Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://amstat.tandfonline.com/loi/uasa20

Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments*

Kosuke Imai, Zhichao Jiang & Anup Malani

To cite this article: Kosuke Imai, Zhichao Jiang & Anup Malani (2020): Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments*, Journal of the American Statistical Association, DOI: 10.1080/01621459.2020.1775612

To link to this article: https://doi.org/10.1080/01621459.2020.1775612

Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments*

Kosuke Imai†, Zhichao Jiang‡, Anup Malani§

Harvard University, Government and Statistics, Cambridge, United States
Harvard University, Cambridge, United States
University of Chicago Law School, Chicago, 60637-2776 United States

*The proposed methodology is implemented via an open-source software package experiment (Imai and Jiang, 2018), which is available at https://cran.r-project.org/package=experiment. We thank Naoki Egami for helpful comments. We thank the Editor, Associate Editor, and three reviewers for careful reading and many constructive comments. This study was partially funded by grants from the Law School, the MacLean Center for Bioethics, and the Becker-Friedman Institute at the University of Chicago, U.S.; the Department for International Development, U.K.; the International Growth Centre, U.K.; the Tata Trusts through the Tata Centre for Development at the University of Chicago; and SRM University, Andhra Pradesh, India.

†Professor, Department of Government and Department of Statistics, Harvard University. 1737 Cambridge Street, Institute for Quantitative Social Science, Cambridge MA 02138. Email: imai@harvard.edu URL: https://imai.fas.harvard.edu

‡Assistant Professor, Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst MA 01003. Email: zhichaojiang@umass.edu

§Lee and Brena Freeman Professor, University of Chicago Law School and Pritzker School of Medicine, Chicago IL 60637, and National Bureau of Economic Research, Cambridge MA 02138.

Corresponding author Kosuke Imai imai@harvard.edu

Abstract

In many social science experiments, subjects often interact with each other and as a result one unit's treatment influences the outcome of another unit. Over the last decade, a significant progress has been made towards causal inference in the presence of such interference between units. Researchers have shown that the two-stage randomization of treatment assignment enables the identification of average direct and spillover effects. However, much of the literature has assumed perfect compliance with treatment assignment. In this paper, we establish the nonparametric identification of the complier average direct and spillover effects in two-stage randomized experiments with interference and noncompliance. In particular, we consider the spillover effect of the treatment assignment on the treatment receipt as well as the spillover effect of the treatment receipt on the outcome. We propose consistent estimators, and derive their randomization-based variances under the stratified interference assumption. We also prove the exact relationships between the proposed randomization-based estimators and the popular two-stage least squares estimators. The proposed methodology is motivated by and applied to our own randomized evaluation of the India's National Health Insurance Program (RSBY), where we find some evidence of spillover effects. The proposed methods are implemented via an open-source software package.

Keywords: complier average causal effects, encouragement design, program evaluation, randomization inference, spillover effects, two-stage least squares

1 Introduction

Early methodological research on causal inference has assumed no interference between units (e.g., Neyman, 1923; Fisher, 1935; Holland, 1986; Rubin, 1990). That is, spillover effects are assumed to be absent. In many social science experiments, however, subjects often interact with each other and as a result one unit's treatment influences the outcome of another unit. Over the last decade, a significant progress has been made towards causal inference in the presence of

such interference between units (e.g., Sobel, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008; Tchetgen Tchetgen and VanderWeele, 2010; Aronow, 2012; Vanderweele *et al.*, 2013; Liu and Hudgens, 2014; Hong, 2015; Forastiere *et al.*, 2016; Aronow and Samii, 2017; Athey *et al.*, 2017; Basse and Feller, 2018).

Much of this literature, however, has not addressed another common feature of social science experiments where some control units decide to take the treatment while others in the treatment group refuse to receive one. Such noncompliance often occurs in these experiments because for ethical and logistical reasons, researchers typically cannot force experimental subjects to adhere to experimental protocol. The existing methods either assume perfect compliance with treatment assignment or focus on intention-to-treat (ITT) analyses by ignoring the information about actual receipt of treatment.

Unfortunately, the ITT analysis is unable to tell, for example, whether a small causal effect arises due to ineffective treatment or low compliance. While researchers have developed methods to deal with noncompliance (e.g., Angrist *et al.*, 1996), they are based on the assumption of no interference between units. This assumption may be unrealistic since there are multiple ways in which spillover effects could arise. For example, one unit's treatment assignment may influence another unit's decision to receive the treatment. It is also possible that one's treatment receipt affects the outcomes of other units.

In this paper, we show how to analyze two-stage randomized experiments with both interference and noncompliance (Section 3). In an influential paper, Hudgens and Halloran (2008) propose two-stage randomized experiments as a general approach to causal inference with interference. We extend their framework so that it is applicable even in the presence of two-sided noncompliance. In particular, we define the complier average direct and spillover effects, propose consistent estimators, and derive their randomization-based

variances under the stratified interference assumption. Like Aronow (2012), we follow Hudgens and Halloran (2008) by referring to the effect of one's own treatment as a direct effect and the effect of another unit's treatment as a spillover effect. In a closely related working paper, Kang and Imbens (2016) also analyze two-stage randomized experiments with interference and noncompliance. We consider a more general pattern of interference by allowing for the spillover effect of the treatment assignment on the treatment receipt as well as the spillover effect of the treatment receipt on the outcome.

Finally, we prove the exact relationships between the proposed randomization-based estimators and the popular two-stage least squares estimators as well as those between their corresponding variance estimators. Thus, our analysis generalizes the results of Angrist *et al.* (1996) to the setting with interference. Our results on randomization inference build upon and extend the work of Basse and Feller (2018) to the case with noncompliance. We also conduct simulation studies to investigate the finite sample performance of the confidence intervals based on the proposed variance estimators (see Appendix D). The proposed methods are implemented via an open-source software package, experiment (Imai and Jiang, 2018), which is available at https://cran.r-project.org/package=experiment.

The proposed methodology is motivated by our own randomized evaluation of the Indian Health Insurance Scheme (known by the acronym RSBY), a study that employed the two-stage randomized design. In Section 2, we briefly describe the background and experimental design of this study. In Section 4, we apply the proposed methodology to this study. We present some evidence concerning the existence of positive spillover effects of treatment assignment on the enrollment in the RSBY. In addition, we estimate the complier average direct effect to be positive under the "low" treatment assignment mechanism, where fewer households in a village are encouraged to enroll in the insurance program. Finally, Section 5 concludes.

2 A Motivating Empirical Application

In this section, we describe the randomized evaluation of the Indian health insurance program, which serves as our motivating empirical application. We provide a brief background of the evaluation and introduce its experimental design.[1]

2.1 Randomized Evaluation of the Indian Health Insurance Program

Each year, 150 million people worldwide face financial catastrophe due to spending on health. According to a 2010 study, more than one third of them live in India (Shahrawat and Rao, 2011). Almost 63 million Indians fall below the poverty line (BPL) due to health spending (Berman *et al.*, 2010). In 2008, the Indian government introduced its first national, public health insurance scheme, Rastriya Swasthya Bima Yojana (RSBY), to address the problem. Its aim was to provide coverage for hospitalization to its BPL population, comprising roughly 250 million persons.

RSBY provides access to an insurance plan that covers inpatient hospital care for up to five members of each household. The plan covers all pre-existing diseases and there is no age limit of the beneficiaries. The rates of most surgical procedures are fixed by the government. Beneficiaries can obtain treatment at any hospital empaneled in the RSBY network. The insurance scheme is cashless, with the plan paying providers directly rather than reimbursing beneficiaries for expenses. The plan also covers INR 100 (or approximately USD 1.53) of transportation costs per hospitalization. The coverage lasts one year starting the month after the first enrollment in a particular district, but is often extended without cost to beneficiaries. The insurance plan is provided by private insurance companies, but the premium are paid by the government. In Karnataka, the state in which the randomized evaluation was conducted, premiums were roughly INR 200 (USD 3.07) per year during the study.

Households only have to pay INR 30 (USD 0.46) per year user fee to obtain an insurance card. There are no deductibles or co-payments and there is an annual cap of INR 30,000 (USD 460) per household.

