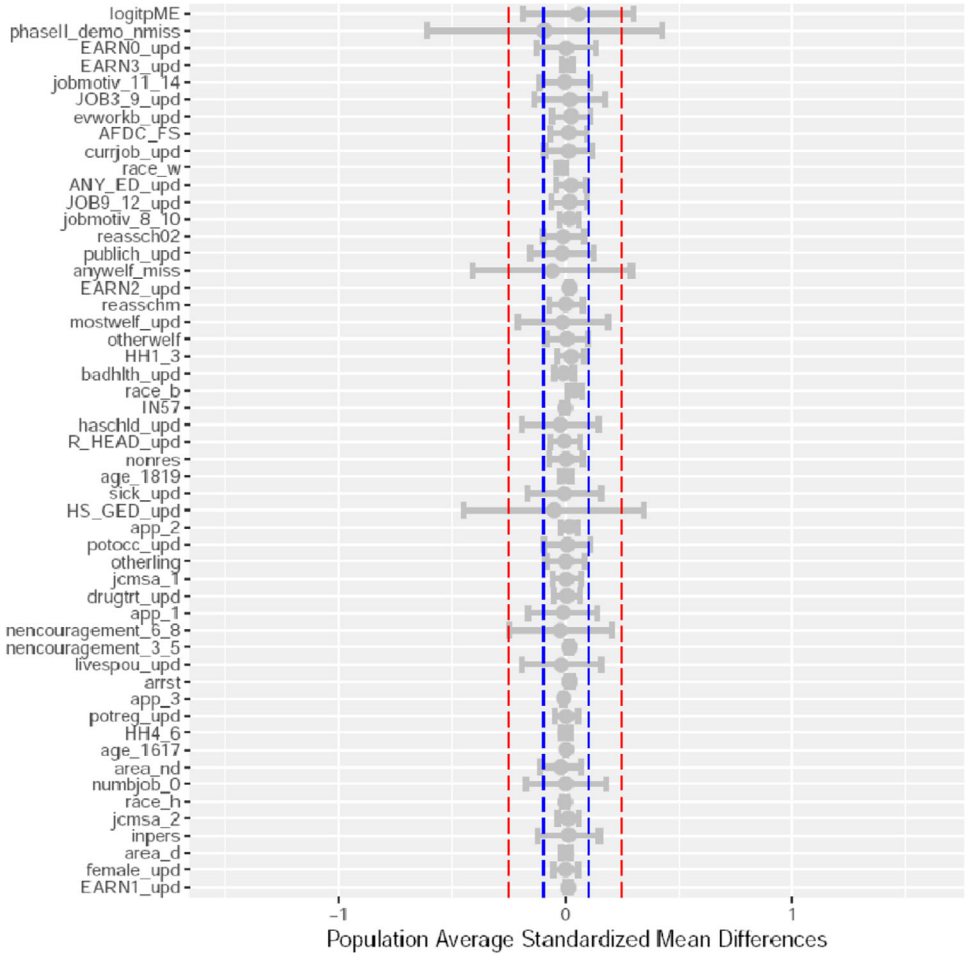
Notes: logitpR is the logit of the propensity score of the response status. Each grey dot indicates the overall weighted mean difference in the corresponding covariate between response levels in the control group, divided by the pooled standard deviation of the covariate in the control group. Each grey interval indicates the 95 percent plausible value range of the site-specific weighted mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D4. Imbalance Between Response Levels After Weighting in the Control Group.



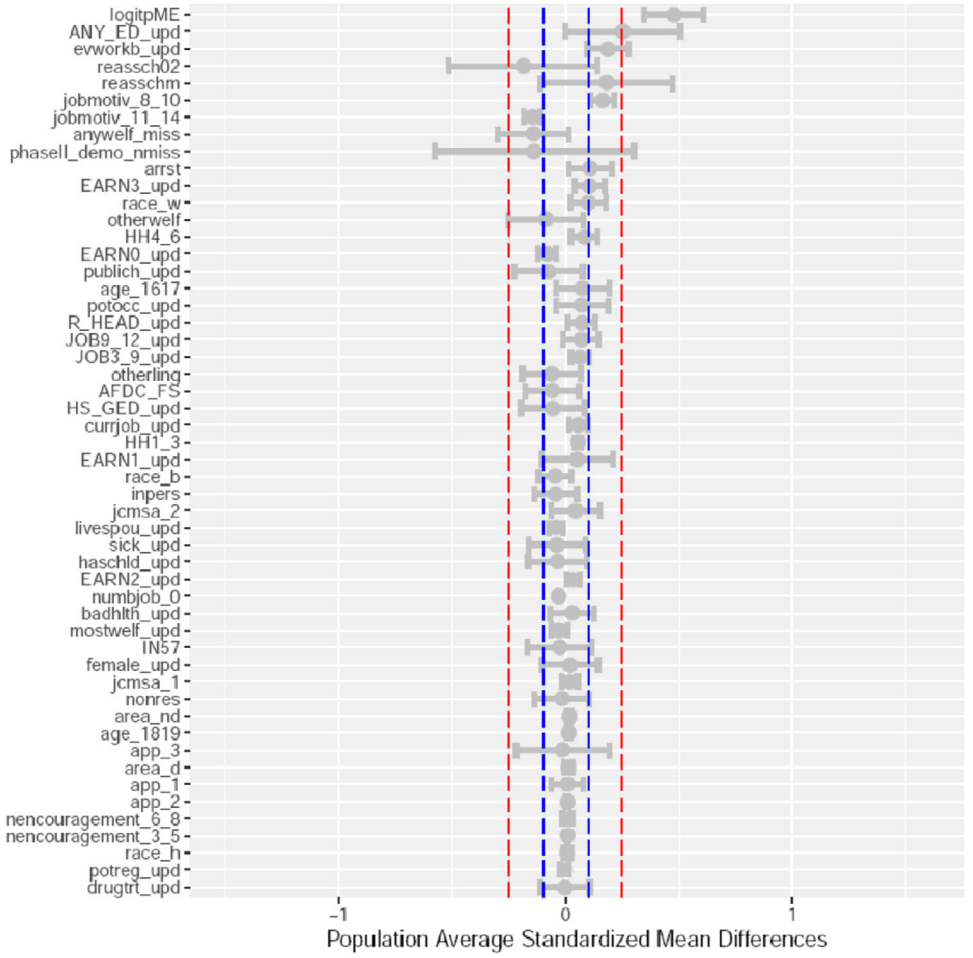
Notes: logitpME is the logit of the propensity score of educational attainment. Each grey dot indicates the overall mean difference in the corresponding covariate between the levels of educational attainment in the Job Corps group, divided by the pooled standard deviation of the covariate in the Job Corps group. Each grey interval indicates the 95 percent plausible value range of the site-specific mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D5. Imbalance Between the Levels of Educational Attainment Before Weighting in the Job Corps Group.