We conducted a randomized controlled trial to determine whether RSBY increases access to hospitalization, and thus health, and reduced impoverishment due to high medical expenses. The findings are policy-relevant because the Indian government has announced a new scheme called the National Health Protection Scheme (NHPS) that seeks to build on RSBY to provide coverage for nearly 500 million Indians, but has not yet decided its design or how much to fund it.

In this evaluation, spillover effects are of concern because formal insurance may crowd out informal insurance, which is a substitute method of smoothing health care shocks (e.g., Jowett, 2003; Lin *et al.*, 2014). That is, the enrollment in RSBY by one household may depend on the treatment assignment of other households. In addition, we also must address noncompliance because some households in the treatment group decided not to enroll in RSBY while others in the control group managed to join the insurance program.

2.2 Experimental Design

Our evaluation study is based on a total of 11,089 above poverty line (APL) households in two districts of Karnataka State who had no pre-existing health insurance coverage and lived within 25 km of an RSBY empaneled hospital. We selected APL households because they are not otherwise eligible for RSBY, but are candidates for any expansion of RSBY. The two districts were Gulbarga and Mysore, which are economically and culturally representative of central and southern India, respectively. We required proximity to a hospital as hospital insurance has little value if there is no local hospital at which to use the insurance.

As shown in Table 1, we employed a two-stage randomized design to study both direct and spillover effects of RSBY. In the first stage, randomly selected 219 villages were assigned to the “High” treatment assignment mechanism whereas the rest of villages were assigned to the “Low” treatment assignment mechanism. In the second stage, under the “High” assignment mechanism, 80% of the households within a cluster are completely randomly assigned to the treatment condition, while the rest of households were assigned to the control group. In contrast, under the “Low” assignment mechanism, 40% of the households within a cluster are completely randomly assigned to the treatment condition. The households in the treatment group are given RSBY essentially for free, whereas some households in the control group were able to buy RSBY at the government price of roughly INR 200.[2]

Households were informed of the assigned treatment conditions and were given the opportunities to enroll in RSBY from April to May, 2015. Approximately 18 months later, we carried out a post-treatment survey and measured a variety of outcomes. Policy makers are interested in the health and financial effects of RSBY. To evaluate the efficacy of RSBY, we must estimate the effects of actual treatment receipt as well as the intention-to-treat effects because some households in the treatment group may not enroll in RSBY while others in the control group may do so.

3 The Proposed Methodology

In this section, we first review the intention-to-treat (ITT) analysis of two-stage randomized experiments proposed by Hudgens and Halloran (2008) and others. We then introduce a new causal quantity of interest, the complier average direct effect (CADE), present a nonparametric identification result, and propose a consistent estimator. We further consider the identification and inference of the CADE under the assumption of stratified interference, and derive the randomization-based variance of the proposed estimator. We also establish the

direct connections between these randomization-based estimators and the two-stage least squares estimators. Finally, we present analogous results for another new causal quantity, the complier average spillover effect (CASE), in Appendix A.

3.1 Two-Stage Randomized Experiments

We consider a two-stage randomized experiment (Hudgens and Halloran, 2008) with a total of N units and J clusters where each unit belongs to one of the clusters. We use n_j to denote the number of units in cluster j with $N = \sum_{j=1}^{J} n_j$. In a two-stage randomized experiment, we first randomly assign each cluster to one of the treatment assignment mechanisms, which in turn assigns different proportions of units within each cluster to the treatment condition. For the sake of simplicity, we consider two assignment mechanisms indicated by $A_j \in \{0,1\}$ where $A_j = 1$ ($A_j = 0$) indicates that a high (low) proportion of units are assigned to the treatment within cluster j. In our application, $A_j = 1$ corresponds to the treatment assignment probability of 80%, whereas $A_j = 0$ represents 40%. We assume complete randomization, in which a total of J_a clusters are assigned to the assignment mechanism a for $a = 0, 1$ with $J_0 + J_1 = J$. Finally, $\boldsymbol{A} = (A_1, A_2, \ldots, A_J)$ denotes the vector of treatment assignment mechanisms for all clusters.

The second stage of randomization concerns the treatment assignment for each unit within cluster j based on the assignment mechanism A_j. Let Z_{ij} be the binary treatment assignment variable for unit i in cluster j where $Z_{ij} = 1$ ($Z_{ij} = 0$) implies that the unit is assigned to the treatment (control) condition. Let $\boldsymbol{Z}_j = (Z_{1j}, \ldots, Z_{n_j j})$ denote the vector of assigned treatments for the n_j units in cluster j and $\Pr(\boldsymbol{Z}_j = \mathbf{z}_j \mid A_j = a)$ represent the distribution of the treatment assignment vector when cluster j is assigned to the assignment mechanism $A_j = a$. We assume the complete randomization such that a total of n_{jz} units in cluster j are assigned to the treatment condition z for $z = 0, 1$, where $n_{j1} + n_{j0} = n_j$.

Assumption 1. (Two-Stage Randomization)

1. *Complete randomization of treatment assignment mechanism at the cluster level:*

$$\Pr(\boldsymbol{A} = \boldsymbol{a}) = \frac{1}{\binom{J}{J_1}}$$

for all $\boldsymbol{a}$ *such that* $\mathbf{1}_J' \boldsymbol{a} = J_1$ *where* $\mathbf{1}_J$ *is the J dimensional vector of ones.*

2. *Complete randomization of treatment assignment within each cluster:*

$$\Pr(\boldsymbol{Z}_j = \boldsymbol{z} \mid A_j = a) = \frac{1}{\binom{n_j}{n_{j1}}}$$

for all $\boldsymbol{z}$ *such that* $\mathbf{1}_{n_j}' \boldsymbol{z} = n_{j1}$.

Following the literature, we adopt the finite population framework, in which potential outcomes are treated as constants and randomness comes from treatment assignment alone. We consider two-stage randomized experiments with noncompliance, in which the actual receipt of treatment may differ from the treatment assignment. Let D_{ij} be the treatment receipt for unit i in cluster j and $\boldsymbol{D}_j = (D_{1j}, \ldots, D_{n_j j})$ be the vector of treatment receipts for the n_j units in the cluster. The outcome variable Y_{ij} is observed for each unit and $\boldsymbol{Y}_j = (Y_{1j}, \ldots, Y_{n_j j})$ denotes the vector of observed outcomes for the n_j units in cluster j.

We use the potential outcomes framework of causal inference (e.g., Neyman, 1923; Holland, 1986; Rubin, 1990). For unit i in cluster j, let $D_{ij}(\boldsymbol{z})$ represent the potential value of treatment receipt, when the treatment assignment vector for all N units in the experiment equals $\boldsymbol{z}$. In addition, we use $Y_{ij}(\boldsymbol{z}; \boldsymbol{d})$ to denote the potential value of outcome, when the treatment assignment

vector equals $\boldsymbol{z}$ and treatment receipt vector equals $\boldsymbol{d}$. Lastly, let $Y_{ij}(\boldsymbol{z})$ represent the potential value of outcome when the treatment assignment vector equals $\boldsymbol{z}$, i.e., $Y_{ij}(\boldsymbol{z}) = Y_{ij}(\boldsymbol{z}; D_{ij}(\boldsymbol{z}))$. The observed values of treatment receipt and outcome are given by $D_{ij} = D_{ij}(\boldsymbol{Z})$ and $Y_{ij} = Y_{ij}(\boldsymbol{Z})$ where $\boldsymbol{Z}$ is the N dimensional vector of treatment assignment for all units. If there were no restriction on the pattern of interference, each unit has 2^N potential values of treatment receipt and outcome, making identification infeasible. Hence, following the literature (e.g., Hong and Raudenbush, 2006; Sobel, 2006; Hudgens and Halloran, 2008), we only allow interference within each cluster.

Assumption 2 (Partial Interference).

$$Y_{ij}(\mathbf{z}) = Y_{ij}(\mathbf{z}') \quad \text{and} \quad D_{ij}(\mathbf{z}) = D_{ij}(\mathbf{z}')$$

for all $\mathbf{z}$ *and* $\mathbf{z}'$ *with* $\mathbf{z}_j = \mathbf{z}_j'$.

Assumption 2 implies that although the treatment receipt and outcome of a unit can be influenced by the treatment assignment of another unit within the same cluster, they cannot be affected by units in other clusters. This assumption substantially reduces the number of potential values of treatment receipt and outcome for each unit in cluster j from 2^N to 2^{n_j}.

3.2 Intention-to-Treat Effects: A Review

We next review the previous results about the ITT analysis of two-stage randomized experiments under the partial interference assumption (Hudgens and Halloran, 2008). Our analysis differs from the existing ones in that we weight each unit equally instead of giving an equal weight to each cluster as done in the literature.

3.2.1 Causal Quantities of Interest

We begin by defining preliminary average quantities. First, we define the average potential value of treatment receipt for unit *i* in cluster *j* when the unit is assigned to the treatment condition *z* under the treatment assignment mechanism *a*. We do so by averaging over the distribution of treatment assignments for the other units within the same cluster,

$$\bar{D}_{ij}(z,a) = \sum_{\boldsymbol{z}_{-i,j} \in \mathsf{Z}_{-i,j}} D_{ij}(Z_{ij} = z, \boldsymbol{Z}_{-i,j} = \boldsymbol{z}_{-i,j}) \Pr(\boldsymbol{Z}_{-i,j} = \boldsymbol{z}_{-i,j} \mid Z_{ij} = z, A_j = a),$$

where $\boldsymbol{Z}_{-i,j} = (Z_{1i}, \ldots, Z_{i-1,j}, Z_{i+1,j}, \ldots, Z_{n_j j})$ represents the $(n_j - 1)$ dimensional subvector of $\boldsymbol{Z}_j$ with the entry for unit *i* removed and $\mathsf{Z}_{-i,j} = \{(z_{1j}, \ldots, z_{i-1,j}, z_{i+1,j}, \ldots, z_{n_j j}) \mid z_{i'j} \in \{0,1\} \text{ for } i' = 1, \ldots, i-1, i+1, \ldots, n_j\}$ is the set of all possible values of the assignment vector $\boldsymbol{Z}_{-i,j}$. Similarly, we define the average potential outcome for unit *i* in cluster *j* as,

$$\bar{Y}_{ij}(z,a) = \sum_{\boldsymbol{z}_{-i,j} \in \mathsf{Z}_{-i,j}} Y_{ij}(Z_{ij} = z, \boldsymbol{Z}_{-i,j} = \boldsymbol{z}_{-i,j}) \Pr(\boldsymbol{Z}_{-i,j} = \boldsymbol{z}_{-i,j} \mid Z_{ij} = z, A_j = a).$$

Given these unit-level average potential outcomes, we consider the cluster-level and population-level average potential values of the treatment receipt and outcome,

$$\bar{D}_j(z,a) = \frac{1}{n_j} \sum_{i=1}^{n_j} \bar{D}_{ij}(z,a), \quad \bar{D}(z,a) = \frac{1}{N} \sum_{j=1}^{J} n_j \bar{D}_j(z,a),$$

$$\bar{Y}_j(z,a) = \frac{1}{n_j} \sum_{i=1}^{n_j} \bar{Y}_{ij}(z,a), \quad \bar{Y}(z,a) = \frac{1}{N} \sum_{j=1}^{J} n_j \bar{Y}_j(z,a).$$

We define the ITT effects, starting with the average direct effect of treatment assignment on the treatment receipt and outcome under the treatment assignment mechanism *a*, as

$$\mathrm{DED}_{ij}(a) = \bar{D}_{ij}(1,a) - \bar{D}_{ij}(0,a), \quad \mathrm{DEY}_{ij}(a) = \bar{Y}_{ij}(1,a) - \bar{Y}_{ij}(0,a).$$

where DED and DEY stand for the average direct effect on D and Y, respectively. These parameters quantify how the treatment assignment of a unit may affect its treatment receipt and outcome by averaging the treatment assignments of other units within the same cluster under a specific assignment mechanism. Finally, averaging these unit-level quantities gives the following average direct effects of treatment assignment for each cluster and for the entire (finite) population,

$$\mathrm{DED}_j(a) = \frac{1}{n_j}\sum_{i=1}^{n_j}\mathrm{DED}_{ij}(a), \quad \mathrm{DED}(a) = \frac{1}{N}\sum_{j=1}^{J} n_j \mathrm{DED}_j(a),$$

$$\mathrm{DEY}_j(a) = \frac{1}{n_j}\sum_{i=1}^{n_j}\mathrm{DEY}_{ij}(a), \quad \mathrm{DEY}(a) = \frac{1}{N}\sum_{j=1}^{J} n_j \mathrm{DEY}_j(a).$$

Another quantity of interest is the spillover effect, which quantifies how one unit's treatment receipt or outcome is affected by other units' treatment assignments. Following Halloran and Struchiner (1995), we define the unit-level spillover effects on the treatment receipt and outcome as,

$$\mathrm{SED}_{ij}(z) = \bar{D}_{ij}(z,1) - \bar{D}_{ij}(z,0), \quad \mathrm{SEY}_{ij}(z) = \bar{Y}_{ij}(z,1) - \bar{Y}_{ij}(z,0),$$

which compare the average potential values under two different assignment mechanisms, i.e., $a = 1$ and $a = 0$, while holding one's treatment assignment at z. We then define the spillover effects on the treatment receipt and outcome at the cluster and population levels,

$$\mathrm{SED}_j(z) = \frac{1}{n_j}\sum_{i=1}^{n_j}\mathrm{SED}_{ij}(z), \quad \mathrm{SED}(z) = \frac{1}{N}\sum_{j=1}^{J} n_j \mathrm{SED}_j(z).$$

$$\mathrm{SEY}_j(z) = \frac{1}{n_j}\sum_{i=1}^{n_j}\mathrm{SEY}_{ij}(z), \quad \mathrm{SEY}(z) = \frac{1}{N}\sum_{j=1}^{J} n_j \mathrm{SEY}_j(z).$$

The quantities defined above differ from those introduced in the literature in that we equally weight each unit (see Basse and Feller, 2018). In contrast, Hudgens

and Halloran (2008) give an equal weight to each cluster regardless of its size. While our analysis focuses on the individual-weighted estimands rather than cluster-weighted estimands, our method can be generalized to any weighting scheme, and as such the proofs in the supplementary appendix are based on general weights.

Finally, in actual policy implementations, the treatment assignment is typically based on a deterministic criterion rather than randomization, suggesting that the causal quantities discussed above may not be of direct interest to policy makers. Even in this situation, however, these causal quantities can provide some policy implications by telling us whether or not spillover effects exist at all. We discuss this issue in the context of our application (see Section 4) and consider a model-based approach to further address this point (see Appendix E).

3.2.2 Nonparametric Identification

Hudgens and Halloran (2008) establish the nonparametric identification of the ITT effects, which equally weight each cluster regardless of its size. Here, we present analogous results by weighting each unit equally as done above. Define the following quantities,

$$\hat{D}(z,a) = \frac{\frac{1}{N}\sum_{j=1}^{J} n_j \hat{D}_j(z,a) I(A_j = a)}{\frac{1}{J}\sum_{j=1}^{J} I(A_j = a)}, \quad \hat{Y}(z,a) = \frac{\frac{1}{N}\sum_{j=1}^{J} n_j \hat{Y}_j(z,a) I(A_j = a)}{\frac{1}{J}\sum_{j=1}^{J} I(A_j = a)},$$

where

$$\hat{D}_j(z,a) = \frac{\sum_{i=1}^{n_j} D_{ij} I(Z_{ij} = z)}{\sum_{i=1}^{n_j} I(Z_{ij} = z)}, \quad \hat{Y}_j(z,a) = \frac{\sum_{i=1}^{n_j} Y_{ij} I(Z_{ij} = z)}{\sum_{i=1}^{n_j} I(Z_{ij} = z)}.$$

Then, we can obtain the unbiased estimators of the direct effects and the spillover effects.