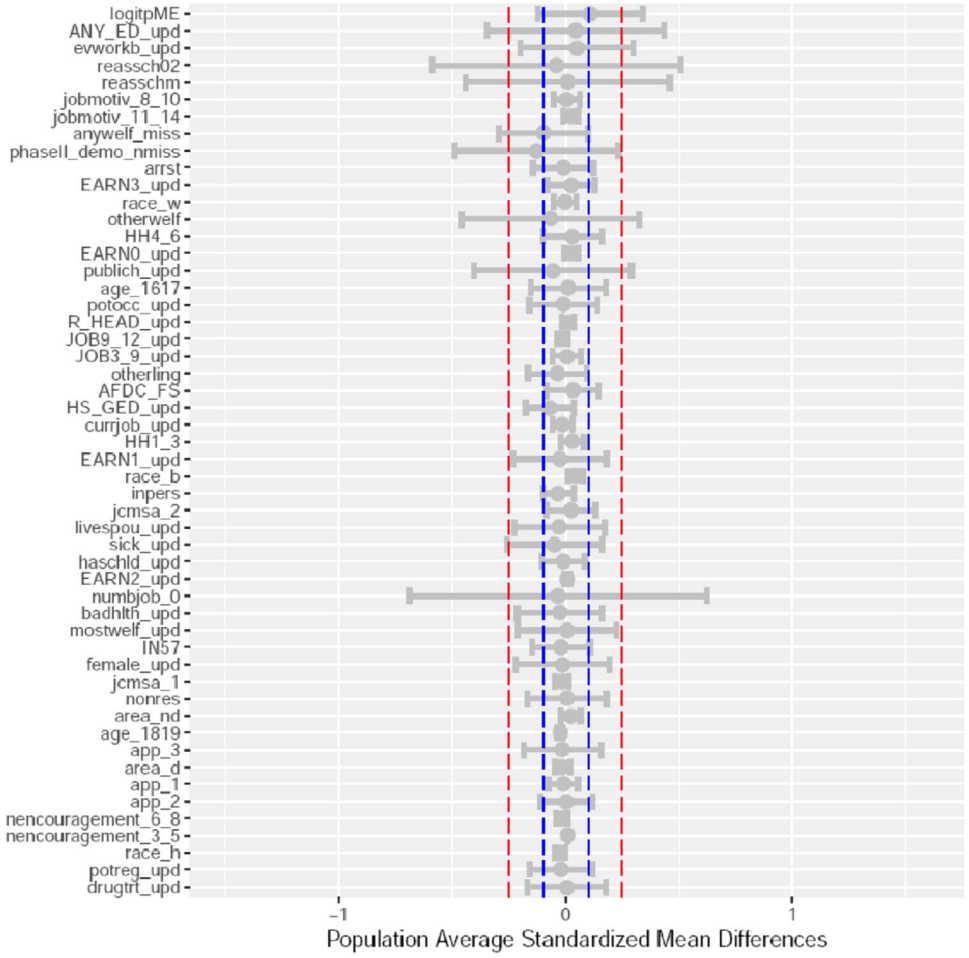
Notes: logitpME is the logit of the propensity score of educational attainment. Each grey dot indicates the overall weighted mean difference in the corresponding covariate between the levels of educational attainment in the Job Corps group, divided by the pooled standard deviation of the covariate in the Job Corps group. Each grey interval indicates the 95 percent plausible value range of the site-specific weighted mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D6. Imbalance Between the Levels of Educational Attainment After Weighting in the Job Corps Group.



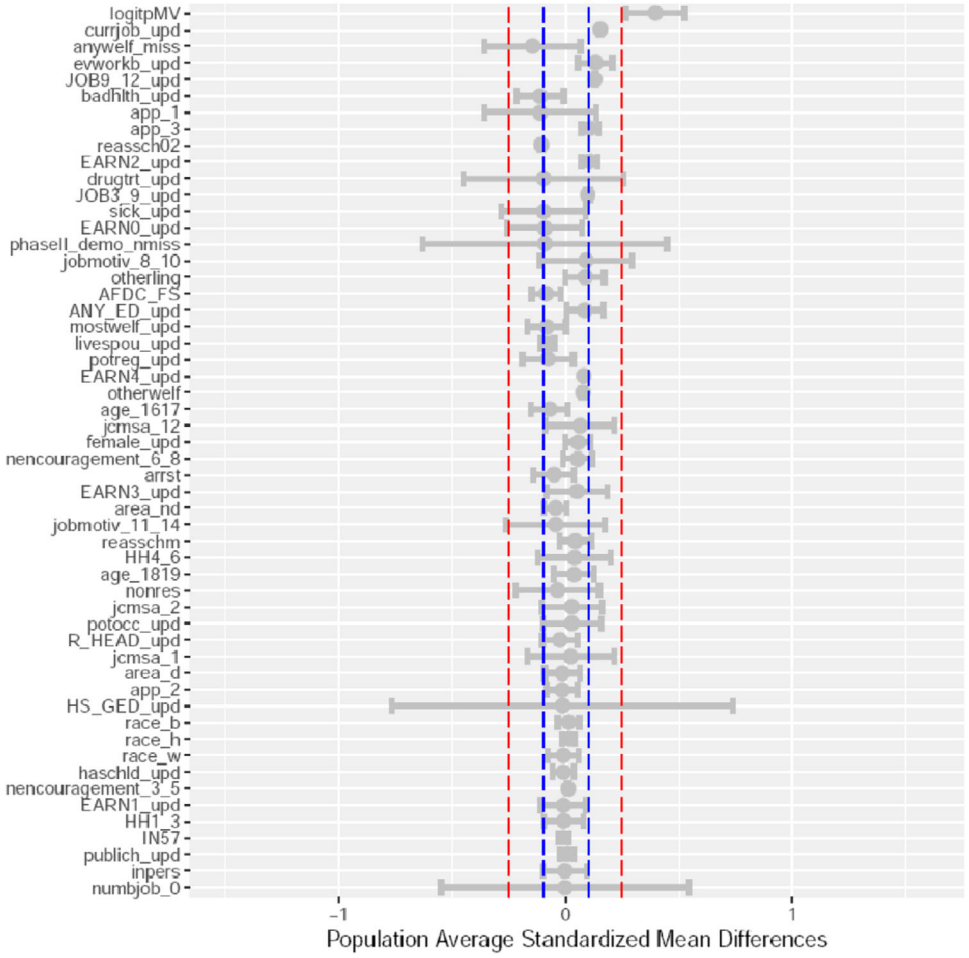
Notes: logitpME is the logit of the propensity score of educational attainment. Each grey dot indicates the overall mean difference in the corresponding covariate between the levels of educational attainment in the control group, divided by the pooled standard deviation of the covariate in the control group. Each grey interval indicates the 95 percent plausible value range of the site-specific mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D7. Imbalance Between the Levels of Educational Attainment Before Weighting in the Control Group.