Theorem 1 (Unbiased Estimation of the ITT Effects). *Define the following estimators,*

$$\widehat{\text{DED}}(a) = \hat{D}(1,a) - \hat{D}(0,a), \quad \widehat{\text{SED}}(z) = \hat{D}(z,1) - \hat{D}(z,0),$$
$$\widehat{\text{DEY}}(a) = \hat{Y}(1,a) - \hat{Y}(0,a), \quad \widehat{\text{SEY}}(z) = \hat{Y}(z,1) - \hat{Y}(z,0).$$

Under Assumptions 1 and 2, these estimators are unbiased for the ITT effects,

$$\text{E}\{\widehat{\text{DED}}(a)\} = \text{DED}(a), \quad \text{E}\{\widehat{\text{SED}}(z)\} = \text{SED}(z),$$
$$\text{E}\{\widehat{\text{DEY}}(a)\} = \text{DEY}(a), \quad \text{E}\{\widehat{\text{SEY}}(z)\} = \text{SEY}(z).$$

Proof is straightforward and hence omitted.

3.3 Complier Average Direct Effects

We now address the issue of noncompliance in the presence of interference between units. In a seminal paper, Angrist *et al.* (1996) show how to identify the complier average causal effect (CACE) in standard randomized experiments under the assumption of no interference. The CACE represents the average effect of treatment receipt among the compliers who would receive the treatment only when assigned to the treatment condition. Below, we introduce the complier average direct effect, which is a generalization of the CACE to settings with interference, and show how to nonparametrically identify and consistently estimate it using the data from two-stage randomized experiments.

3.3.1 Causal Quantity of Interest

We first generalize the definition of compliers to settings with interference between units. Under the assumption of no interference, compliers are those who receive the treatment only when assigned to the treatment condition. However, in the presence of partial interference, the treatment receipt is also affected by the

treatment assignment of other units in the same cluster. Thus, the compliance status of a unit is a function of the treatment assignment of other units in the same cluster,

$$C_{ij}(\mathbf{z}_{-i,j}) = I\{D_{ij}(1, \mathbf{z}_{-i,j}) = 1, D_{ij}(0, \mathbf{z}_{-i,j}) = 0\}. \quad (1)$$

We consider a measure of compliance behavior for each unit by averaging over the distribution of treatment assignments of the other units within the same cluster under the treatment assignment mechanism *a*. This general measure of compliance behavior ranges from 0 to 1 and is defined as,

$$\sum_{\mathbf{z}_{-i,j} \in \mathcal{Z}_{-i,j}} C_{ij}(\mathbf{z}_{-i,j}) \Pr(\mathbf{Z}_{-i,j} = \mathbf{z}_{-i,j} \mid A_j = a) \qquad (2)$$

for $a = 0, 1$. Given this compliance measure, we now define the complier average direct effect (CADE) as the average direct effect of treatment assignment among compliers,

$$\text{CADE}(a) = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{\mathbf{z}_{-i,j} \in \mathcal{Z}_{-i,j}} \{Y_{ij}(1, \mathbf{z}_{-i,j}) - Y_{ij}(0, \mathbf{z}_{-i,j})\} C_{ij}(\mathbf{z}_{-i,j}) \Pr(\mathbf{Z}_{-i,j} = \mathbf{z}_{-i,j} \mid A_j = a)}{\sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{\mathbf{z}_{-i,j} \in \mathcal{Z}_{-i,j}} C_{ij}(\mathbf{z}_{-i,j}) \Pr(\mathbf{Z}_{-i,j} = \mathbf{z}_{-i,j} \mid A_j = a)}.$$

The definition requires that there exists at least one complier in the population. If units do not influence each other, we have $Y_{ij}(z_{ij}, \mathbf{z}_{-i,j}) = Y_{ij}(z_{ij})$ and $D_{ij}(z_{ij}, \mathbf{z}_{-i,j}) = D_{ij}(z_{ij})$. Hence, the compliance status for each unit in equations (1) and (2) no longer depends on the treatment assignment of the other units. As a result, under this setting, the CADE equals the finite sample version of the complier average causal effect defined in Angrist *et al.* (1996). Finally, in the absence of noncompliance, i.e., $C_{ij}(\mathbf{z}_{-i,j}) = 1$ for all $\mathbf{z}_{-i,j}$ and *i*, *j*, then $\text{CADE}(a)$ asymptotically equals $\text{DEY}(a)$ as the cluster size grows.

The CADE combines two causal pathways: a unit's treatment assignment Z_{ij} can affect its outcome Y_{ij} either through its own treatment receipt D_{ij} or that of the

other units $\boldsymbol{D}_{-i,j} = (D_{1j}, \ldots, D_{i-1,j}, D_{i+1,j}, \ldots, D_{n_j j})$. If there is either no spillover effect of encouragement on treatment receipt or no spillover effect of treatment receipt on outcome, then the second causal pathway no longer exists. Under this scenario, the CADE corresponds to the average direct effect of one's own treatment receipt among compliers because the treatment assignment is the same as the treatment receipt. In contrast, when both types of spillover effects exist, the CADE includes the indirect effect of one's own encouragement on the outcome through the treatment receipt of other units in the same village as well as the direct effect of one's own treatment receipt on the outcome. Unfortunately, without additional assumptions, the CADE is not identifiable. We therefore propose a set of assumptions for nonparametric identification. In addition, Appendix E.2 considers a model-based approach to the identification and estimation for further distinguishing the two causal pathways.

3.3.2 Nonparametric Identification

To establish the nonparametric identification of the CADE, we begin by generalizing the exclusion restriction of Angrist *et al.* (1996), which assumes no interference between units.

Assumption 3 (Exclusion restriction with Interference between Units).

$$Y_{ij}(\boldsymbol{z}_j; \boldsymbol{d}_j) = Y_{ij}(\boldsymbol{z'}_j; \boldsymbol{d}_j) \quad \text{for any } \boldsymbol{z}_j, \boldsymbol{z}_j' \text{ and } \boldsymbol{d}_j.$$

Assumption 3 states that the outcome of a unit does not depend on the treatment assignment of any unit within the same cluster (including itself) so long as the treatment receipt for all the units of the cluster remains identical. In other words, the outcome of a unit depends only on the treatment receipt vector of all units within its own cluster. The assumption is violated if the outcome of one unit is influenced by its own treatment assignment or that of another unit within the same cluster even when the treatment receipts of all the units in the cluster including itself are held constant. In our application, the assumption is plausible

since the encouragement to enroll in the RSBY is unlikely to affect the hospital expenditure other than through the actual enrollment itself.

Under Assumption 3, we can write the potential outcome as the function of treatment receipt alone, $Y_{ij}(\boldsymbol{d}_j)$. Thus, the observed outcome is written as $Y_{ij}(\boldsymbol{D}_j)$ where $\boldsymbol{D}_j = \boldsymbol{D}_j(\boldsymbol{Z}_j)$. We maintain Assumption 3 for the remainder of the paper. To avoid confusion, we will explicitly write out the treatment receipt as the argument of potential outcome. For example, $Y_{ij}(\boldsymbol{D}_j = \boldsymbol{1}_{n_j})$ represents the potential outcome when D_{ij} = 1 for $j = 1, \ldots, n_j$, while $Y_{ij}(\boldsymbol{1}_{n_j})$ represents the potential outcome when Z_{ij} = 1 for $j = 1, \ldots, n_j$.

We next generalize the monotonicity assumption of Angrist *et al.* (1996).

Assumption 4 (Monotonicity with Interference between Units).

$$D_{ij}(1, \boldsymbol{z}_{-i,j}) \geq D_{ij}(0, \boldsymbol{z}_{-i,j}) \quad \text{for all } \boldsymbol{z}_{-i,j} \in \mathsf{Z}_{-i,j}.$$

The assumption states that being assigned to the treatment condition never negatively affects the treatment receipt of a unit, regardless of how the other units within the same cluster are assigned to the treatment/control conditions. Assumption 4 is plausible in our application because the encouragement is expected to increase the enrollment in the RSBY.

In the absence of interference between units, exclusion restriction and monotonicity are sufficient for the nonparametric identification of the complier average causal effect. However, when interference exists, an additional restriction on the interference structure is necessary. The reason is that there are two types of possible spillover effects: the spillover effect of treatment assignment on the treatment receipt and the spillover effect of treatment receipt on the outcome. As a result, even under exclusion restriction, the treatment assignment of a noncomplier can still affect its outcome through the treatment receipts of other units within in the same cluster.

To address this problem, we propose the following identification assumption.

Assumption 5 (Restricted Interference under Noncompliance). *For any unit i in cluster j, if* $D_{ij}(1, \mathbf{z}_{-i,j}) = D_{ij}(0, \mathbf{z}_{-i,j})$ *for some* $\mathbf{z}_{-i,j} \in \mathsf{Z}_{-i,j}$*, then* $Y_{ij}(\boldsymbol{D}_j(1, \mathbf{z}_{-i,j})) = Y_{ij}(\boldsymbol{D}_j(0, \mathbf{z}_{-i,j}))$ *holds.*

The assumption states that if the treatment receipt of a unit is not affected by its own treatment assignment (i.e., the unit is a noncomplier), then its outcome should also not be affected by its own treatment assignment through the treatment receipts of other units in the same cluster. Although Assumption 5 appears to be concerned only with the spillover effects of treatment receipt on the outcome, its plausibility also depends on the spillover effects of treatment assignment on the treatment receipt.

To facilitate the understanding of this assumption, we consider the following three scenarios under which Assumption 5 is satisfied. First, assume no spillover effect of treatment receipt on the outcome (Scenario I of Figure 1(a)),

$$Y_{ij}(d_{ij}, \boldsymbol{d}_{-i,j}) = Y_{ij}(d_{ij}, \boldsymbol{d}'_{-i,j}) \text{ for } d_{ij} = 0,1, \text{ and any } \boldsymbol{d}_{-i,j}, \boldsymbol{d}'_{-i,j}. \quad (3)$$

Testable conditions for this scenario are given in Appendix B.1.

Second, suppose that the treatment assignment has no spillover effect on the treatment receipt (Scenario II of Figure 1(b)),

$$D_{ij}(z_{ij}, \mathbf{z}_{-i,j}) = D_{ij}(z_{ij}, \mathbf{z}'_{-i,j}) \text{ for } z_{ij} = 0,1, \text{ and any } \mathbf{z}_{-i,j}, \mathbf{z}'_{-i,j}. \quad (4)$$

Such an assumption is made by Kang and Imbens (2016) in the context of online experiments, in which the assignment of treatment (e.g., social media messaging) can be individualized but units may interact with each other once they receive the treatment. We can test this scenario by estimating $\text{SED}(1)$ and $\text{SED}(0)$.

Third, we can weaken the condition in equation (4) by considering an alternative condition that if a unit's treatment receipt is not affected by its own treatment assignment (i.e., the unit is a noncomplier), then the treatment assignment of this unit has no effect on the treatment receipts of the other units in the same cluster (the absence of dotted edges in Scenario III of Figure 1(c)),

$$\text{if } D_{ij}(1, \boldsymbol{z}_{-i,j}) = D_{ij}(0, \boldsymbol{z}_{-i,j}), \text{ then } \boldsymbol{D}_{-i,j}(1, \boldsymbol{z}_{-i,j}) = \boldsymbol{D}_{-i,j}(0, \boldsymbol{z}_{-i,j}).$$

In our application, this scenario is violated, for example, if a household that already has insurance and is not going to be affected by the encouragement influences the enrollment decision of another household by recommending the RSBY to it. To increase the plausibility of this scenario in our application, we excluded all the households with pre-existing insurance from the experiment. As a result, this scenario is plausible because one's encouragement is expected to have a much greater influence on his/her own enrollment than the enrollment of another unit.

Although all three scenarios above satisfy Assumption 5, the interpretation of the CADE is different. In particular, under Scenarios I and II, we can interpret the CADE as the average direct effect of one's own treatment receipt on the outcome among compliers. In contrast, under Scenario III, the CADE also includes the average direct effect of one's own encouragement on the outcome through the treatment receipts of other units. Nevertheless, this combined direct effect of encouragement may be of interest to policy makers because most government programs including the RSBY are based on the encouragement design. In Appendix E.2, we address this issue using a model-based approach.

The next theorem establishes the nonparametric identification of the CADE as the cluster size tends to infinity. Under Assumptions 1–5, we show that in the limit, the CADE equals the ratio of the average direct effects of treatment assignment on the outcome and on the treatment receipt while holding the treatment assignment mechanism fixed. Although the unbiased estimation of

$\mathrm{DEY}(a)$ and $\mathrm{DED}(a)$ is readily available (Hudgens and Halloran, 2008), for the consistent estimation of the CADE, we need an additional restriction on the structure of interference. We follow Sävje *et al.* (2017)'s result on the consistency of average causal effect in finite population framework, and assume that the average amount of interference per unit does not grow proportionally to the cluster size (see Appendix B.2 for a proof of the theorem and the details).

Theorem 2. (Nonparametric Identification and Consistent Estimation of the Complier Average Direct Effect)

1. *Under Assumptions 1–5, we have*

$$\lim_{n_j \to \infty} \frac{\mathrm{DEY}(a)}{\mathrm{DED}(a)} = \lim_{n_j \to \infty} \mathrm{CADE}(a).$$

2. *Suppose that the outcome is bounded and the restriction on interference in Sävje et al. (2017) holds for both the treatment receipt and the outcome. Then, as both the cluster size n_j and the number of clusters J go to infinity, we can consistently estimate the* CADE*,*

$$\mathrm{plim}_{n_j \to \infty, J \to \infty} \frac{\mathrm{DEY}(a)}{\mathrm{DED}(a)} = \lim_{n_j \to \infty, J \to \infty} \mathrm{CADE}(a)$$

for each a = 0, 1.

The CADE is nonparametrically identifiable as the cluster size and the number of clusters tend to infinity, and can be consistently estimated by the ratio of two estimated ITT effects. The asymptotic properties are derived within the finite population framework, approximating the sampling distribution of an estimator by embedding it in an asymptotically stable sequence of finite populations (Hájek, 1960; Lehmann, 2004).

3.4 Stratified Interference

Unfortunately, as pointed out by Hudgens and Halloran (2008), a valid estimator of the variances of these ITT effect estimators is unavailable without an additional assumption. Hudgens and Halloran (2008) rely upon the stratified interference assumption that the outcome of one unit depends on the treatment assignment of other units only through the number of those who are assigned to the treatment condition within the same cluster. In other words, what matters is the number of units rather than which units are assigned to the treatment condition.

We assume that stratified interference applies to both the outcome and treatment receipt.

Assumption 6 (Stratified Interference).

$$D_{ij}(\boldsymbol{z}_j) = D_{ij}(\boldsymbol{z}_j') \quad \text{and} \quad Y_{ij}(\boldsymbol{z}_j) = Y_{ij}(\boldsymbol{z}_j') \quad \text{if} \quad z_{ij} = z_{ij}' \text{ and } \sum_{i=1}^{n_j} z_{ij} = \sum_{i=1}^{n_j} z_{ij}'.$$

In our application, stratified interference for the treatment receipt requires that the enrollment decisions of households depend only on their own encouragement and the number of encouraged households in their village. Under the assumption of no spillover effect of treatment receipt on the outcome, stratified interference for the outcome holds so long as it is applicable to the treatment receipt. However, for more general scenarios, Assumption 6 may not be satisfied for the outcome even if it holds for the treatment receipt.

3.4.1 Nonparametric Identification

Under Assumption 6, we can simplify the CADE because the number of the units assigned to the treatment condition in each cluster is fixed given treatment assignment mechanism. This implies that we can write $D_{ij}(\boldsymbol{z}_j)$ and $Y_{ij}(\boldsymbol{z}_j)$ as $D_{ij}(z,a)$ and $Y_{ij}(z,a)$, respectively, and as a result CADE(a) equals

$$\text{CADE}(a) = \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}\{Y_{ij}(1,a) - Y_{ij}(0,a)\}I\{D_{ij}(1,a) - D_{ij}(0,a) = 1\}}{\sum_{j=1}^{J}\sum_{i=1}^{n_j} I\{D_{ij}(1,a) - D_{ij}(0,a) = 1\}},$$

where the complier status can also be simplified as a function of assignment mechanism alone, i.e., $C_{ij}(a) = I\{D_{ij}(1,a) = 1, D_{ij}(0,a) = 0\}$.