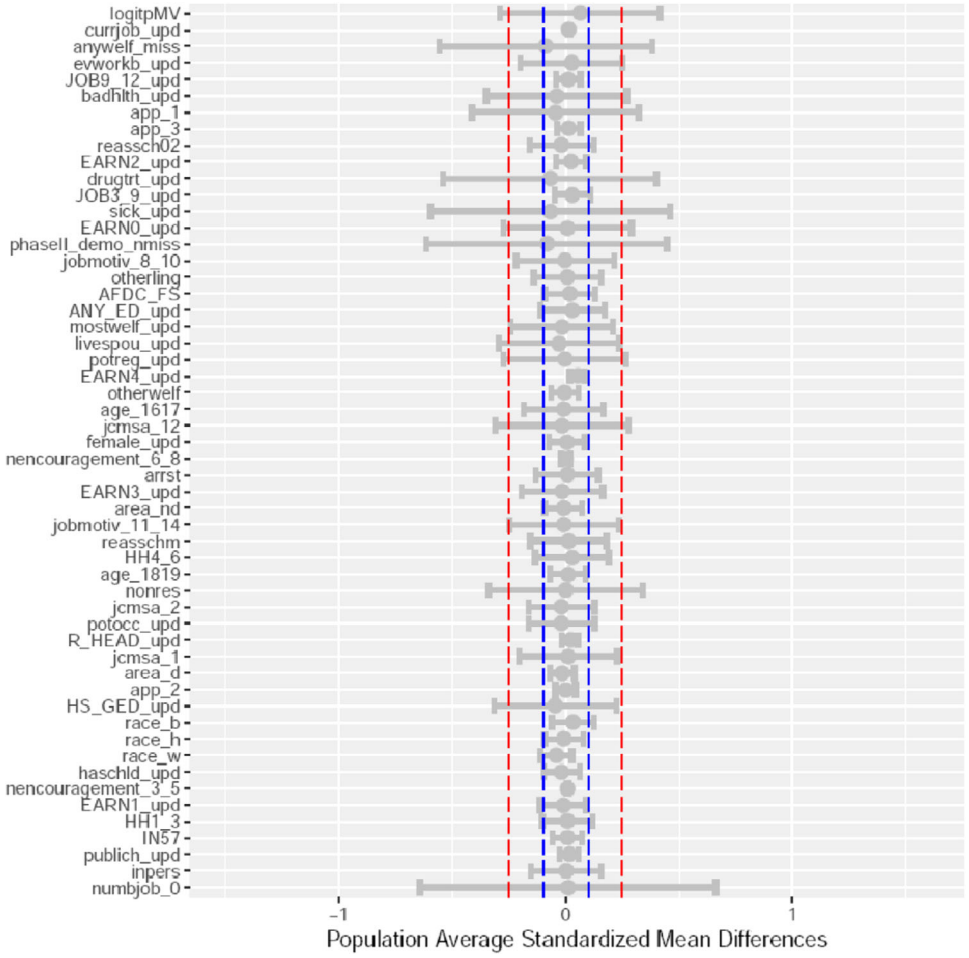
Notes: logitpME is the logit of the propensity score of educational attainment. Each grey dot indicates the overall weighted mean difference in the corresponding covariate between the levels of educational attainment in the control group, divided by the pooled standard deviation of the covariate in the control group. Each grey interval indicates the 95 percent plausible value range of the site-specific weighted mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D8. Imbalance Between the Levels of Educational Attainment After Weighting in the Control Group.



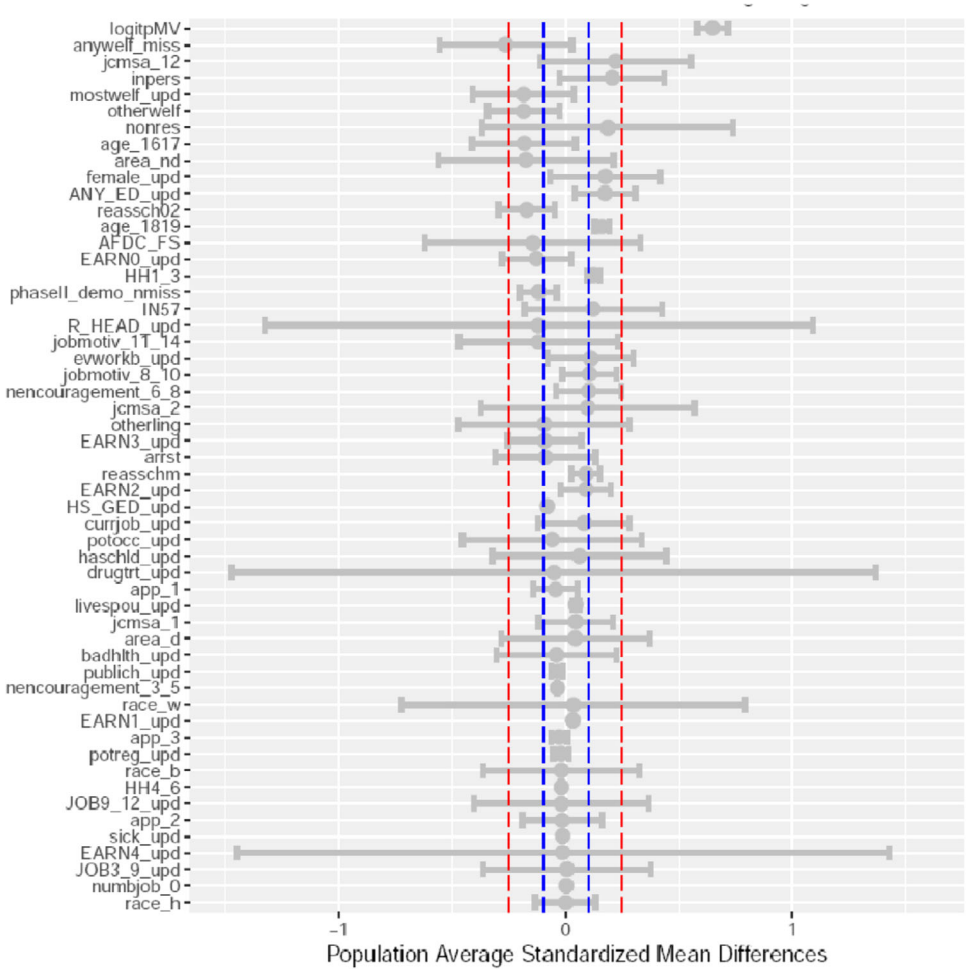
Notes: logitpMV is the logit of the propensity score of vocational attainment. Each grey dot indicates the overall mean difference in the corresponding covariate between the levels of vocational attainment in the Job Corps group, divided by the pooled standard deviation of the covariate in the Job Corps group. Each grey interval indicates the 95 percent plausible value range of the site-specific mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D9. Imbalance Between the Levels of Vocational Attainment Before Weighting in the Job Corps Group.