We now present the results on nonparametric identification and consistent estimation under stratified interference.

Theorem 3. (Nonparametric Identification and Consistent Estimation of the Complier Average Direct Effect under Stratified Interference) *Suppose that the outcome is bounded. Then, under Assumptions 1–6, we have*

$$\lim_{n_j \to \infty, J \to \infty} \text{CADE}(a) = \text{plim}_{n_j \to \infty, J \to \infty} \frac{\text{DEY}(a)}{\text{DED}(a)}$$

for a = 0, 1.

Proof is in Appendix B.4. Under the stratified interference assumption, the consistent estimation of CADE no longer requires the restrictions on interference in Sävje *et al.* (2017).

3.4.2 Effect Decomposition

Under stratified interference, we can decompose the average direct effect of treatment assignment as the sum of the average direct effects for compliers and noncompliers,

$$\text{DEY}(a) = \text{CADE}(a) \cdot \pi_c(a) + \text{NADE}(a) \cdot \{1 - \pi_c(a)\}, \quad (5)$$

where $\text{NADE}(a)$ is the noncomplier average direct effect and is defined as,

$$\text{NADE}(a) = \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}\{Y_{ij}(1,a) - Y_{ij}(0,a)\}I\{D_{ij}(1,a) = D_{ij}(0,a)\}}{\sum_{j=1}^{J}\sum_{i=1}^{n_j} I\{D_{ij}(1,a) = D_{ij}(0,a)\}},$$

and the proportion of compliers is given by,

$$\pi_c(a) = \frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{n_j} I\{D_{ij}(1,a) = 1, D_{ij}(0,a) = 0\}.$$

According to the exclusion restriction given in Assumption 3, for compliers with $D_{ij}(1,a) = 1$ and $D_{ij}(0,a) = 0$, we can write the unit-level direct effect on the outcome as the sum of the direct effect through its own treatment receipt and the indirect effect through the treatment receipts of other units within the same cluster,

$$\begin{aligned} & Y_{ij}(Z_{ij} = 1, a) - Y_{ij}(Z_{ij} = 0, a) \\ & = \{Y_{ij}(D_{ij} = 1, \boldsymbol{D}_{-i,j}(Z_{ij} = 1, a)) - Y_{ij}(D_{ij} = 0, \boldsymbol{D}_{-i,j}(Z_{ij} = 1, a))\} \\ & + \{Y_{ij}(D_{ij} = 0, \boldsymbol{D}_{-i,j}(Z_{ij} = 1, a)) - Y_{ij}(D_{ij} = 0, \boldsymbol{D}_{-i,j}(Z_{ij} = 0, a))\}. \end{aligned} \qquad (6)$$

Thus, the treatment assignment can affect its outcome either directly through its own treatment or indirectly through the treatment receipts of the other units in the same cluster.

For noncompliers ($D_{ij}(1,a) = D_{ij}(0,a) = d$), the exclusion restriction implies,

$$\begin{aligned} & Y_{ij}(Z_{ij} = 1, a) - Y_{ij}(Z_{ij} = 0, a) \\ & = Y_{ij}(D_{ij} = d_{ij}, \boldsymbol{D}_{-i,j}(Z_{ij} = 1, a)) - Y_{ij}(D_{ij} = d_{ij}, \boldsymbol{D}_{-i,j}(Z_{ij} = 0, a)). \end{aligned} \qquad (7)$$

The treatment assignment affects its own outcome only through the treatment receipt of the other units in the same cluster. Furthermore, Assumption 5 implies $Y_{ij}(D_{ij} = d_{ij}, \boldsymbol{D}_{-i,j}(Z_{ij} = 1, a)) = Y_{ij}(D_{ij} = d_{ij}, \boldsymbol{D}_{-i,j}(Z_{ij} = 0, a))$. Under this assumption, equation (7) equals zero, implying $\text{NADE}(a) = 0$ and the identification of $\text{CADE}(a)$.

3.4.3 Randomization-based Variances

We derive the randomization-based variances of the proposed estimators within the finite population framework, in which the uncertainty comes solely from the two-stage randomization. As shown by Hudgens and Halloran (2008) in the context of ITT analysis, stratified interference enables the estimation of variance. Here, we first derive the variances of the ITT effect estimators and then derive the variance of the proposed CADE estimator. We begin by defining the following quantities,

$$\sigma_j^2(z,a)=\frac{1}{n_j-1}\sum_{i=1}^{n_j}\{Y_{ij}(z,a)-\bar{Y}_j(z,a)\}^2,\quad \sigma_{DE}^2(a)=\frac{1}{J-1}\sum_{j=1}^{J}\left\{\frac{n_jJ}{N}\mathrm{DEY}_j(a)-\mathrm{DEY}(a)\right\}^2,$$

$$\omega_j^2(a)=\frac{1}{n_j-1}\sum_{i=1}^{n_j}\ [\{Y_{ij}(1,a)-Y_{ij}(0,a)\}-\{\bar{Y}_j(1,a)-\bar{Y}_j(0,a)\}]^2,$$

where $\sigma_j^2(z,a)$ is the within-cluster variance of potential outcomes, $\sigma_{DE}^2(a)$ is the between-cluster variance of $\mathrm{DEY}_{ij}(a)$, and $\omega_j^2(a)$ is the within-cluster variance of $\mathrm{DEY}_{ij}(a)$. Using this notation, we give the results for the ITT effects of treatment assignment on the outcome. The results for the ITT effects of treatment assignment on the treatment receipt can be obtained in the same way.

Theorem 4 (Randomization-based Variances of the ITT Effect Estimators). *Under Assumptions 1, 2 and 6, we have*

$$\mathrm{var}\left\{\mathrm{DEY}(a)\right\}=\left(1-\frac{J_a}{J}\right)\frac{\sigma_{DE}^2(a)}{J_a}+\frac{1}{J_aJ}\sum_{j=1}^{J}\mathrm{var}\left\{\frac{n_jJ}{N}\mathrm{DEY}_j(a)\mid A_j=a\right\},$$

where

$$\mathrm{var}\left\{\mathrm{DEY}_j(a)\mid A_j=a\right\}=\frac{\sigma_j^2(1,a)}{n_{j1}}+\frac{\sigma_j^2(0,a)}{n_{j0}}-\frac{\omega_j^2(a)}{n_j}.$$

Proof is given in Appendix B.5. Because we cannot observe $Y_{ij}(1,a)$ and $Y_{ij}(0,a)$ simultaneously, no unbiased estimator exists for $\omega_j^2(a)$, implying that no

unbiased estimation of the variances is possible. Thus, following Hudgens and Halloran (2008), we propose a conservative estimator,

$$\text{var}\left\{\text{DEY}(a)\right\} = \left(1-\frac{J_a}{J}\right)\frac{\hat{\sigma}^2_{DE}(a)}{J_a} + \frac{1}{J_a J}\sum_{j=1}^{J}\frac{n_j^2 J^2}{N^2}\left(\frac{\hat{\sigma}^2_j(1,a)}{n_{j1}} + \frac{\hat{\sigma}^2_j(0,a)}{n_j - n_{j1}}\right) I(A_j = a), \quad (8)$$

where

$$\hat{\sigma}^2_j(z,a) = \frac{\sum_{i=1}^{n_j}\{Y_{ij} - \hat{Y}_j(z,a)\}^2 I(Z_{ij} = z)}{n_{jz} - 1},$$

$$\hat{\sigma}^2_{DE}(a) = \frac{\sum_{i=1}^{J}\left\{\frac{n_j J}{N}\text{DEY}_j(a) - \text{DEY}(a)\right\}^2 I(A_j = a)}{J_a - 1}.$$

In equation (8), $\hat{\sigma}^2_{DE}(a)$ represents the between-cluster sample variance, and $\hat{\sigma}^2_j(z,a)$ is the within-cluster variance in cluster *j*. Thus, the variance of the ITT direct effect estimator is a weighted average of the between-cluster sample variance and the within-cluster sample variance.