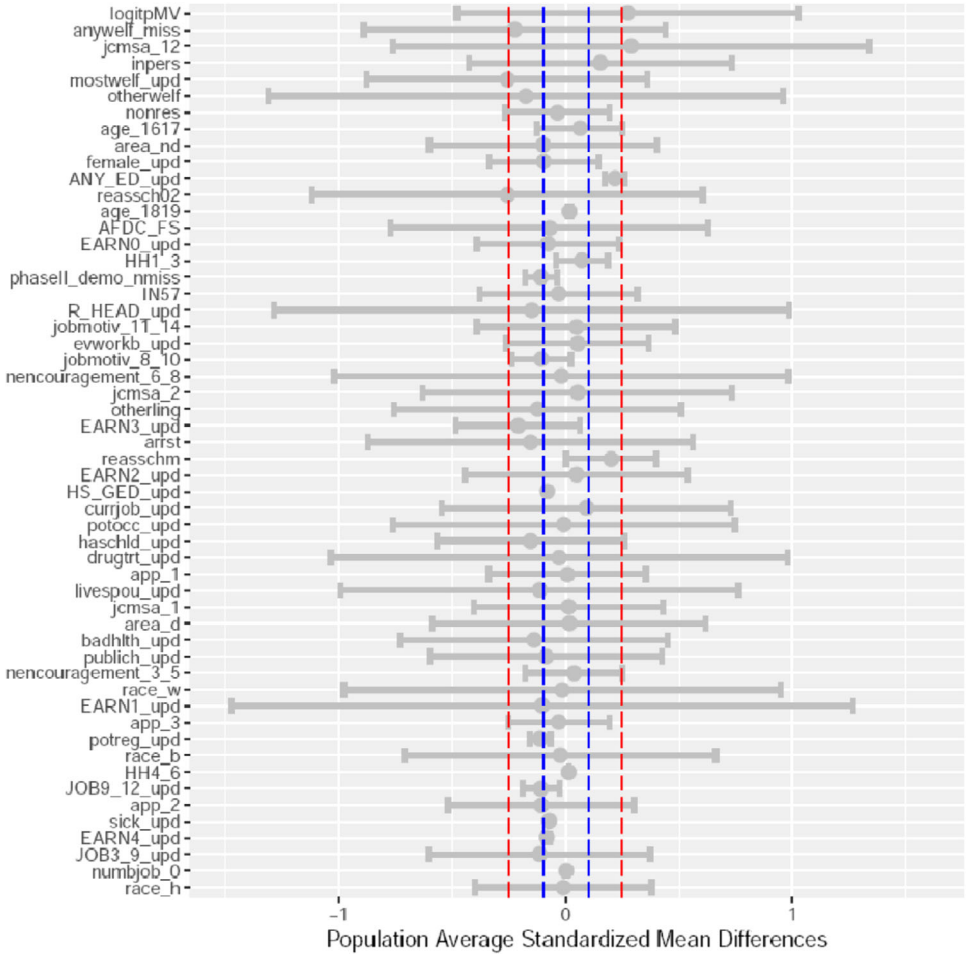
Notes: logitpMV is the logit of the propensity score of vocational attainment. Each grey dot indicates the overall weighted mean difference in the corresponding covariate between the levels of vocational attainment in the Job Corps group, divided by the pooled standard deviation of the covariate in the Job Corps group. Each grey interval indicates the 95 percent plausible value range of the site-specific weighted mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D10. Imbalance Between the Levels of Vocational Attainment After Weighting in the Job Corps Group.



Notes: logitpMV is the logit of the propensity score of vocational attainment. Each grey dot indicates the overall mean difference in the corresponding covariate between the levels of vocational attainment in the control group, divided by the pooled standard deviation of the covariate in the control group. Each grey interval indicates the 95 percent plausible value range of the site-specific mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

Figure D11. Imbalance Between the Levels of Vocational Attainment Before Weighting in the Control Group.



Notes: logitpMV is the logit of the propensity score of vocational attainment. Each grey dot indicates the overall weighted mean difference in the corresponding covariate between the levels of vocational attainment in the control group, divided by the pooled standard deviation of the covariate in the control group. Each grey interval indicates the 95 percent plausible value range of the site-specific weighted mean differences. The red dotted lines indicate the threshold of ±0.25. The blue dotted lines indicate the threshold of ±0.1.

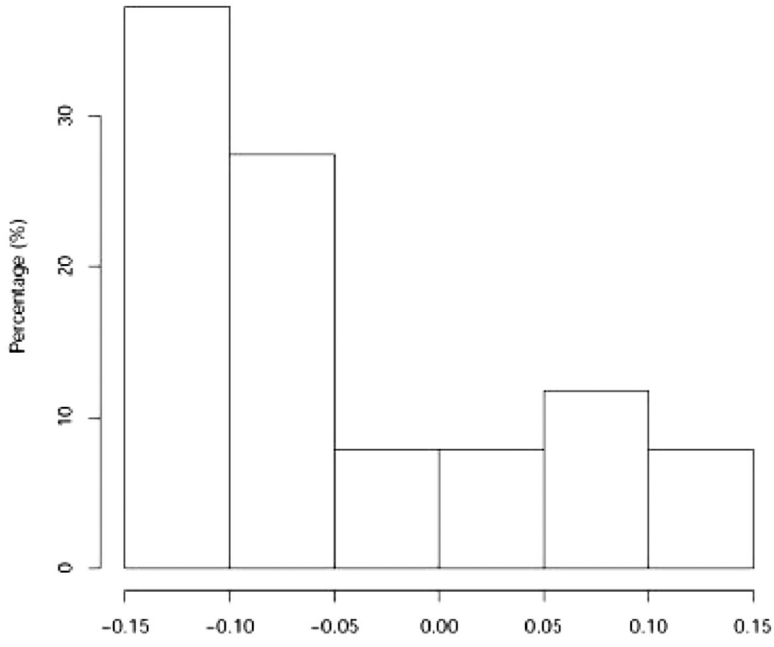
Figure D12. Imbalance Between the Levels of Vocational Attainment After Weighting in the Control Group.

APPENDIX E

This appendix presents the technical details for sensitivity analysis in considering potential violations of Assumptions 3, 4, and 5.

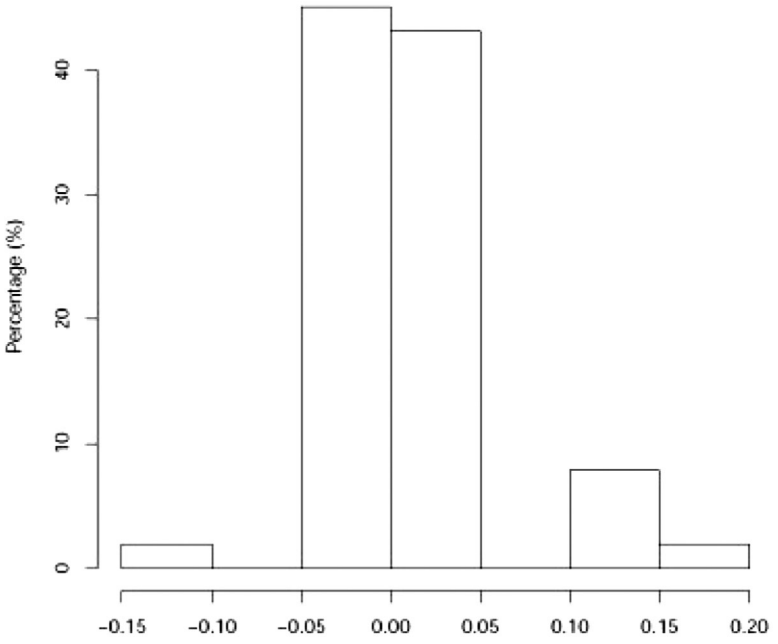
Potential Omissions of Pre-Treatment Confounders

The weighting-based sensitivity analysis for single-site causal mediation analysis summarizes the hidden bias associated with one or more omitted confounders in a function of two sensitivity parameters: one is the standard deviation of the discrepancy between a new weight that adjusts for a confounder and an initial weight that omits the confounder; and the other is the correlation between the weight discrepancy and the outcome within a treatment group. The former is associated with the degree to which the omitted confounder predicts the mediator and the latter is associated with the degree to which it predicts the outcome (Hong, Qin, & Yang, 2018). This new approach to sensitivity analysis has been extended to multisite causal mediation analysis (Qin et al., 2019). In the current study, omitted pretreatment confounders of the response-mediator or response-outcome relationships may potentially violate Assumption 3 (strongly ignorable non-response); and omitted pre-treatment confounders of the mediator-outcome relationships may potentially violate Assumption 4 (strongly ignorable mediator value assignment). Both assumptions need to hold within each site. A violation of Assumption 3 would bias the ITT effect estimate; and a violation of Assumption 3 or 4 would bias the indirect



Note: The first quartile, median, and third quartile of $\gamma^{(T)}$ are -0.12, -0.09, and 0.01, respectively.

Figure E1. Effect Sizes of Bias Values in $\gamma^{(T)}$ Due to Omissions of Pre-Treatment Confounders.



Note: The first quartile, median, and third quartile of $\gamma^{(I.V)}(0)$ are -0.01, 0.00, and 0.02, respectively.

Figure E2. Effect Sizes of Bias Values in $\gamma^{(I.V)}(0)$ Due to Omissions of Pre-Treatment Cofounders.

and direct effect estimates. The weighting-based approach to sensitivity analysis assesses the consequences of such omissions. It quantifies the amount of bias due to the omission by comparing an initial weight with a new weight that adjusts for the omissions.

For example, suppose that an unobserved pre-treatment covariate U has been omitted from the initial analysis. The site-specific NIE, denoted by $\beta_j^{(I)}(1)$, is to be identified as follows when both $\mathbf{X}$ and U are adjusted for in the propensity score models:

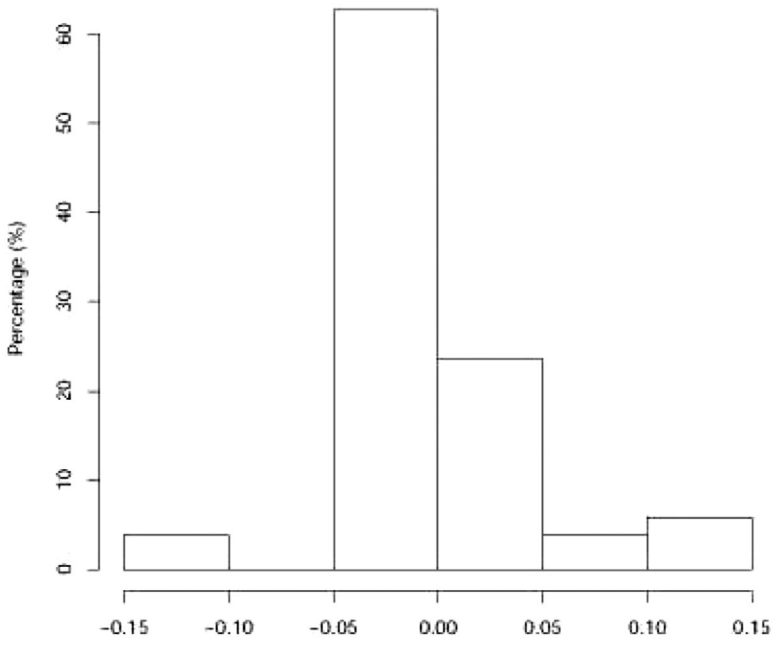
$$E\left[W_{Dij}W_{Tij}W_{U.Rij}Y_{ij}\left|T_{ij}=1,D_{ij}=1,R_{ij}=1,S_{ij}=j\right.\right] - E\left[W_{Dij}W_{Tij}W_{U.Rij}W_{U.Mij}Y_{ij}\left|T_{ij}=1,D_{ij}=1,R_{ij}=1,S_{ij}=j\right.\right],$$

where for individual i in site j,

$$W_{U.Rij}=\frac{Pr\left(R_{ij}=1\left|T_{ij}=1,D_{ij}=1,S_{ij}=j\right.\right)}{Pr\left(R_{ij}=1\left|\mathbf{X}_{ij}=x,U_{ij}=u,T_{ij}=1,D_{ij}=1,S_{ij}=j\right.\right)},$$

$$W_{U.Mij}=\frac{Pr\left(M_{ij}=m\left|R_{ij}=1,T_{ij}=0,D_{ij}=1,\mathbf{X}_{ij}=x,U_{ij}=u,S_{ij}=j\right.\right)}{Pr\left(M_{ij}=m\left|R_{ij}=1,T_{ij}=1,D_{ij}=1,\mathbf{X}_{ij}=x,U_{ij}=u,S_{ij}=j\right.\right)}.$$

If U only confounds the mediator-outcome relationship, the adjusted non-response weight $W_{U.Rij}$ is equal to the initial weight W_{Rij}. Similarly, if U only confounds the response-mediator or response-outcome relationship, the adjusted



Note: The first quartile, median, and third quartile of $\gamma^{(I.E)}(1)$ are -0.03, -0.01, and 0.01, respectively.

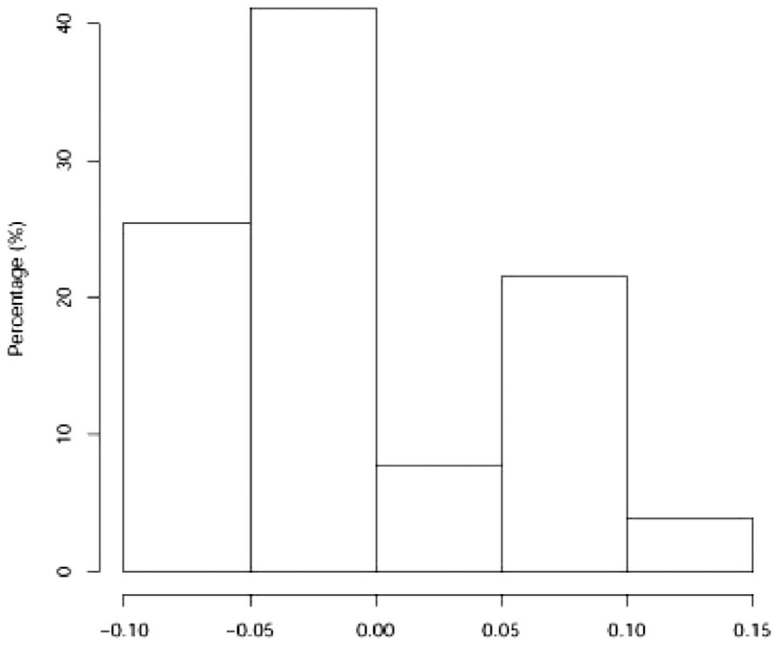
Figure E3. Effect Sizes of Bias Values in $\gamma^{(I.E)}(1)$ Due to Omissions of Pre-Treatment Confounders.

RMPW weight $W_{U.Mij}$ is equal to the initial weight W_{Mij}. In the case that an observed U has been omitted from the initial analysis, the analyst may simply compare the initial result that excludes U with a new result that additionally adjusts for U and determine whether there is a qualitative change in the analytic conclusion. In the case that U is unobserved while theory or past research may indicate that its confounding impact is likely comparable to that of an observed covariate, the analyst may use the latter to generate referent values of bias for the former.

The NJCS data set contains very comprehensive measurements of participants' pre-treatment background. Our analysis has included as many as 51 covariates. Each of the following figures displays, for each estimated causal effect, a distribution of the effect size of bias that would be contributed by each observed pre-treatment covariate if it were to be omitted. It may seem reasonable to speculate that plausible values of effect size of bias associated with some of the unmeasured confounders would likely fall within the range of these empirical distributions.