It can be shown that this variance estimator is on average no less than the true variance,

$$\text{E}\left[\text{var}\left\{\text{DEY}(a)\right\}\right] \geq \text{var}\left\{\text{DEY}(a)\right\},$$

where the inequality becomes equality when the unit-level direct effect, i.e., $Y_{ij}(1,a) - Y_{ij}(0,a)$, is constant within each cluster (see Appendix B.7 for a proof). In Appendix B.3, we provide the asymptotic normality result of the ITT effect estimators under additional regularity conditions based on the finite population central limit theorems in Hájek (1960), Ohlsson (1989) and Li and Ding (2017). These conditions are satisfied for a bounded outcome as the cluster size and the number of clusters go to infinity. See Chin (2018) for more refined results on the asymptotic normality of the ITT effect estimators without stratified interference.

We next derive the asymptotic randomization-based variance of the proposed estimator.

Theorem 5 (Randomization-based Variance of the CADE Estimator). *Under Assumptions 1–6, the asymptotic variance of* $\mathrm{CADE}(a)$ *is*

$$\frac{1}{\mathrm{DED}(a)^2}\left[\mathrm{var}\left\{\mathrm{DEY}(a)\right\} - 2\frac{\mathrm{DEY}(a)}{\mathrm{DED}(a)}\mathrm{cov}\left\{\mathrm{DEY}(a),\mathrm{DED}(a)\right\} + \frac{\mathrm{DEY}(a)^2}{\mathrm{DED}(a)^2}\mathrm{var}\left\{\mathrm{DED}(a)\right\}\right].$$

Proof of Theorem 5 is a direct application of the Delta method based on the asymptotic normality of the ITT effect estimators shown in Appendix B.3. Due to the space limitation, we give the expression of $\mathrm{cov}\left\{\mathrm{DEY}(a),\mathrm{DED}(a)\right\}$ in Appendix B.6. Because the proposed CADE estimator is a ratio estimator, its variance blows up when DED is close to zero. This is similar to the weak instrument problem in the standard instrumental variable settings.

We obtain the following variance estimator by replacing each term in the brackets with its conservative estimator,

$$\mathrm{var}\left\{\mathrm{CADE}(a)\right\} = \frac{1}{\mathrm{DED}(a)^2}\left[\mathrm{var}\left\{\mathrm{DEY}(a)\right\} - 2\frac{\mathrm{DEY}(a)}{\mathrm{DED}(a)}\mathrm{cov}\left\{\mathrm{DEY}(a),\mathrm{DED}(a)\right\} + \frac{\mathrm{DEY}(a)^2}{\mathrm{DED}(a)^2}\mathrm{var}\left\{\mathrm{DED}(a)\right\}\right], \quad (9)$$

where $\mathrm{var}\left\{\mathrm{DED}(a)\right\}$ and $\mathrm{cov}\left\{\mathrm{DEY}(a),\mathrm{DED}(a)\right\}$ are obtained by replacing *Y* with *D* and the sample variances with the sample covariances in equation (8), respectively. Similar to the ITT analysis, each of the three terms in the brackets of $\mathrm{var}\left\{\mathrm{CADE}(a)\right\}$ is a weighted average of between-cluster and within-cluster sample variances.

Because the expectation of product is generally not equal to the product of expectations, unlike the ITT analysis, $\mathrm{var}\left\{\mathrm{CADE}(a)\right\}$ is not a conservative

variance estimator in finite samples. In Appendix B.8, however, we show that it is asymptotically conservative. Finally, to evaluate the robustness of the variance estimator based on Assumption 6, we conduct simulation studies and find that the proposed variance estimator works well so long as the number of clusters is relatively large (see Appendix D).

3.5 Connections to Two-stage Least Squares Regression

In this section, we establish direct connections between the proposed estimator of the CADE and the two-stage least squares estimator, which is popular among applied researchers. Basse and Feller (2018) study the relationships between the ordinary least squares and randomization-based estimators for the ITT analysis under a particular two-stage randomized experiment design. Here, we further extend these previous results.

3.5.1 Point Estimates

We begin with the ITT analysis. To account for different cluster sizes, we transform the treatment and outcome variables so that each unit, rather than each cluster, is equally weighted. Specifically, we multiply them by the weights proportional to the cluster size, i.e., $D_{ij}^{*} = n_j J D_{ij} / N$ and $Y_{ij}^{*} = n_j J Y_{ij} / N$ (see Appendix C for the results with general weights). We consider the following linear models for the treatment receipt and outcome,

$$D_{ij}^{*} = \sum_{a=0,1} \gamma_a I(A_j = a) + \sum_{a=0,1} \gamma_{1a} Z_{ij} I(A_j = a) + \xi_{ij}, \qquad (10)$$

$$Y_{ij}^{*} = \sum_{a=0,1} \alpha_a I(A_j = a) + \sum_{a=0,1} \alpha_{1a} Z_{ij} I(A_j = a) + \epsilon_{ij}, \qquad (11)$$

where ξ_{ij} and ϵ_{ij} are error terms.

Unlike the two-step procedure in Basse and Feller (2018), we fit the weighted least squares regression with the following inverse probability weights,

$$w_{ij} = \frac{1}{J_{A_j}} \cdot \frac{1}{n_{jZ_{ij}}}. \qquad (12)$$

The next theorem shows that the resulting weighted least squares estimators are equivalent to the randomization-based ITT effect estimators. Proof is given in Appendix C.1.

Theorem 6 (Weighted Least Squares Regression Estimators for the ITT Analysis). *Let* $\hat{\gamma}^{\text{wls}}$ *and* α^{wls} *be the weighted least squares estimators of the coefficients in the models given in equations (10) and (11), respectively. The regression weights are given in equation (12). Then,*

$$\hat{\gamma}_{1a}^{\text{wls}} = \text{DED}(a), \quad \hat{\gamma}_{a}^{\text{wls}} = \hat{D}(0,a), \quad \hat{\alpha}_{1a}^{\text{wls}} = \text{DEY}(a), \quad \hat{\alpha}_{a}^{\text{wls}} = \hat{Y}(0,a).$$

For the CADE, we consider the weighted two-stage least squares regression where the weights are the same as before and given in equation (12). In our setting, the first-stage regression model is given by equation (10) while the second-stage regression is given by

$$Y_{ij}^{*} = \sum_{a=0,1} \beta_a I(A_j = a) + \sum_{a=0,1} \beta_{1a} D_{ij}^{*} I(A_j = a) + \eta_{ij}, \qquad (13)$$

where η_{ij} is an error term. The weighted two-stage least squares estimators of the coefficients for the model in equation (13) can be obtained by first fitting the model in equation (10) with weighted least squares and then fitting the model in equation (13) again via weighted least squares, in which D_{ij}^{*} is replaced by its predicted values based on the first stage regression model. The following theorem establishes the equivalence between the resulting weighted two-stage least squares regression and randomization-based estimators. Proof is given in Appendix C.2.

Theorem 7 (Weighted Two-stage Least Squares Regression Estimator for the CADE). L*et* β_a^{w2sls} *and* $\beta_{1a}^{\text{w2sls}}$ *be the weighted two-stage least squares estimators*

of the coefficients for the model given in equation (13). The first stage regression model is given in equation (10), and the regression weights are given in equation (12). Then,

$$\hat{\beta}_{1a}^{\text{w2sls}} = \text{CADE}(a), \quad \hat{\beta}_{a}^{\text{w2sls}} = \hat{Y}(0,a) - \text{CADE}(a)\cdot\hat{D}(0,a).$$

3.5.2 Variances

Basse and Feller (2018) show that the cluster-robust HC2 variance (Bell and McCaffrey, 2002) is equal to the randomization-based variance of the average spillover effect estimator under the assumption of equal cluster size. We first generalize this equivalence result to the case where the cluster size varies and then proposes a regression-based variance estimator for the CADE estimator that is equivalent to the randomization-based variance estimator.