Omission of a Post-Treatment Confounder

In a randomized trial with noncompliance, a decomposition of the ITT effect into NIE and NDE will likely contain bias. Bias may arise in identifying the population average counterfactual outcomes $E[Y(1, M_V(1), M_E(0))]$, $E[Y(1, M_V(0), M_E(1))]$ and $E[Y(1, M_V(0), M_E(0))]$. We can derive the following result under the ignorability of treatment assignment, the ignorability of mediator value assignment given T, $Z(t)$,

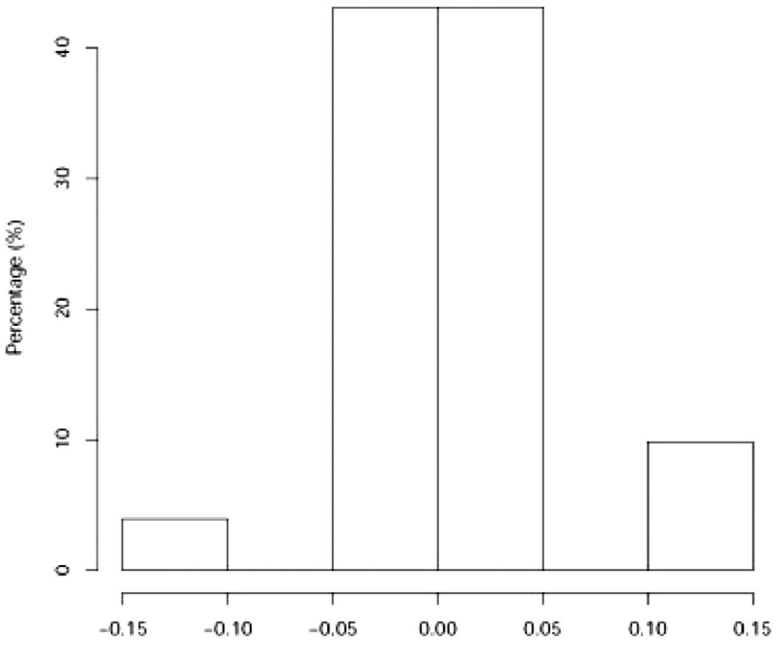


Note: The first quartile, median, and third quartile of $\gamma^{(D)}(0)$ are -0.05, -0.03, and 0.04, respectively.

Figure E4. Effect Sizes of Bias Values in $\gamma^{(D)}(0)$ Due to Omissions of Pre-Treatment Confounders.

and X, and the conditional independence of $M_V(t')$ and $M_V(t'')$. For $t, t', t'' = 0, 1$.

$$
\begin{aligned}
&E\left[Y\left(t, Z(t), M_V\left(t', Z\left(t'\right)\right), M_E\left(t'', Z\left(t''\right)\right)\right)\right] \\
&= E\left[Y\left(t, Z(t), M_V\left(t', Z\left(t'\right)\right), M_E\left(t'', Z\left(t''\right)\right)\right) | T = 1\right] \\
&= E\left\{E\left[Y\left(t, Z(t), M_V\left(t', Z\left(t'\right)\right), M_E\left(t'', Z\left(t''\right)\right)\right) | \mathbf{X} = x\right] T = 1\right\} \\
&= \int\int\int\int\int y f(y|T = t, Z(t) = z, M_V(t', Z(t')) = m_V, M_E(t'', Z(t'')) \\
&= m_E, \mathbf{X} = x) pr\left(M_V\ \left(t', Z\left(t'\right)\right)\right. \\
&= m_V \left| T = t, Z(t) = z, \mathbf{X} = \mathbf{x}\right) pr\left(M_E\ \left(t'', Z\left(t''\right)\right) = m_E \left| T = t, Z(t)\right.\right. \\
&= z, \mathbf{X} = \mathbf{x}) \, pr\left(Z(t) = z \,|T = t, \mathbf{X} = \mathbf{x}\right) f\left(|T = t\right) dydzdm_V dm_E dx \\
&= \int\int\int\int y f(y|T = t, Z(t) = z, M_V(t, Z(t)) = m_V, M_E(t, Z(t)) \\
&= m_E, \mathbf{X} = \mathbf{x}) pr(M_V(t', Z(t')) \\
&= m_V \left| T = t', Z\left(t'\right) = z, \mathbf{X} = \mathbf{x}\right) pr\left(M_E\ \left(t'', Z\left(t''\right)\right) = m_E \left| T = t'', Z\left(t''\right)\right.\right. \\
&= z, \mathbf{X} = \mathbf{x}) \, pr\left(Z(t)) = z \,|T = t, \mathbf{X} = \mathbf{x}\right) f\left(\mathbf{x} | T = t\right) dydzdm_V dm_E dx \\
&= \int\int\int\int\int W_V^* W_E^* y f(y|T = t, Z(t) = z, M_V(t, Z(t)) = m_V, M_E(t, Z(t))
\end{aligned}
$$



Note: The first quartile, median, and third quartile of $\gamma^{(I.V)}(1)$ are -0.01, 0.00, and 0.02, respectively.

Figure E5. Effect Sizes of Bias Values in $\gamma^{(I.V)}(1)$ Due to Omissions of Pre-Treatment Confounders.

$$
\begin{aligned}
&= m_E, \mathbf{X} = \mathbf{x}) pr(M_V(t', Z(t')) \\
&= m_V \left| T = t', Z\left(t'\right) = z, \mathbf{X} = \mathbf{x}\right) pr\left(M_E\left(t'', Z\left(t''\right)\right) = m_E \left| T = t'', Z\left(t''\right)\right.\right. \\
&= z, \mathbf{X} = \mathbf{x}) \, pr\left(Z(t) = z \,|T = t, \mathbf{X} = \mathbf{x}\right) f\left(\mathbf{x} \,|T = t\right) dydzdm_V dm_E dx \\
&= E\left[W_V^* W_E^* Y \,|T = t\right],
\end{aligned}
$$

where

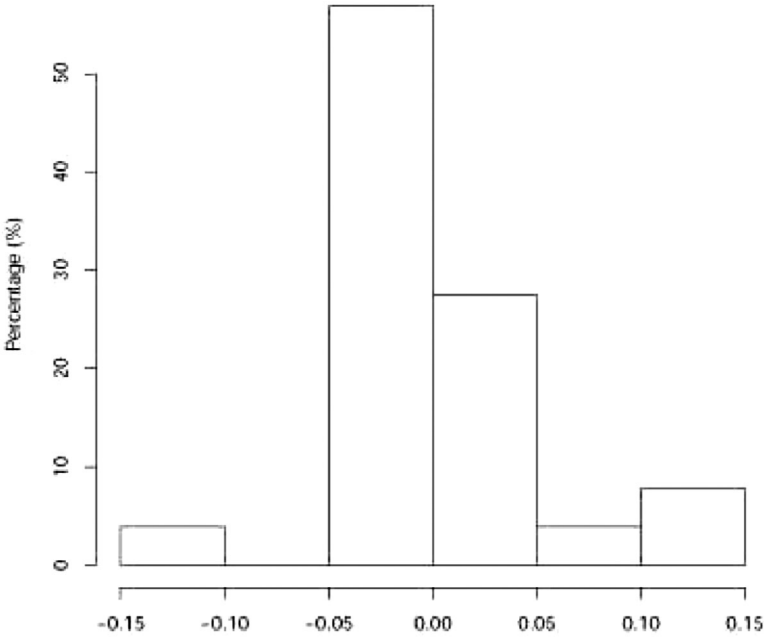
$$
W_V^* = \frac{pr\left(M_V\left(t', Z\left(t'\right)\right) = m_V \left| T = t', Z\left(t'\right) = z, \mathbf{X} = \mathbf{x}\right.\right)}{pr\left(M_V\left(t, Z(t)\right) = m_V \,|T = t, Z(t) = z, \mathbf{X} = \mathbf{x}\right)};
$$

$$
W_E^* = \frac{pr\left(M_E\left(t'', Z\left(t''\right)\right) = m_E \left| T = t'', Z\left(t''\right) = z, \mathbf{X} = \mathbf{x}\right.\right)}{pr\left(M_E\left(t, Z(t)\right) = m_E \,|T = t, Z(t) = z, \mathbf{X} = \mathbf{x}\right)}.
$$

In the case of one-sided noncompliance—that is, noncompliance occurred in only the experimental group but not in the control group, $Z(t')$ is a constant zero when $t' = 0$; and $Z(t'')$ is a constant zero when $t'' = 0$.

In a multisite randomized trial, under the assumptions of ignorable sampling mechanism, ignorable treatment assignment, and strongly ignorable non-response, we can further show that for individual i in site j,

$$
\begin{aligned}
&E\left[Y\left(t, Z(t)\right), M_V\left(t', Z\left(t'\right)\right), M_E\left(t'', Z\left(t''\right)\right)\right] \\
&\quad = E\left[W_{ITT} W_V^* W_E^* Y \,|R = 1, T = t, D = 1\right],
\end{aligned}
$$



Note: The first quartile, median, and third quartile of $\gamma^{(I,E)}(0)$ are -0.03, 0.00, and 0.02, respectively.

Figure E6. Effect Sizes of Bias Values in $\gamma^{(I,E)}(0)$ Due to Omissions of Pre-Treatment Confounders.

where

$$W^*_{Vij} = \frac{pr\left(M_{Vij} = m_V \left| R_{ij} = 1, T_{ij} = 0, D_{ij} = 1, \mathbf{X}_{ij} = \mathbf{x}, S_{ij} = j\right.\right)}{pr\left(M_{Vij} = m_V \left| R_{ij} = 1, T_{ij} = 0, D_{ij} = 1, Z_{ij} = z, \mathbf{X}_{ij} = \mathbf{x}, S_{ij} = j\right.\right)};$$
$$W^*_{Eij} = \frac{pr\left(M_{Eij} = m_E \left| R_{ij} = 1, T_{ij} = 0, D_{ij} = 1, \mathbf{X}_{ij} = \mathbf{x}, S_{ij} = j\right.\right)}{pr\left(M_{Eij} = m_E \left| R_{ij} = 1, T_{ij} = 0, D_{ij} = 1, Z_{ij} = z, \mathbf{X}_{ij} = \mathbf{x}, S_{ij} = j\right.\right)}.$$

After adjusting for compliance behavior, we obtain the sensitivity analysis results summarized in Table E1.

Potential Violations of Assumption 5

Indirect Effects Transmitted Solely via M_V

To disentangle the indirect effect transmitted via M_V and that via M_E, our strategy is to regress M_V on M_E within a treatment group. The regression model may be specified as follows for individual i at site j in treatment t:

$$M_{Vij}(t) = \alpha^{(t)}_{0j} + \alpha^{(t)}_{1j} M_{Eij}(t) + \boldsymbol{\alpha}^{t\prime}_2 \mathbf{X}_{VEij} + \varepsilon_{M_V ij}(t).$$

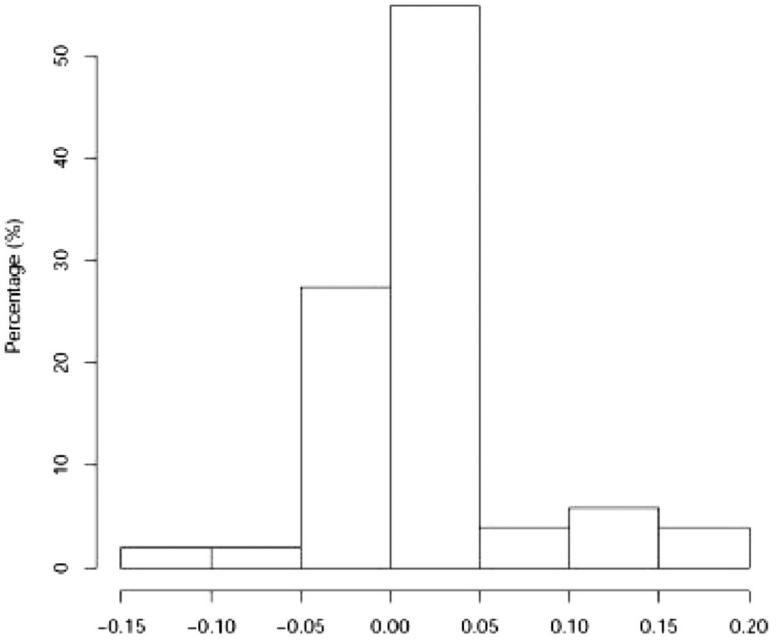
where $\mathbf{X}_{VEij} = \mathbf{X}_{Vij} \cup \mathbf{X}_{Eij}$. Let $M^*_{Vij} = M_{Vij}(t) - \alpha^{(t)}_{0j} - \alpha^{(t)}_{1j} M_{Eij}(t) - \boldsymbol{\alpha}^{t\prime}_2 \mathbf{X}_{VEij}$ be the residual. We have that

$$M_{Eij}(t) \perp\!\!\!\perp M^*_{Vij}(t) | R_{ij} = 1, T_{ij} = t, D_{ij} = 1, \mathbf{X}_{Vij} = \mathbf{x}_V, \mathbf{X}_{Eij} = \mathbf{x}_E, S_{ij} = j,$$

Table E1. Sensitivity analysis for the omission of a post-treatment confounder.

	Population average (dollars)	Between-site standard deviation (dollars)	Population average (dollars)	Between-site standard deviation (dollars)
Total treatment effect on the outcome	21.743 (5.944)	26.379	21.743 (5.944)	26.379
Indirect effect via M_V given $M_E(0)$	3.152 (1.625)	8.690	3.049 (1.616)	7.451
Indirect effect via M_E given $M_V(1)$	7.518 (1.630)	2.666	7.957 (1.619)	0.000
Direct effect	11.073 (6.322)	29.288	10.736 (6.243)	26.585
Indirect effect via M_V given $M_E(1)$	4.050 (1.723)	9.175	4.362 (1.781)	9.127
Indirect effect via M_E given $M_V(0)$	6.620 (1.761)	6.498	6.645 (1.684)	0.000
Interaction effect between M_V and M_E	0.898 (0.700)	0.940	1.313 (0.793)	1.009

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management



Note: The first quartile, median, and third quartile of $\gamma^{(I.V\times E)}$ are 0.00, 0.00, and 0.04, respectively.

Figure E7. Effect Sizes of Bias Values in $\gamma^{(I.V\times E)}$ Due to Omissions of Pre-Treatment Confounders.

which implies the following,

$$M_{Eij}\left(t'\right) \perp\!\!\!\perp M^{*}_{Vij}(t) \left| R_{ij}=1\ , T_{ij}=t, D_{ij}=1, \mathbf{X}_{Vij}=\mathbf{x}_{V}, \mathbf{X}_{Eij}=\mathbf{x}_{E}, S_{ij}=j.\right.$$

The individual-specific indirect effect via M^{*}_{V} without improvement in general education is

$$\beta^{(I.V*)}_{ij}(0) = Y_{ij}\left(1, M^{*}_{Vij}(1), M_{Eij}(0)\right) - Y_{ij}\left(1, M^{*}_{Vij}(0), M_{Eij}(0)\right);$$

and the individual-specific indirect effect via M^{*}_{V} with improvement in general education is

$$\beta^{(I.V*)}_{ij}(1) = Y_{ij}\left(1, M^{*}_{Vij}(1), M_{Eij}(1)\right) - Y_{ij}\left(1, M^{*}_{Vij}(0), M_{Eij}(1)\right).$$

Averaged over all individuals within a site and averaged over all the sites, the alternative parameters $\gamma^{(I.V*)}(0) = E[E(\beta^{(I.V*)}_{ij}(0)|S_{ij}=j)]$ and $\gamma^{(I.V*)}(1) = E[E(\beta^{(I.V*)}_{ij}(1)|S_{ij}=j)]$ define the population average indirect effects transmitted solely through M_V. When Assumption 5 holds, $\gamma^{(I.V*)}(0)$ and $\gamma^{(I.V^{*})}(1)$ converge to $\gamma^{(I.V)}(0)$ and $\gamma^{(I.V)}(1)$, respectively. As before, we employ RMPW to estimate these indirect effects. This involves obtaining

$$E\left[Y_{ij}\left(1, M^{*}_{Vij}(0), M_{Eij}\ (1)\right)\left|S_{ij}=j\right.\right]$$

Table E2. Sensitivity analysis for potential violations of Assumption 5.

	Indirect effect undifferentiated between M_V and M_E is transmitted through M_V		Indirect effect undifferentiated between M_V and M_E is transmitted through M_E	
	Effective Size of Bias	Effect Size of the Adjusted Estimate	Effective Size of Bias	Effect Size of the Adjusted Estimate
Indirect effect via M_V given $M_E(0)$	0.005	0.013	−0.019	0.037
Indirect effect via M_E given $M_V(1)$	0.001	0.042	0.025	0.018
Direct effect	−0.006	0.070	−0.006	0.070
Indirect effect via M_V given $M_E(1)$	0.012	0.011	−0.028	0.051
Indirect effect via M_E given $M_V(0)$	−0.006	0.044	0.034	0.004
Interaction effect between M_V and M_E	0.007	−0.002	−0.009	0.014

$$= E\left[W_{ITTij}W_{V*ij}Y_{ij}\left|R_{ij}=1,T_{ij}=1,D_{ij}=1,S_{ij}=j\right.\right],$$

$$E\left[Y_{ij}\left(1,M^*_{Vij}(0),M_{Eij}\ (0)\right)\left|S_{ij}=j\right.\right]$$

$$'quad = E\left[W_{ITTij}W_{V*ij}W_{Eij}Y_{ij}\left|R_{ij}=1,T_{ij}=1,D_{ij}=1,S_{ij}=j\right.\right],$$

where

$$W_{V*ij}=\frac{Pr\left(M_{V*ij}=m_V|\mathbf{X}_{Vij}=\mathbf{x}_V,R_{ij}=1,T_{ij}=0,D_{ij}=1,S_{ij}=j\right)}{Pr\left(M_{V*ij}=m_V|\mathbf{X}_{Vij}=\mathbf{x}_V,R_{ij}=1,T_{ij}=1,D_{ij}=1,S_{ij}=j\right)}.$$

Indirect Effects Transmitted Solely via M_E

The population average indirect effects transmitted solely via M_E is denoted by $\gamma^{(I.E*)}(t)=E[E(\beta_{ij}^{(I.E*)}(t)|S_{ij}=j)]$ for $t=0,1$ in which

$$\beta_{ij}^{(I.E*)}(0)=Y_{ij}\left(1,M_{Vij}(0),M^*_{Eij}(1)\right)-Y_{ij}\left(1,M_{Vij}(0),M^*_{Eij}(0)\right);$$

$$\beta_{ij}^{(I.E*)}(1)=Y_{ij}\left(1,M_{Vij}(1),M^*_{Eij}(1)\right)-Y_{ij}\left(1,M_{Vij}(1),M^*_{Eij}(0)\right).$$

To obtain M^*_E that is orthogonal to M_V, we regress M_E on M_V and obtain the residual. Specifically, the regression model may be specified as follows for individual i at site j in treatment t:

$$M_{Eij}(t)=\alpha_{0j}^{(t)}+\alpha_{1j}^{(t)}M_{Vij}(t)+\boldsymbol{\alpha}_2^{(t)\prime}\mathbf{X}_{VEij}+\varepsilon_{M_Eij}(t).$$

Let $M^*_{Eij}(t)=M_{Eij}(t)-\alpha_{0j}^{(t)}-\alpha_{1j}^{(t)}M_{Vij}(t)-\boldsymbol{\alpha}_2^{(t)\prime}X_{VEij}$ be the residual. We then obtain

$$E[Y_{ij}(1,M_{Vij}(1),M^*_{Eij}(0))|S_{ij}=j]$$

$$= E[W_{ITTij} W_{E*ij} Y_{ij} | R_{ij} = 1, T_{ij} = 1, D_{ij} = 1, S_{ij} = j],$$

$$E[Y_{ij}(1, M_{Vij}(0), M^*_{Eij}(0)) | S_{ij} = j]$$
$$= E[W_{ITTij} W_{Vij} W_{E*ij} Y_{ij} | R_{ij} = 1, T_{ij} = 1, D_{ij} = 1, S_{ij} = j],$$

where

$$W_{E*ij} = \frac{Pr\left(M_{E*ij} = m_E | \mathbf{X}_{Eij} = \mathbf{x}_E, R_{ij} = 1, T_{ij} = 0, D_{ij} = 1, S_{ij} = j\right)}{Pr\left(M_{E*ij} = m_E | \mathbf{X}_{Eij} = \mathbf{x}_E, R_{ij} = 1, T_{ij} = 1, D_{ij} = 1, S_{ij} = j\right)}.$$

Table E2 summarizes the results of this sensitivity analysis. It displays the effect size of bias and the estimate after adjusting for the bias in each causal effect.

REFERENCES

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica: Journal of the Econometric Society, 50, 1029–1054.

Hedeker, D., & Gibbons, R.D. (2006). Longitudinal data analysis (Vol. 451). Hoboken, NJ: John Wiley & Sons.

Hong, G., Qin, X., & Yang, F. (2018). Weighting-based sensitivity analysis in causal mediation studies. Journal of Educational and Behavioral Statistics, 43, 32–56.

Qin, X., & Hong, G. (2017). A weighting method for assessing between-site heterogeneity in causal mediation mechanism. Journal of Educational and Behavioral Statistics, 42, 308–340.

Qin, X., Hong, G., Deutsch, J., & Bein, E. (2019). Multisite causal mediation analysis in the presence of complex sample and survey designs and non-random attrition. Journal of the Royal Statistical Society: Series A, 182, 1343–1370.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). Sage.

Stroud, A. H., & Secrest, D. (1966). Gaussian quadrature formulas (Vol. 39). Englewood Cliffs, NJ: Prentice-Hall.