We begin by introducing additional notation. Let $\mathbf{X}_j = (\mathbf{X}_{1j},\ldots,\mathbf{X}_{n_j j})^{\cdot}$ be the design matrix of cluster *j* for the model given in equations (10) and (11) with $\mathbf{X}_{ij} = (I(A_j = 1), I(A_j = 0), Z_{ij}I(A_j = 1), Z_{ij}I(A_j = 0))^{\cdot}$. Let $\mathbf{X} = (\mathbf{X}_1^{\cdot},\ldots,\mathbf{X}_J^{\cdot})^{\cdot}$ be the entire design matrix, and $\mathbf{W}_j = \text{diag}(w_{1j},\ldots,w_{n_j j})$ be the weight matrix in cluster *j*, $\mathbf{W} = \text{diag}(\mathbf{W}_1,\ldots,\mathbf{W}_J)$ be the entire weight matrix. We use $\grave{\mathbf{o}}_j = (\grave{o}_{1j},\ldots,\grave{o}_{n_j j})$ to denote the residual vector in cluster *j* obtained from the weighted least squares fit of the model given in equation (11), and $\grave{\mathbf{o}} = (\grave{\mathbf{o}}_1,\ldots,\grave{\mathbf{o}}_J)$ to represent the residual vector for the entire sample.

Using the weights, the cluster-robust generalization of HC2 variance, $\text{var}_{\text{hc2}}^{\text{cluster}}(\boldsymbol{\alpha}^{\text{wls}})$, is given by

$$(\mathbf{X}^{\cdot}\mathbf{W}\mathbf{X})^{-1}\left\{\sum_j \mathbf{X}_j^{\cdot}\mathbf{W}_j(\mathbf{I}_{n_j} - \mathbf{P}_j)^{-1/2}\grave{\mathbf{o}}_j\grave{\mathbf{o}}_j^{\cdot}(\mathbf{I}_{n_j} - \mathbf{P}_j)^{-1/2}\mathbf{W}_j\mathbf{X}_j\right\}(\mathbf{X}^{\cdot}\mathbf{W}\mathbf{X})^{-1},$$

where $\mathbf{I}_{n_j}$ is the $n_j \times n_j$ identity matrix and $\mathbf{P}_j$ is the following cluster leverage matrix,

$$\mathbf{P}_j = \mathbf{W}_j^{1/2}\mathbf{X}_j(\mathbf{X}^{\cdot}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}_j^{\cdot}\mathbf{W}_j^{1/2}.$$

It can be shown that $\text{var}_{\text{hc2}}^{\text{cluster}}(\hat{\alpha}_{1a}^{\text{wls}}) = \hat{\sigma}_{DE}^2(a)/J_a$, representing the between-cluster sample variance. However, as shown in Theorem 4, $\text{var}\left\{\text{DEY}(a)\right\}$ is a weighted average of between-cluster and within-cluster sample variances. Thus, unlike the results in Basse and Feller (2018), the cluster-robust HC2 variance no longer equals the randomization-based variance estimator, because it only takes into account the between-cluster variance.

To address this problem, we introduce the following individual-robust HC2 variance,

$$\text{var}_{\text{hc2}}^{\text{ind}}(\boldsymbol{\alpha}^{\text{wls}}) = (\mathbf{X}^{\cdot}\mathbf{W}\mathbf{X})^{-1}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n_j} w_{ij}^2\hat{o}_{ij}^{*2}(1-P_{ij})^{-1}\mathbf{X}_{ij}\mathbf{X}_{ij}^{\cdot}\right\}(\mathbf{X}^{\cdot}\mathbf{W}\mathbf{X})^{-1},$$

where $P_{ij} = w_{ij}\mathbf{X}_{ij}^{\cdot}(\mathbf{X}_j^{\cdot}\mathbf{W}_j\mathbf{X}_j)^{-1}\mathbf{X}_{ij}$ is the individual leverage and $\hat{o}_{ij}^{*} = \hat{o}_{ij} - \sum_{i'=1}^{n_j}\hat{o}_{i'j}I(Z_{i'j}=z)/n_{jz}$ is the adjusted residuals for $Z_{ij}=z$ so that we have $\mathbf{X}_j^{\cdot}\hat{\boldsymbol{o}}_j^{*} = \boldsymbol{0}_4$. The next theorem establishes that the weighted average of the cluster-robust and individual-robust HC2 variance estimators is numerically equivalent to the randomization-based variance estimator.

Theorem 8 (Regression-based Variance Estimators for the ITT Effects). *The randomization-based variance estimator of the direct effect is a weighted average of the cluster-robust and individual-robust HC2 variances,*

$$\text{var}\left\{\text{DED}(a)\right\} = \left(1-\frac{J_a}{J}\right)\text{var}_{\text{hc2}}^{\text{cluster}}(\hat{\gamma}_{1a}^{\text{wls}}) + \frac{J_a}{J}\text{var}_{\text{hc2}}^{\text{ind}}(\hat{\gamma}_{1a}^{\text{wls}}),$$

$$\text{var}\left\{\text{DEY}(a)\right\} = \left(1-\frac{J_a}{J}\right)\text{var}_{\text{hc2}}^{\text{cluster}}(\hat{\alpha}_{1a}^{\text{wls}}) + \frac{J_a}{J}\text{var}_{\text{hc2}}^{\text{ind}}(\hat{\alpha}_{1a}^{\text{wls}}).$$

Proof is given in Appendix C.3.

To gain some intuition about the weighted average of two robust variances, consider the following model commonly used for split-plot designs,

$$Y_{ij}^{*} = \sum_{a=0,1} \alpha_a I(A_j = a) + \sum_{a=0,1} \alpha_{1a} Z_{ij} I(A_j = a) + \epsilon_{Bj} + \epsilon_{Wij},$$

where ϵ_{Bj} represents the random effects for whole plots (or clusters), and ϵ_{Wij} is the random effects for split-plots (or individuals). The cluster-robust HC2 variance is related to ϵ_{Bj} and the individual-robust HC2 variance is related to ϵ_{Wij}. In Appendix C.4, we discuss the connection between the random effects model and the randomization-based inference and explain why the adjustment for δ_{ij}^{*} is necessary.

Finally, we consider the weighted two-stage least squares regression given in equations (10) and (13). Let $\mathbf{M}_j = (\mathbf{M}_{1j}, \ldots, \mathbf{M}_{n_j j})^{\cdot}$ be the design matrix for cluster j in the second-stage regression with $\mathbf{M}_{ij}^{\cdot} = (I(A_j = 1), I(A_j = 0), \hat{D}_{ij}^{*} I(A_j = 1), \hat{D}_{ij}^{*} I(A_j = 0))$ where $\hat{D}_{ij}^{*}$ represents the fitted value given in equation (10). Let $\mathbf{M} = (\mathbf{M}_1^{\cdot}, \ldots, \mathbf{M}_J^{\cdot})^{\cdot}$ be the entire design matrix. We define the cluster-robust HC2 variance, $\mathrm{var}_{\mathrm{hc2}}^{\mathrm{cluster}}(\boldsymbol{\beta}^{\mathrm{w2sls}})$, as

$$(\mathbf{M}^{\cdot}\mathbf{W}\mathbf{M})^{-1} \left\{ \sum_{j=1}^{J} \mathbf{M}_j^{\cdot} \mathbf{W}_j (\mathbf{I}_{n_j} - \mathbf{Q}_j)^{-1/2} \boldsymbol{\eta}_j \boldsymbol{\eta}_j^{\cdot} (\mathbf{I}_{n_j} - \mathbf{Q}_j)^{-1/2} \mathbf{W}_j \mathbf{M}_j \right\} (\mathbf{M}^{\cdot}\mathbf{W}\mathbf{M})^{-1}, \qquad (14)$$

where $\mathbf{Q}_j$ is the cluster leverage matrix,

$$\mathbf{Q}_j = \mathbf{W}_j^{1/2} \mathbf{M}_j (\mathbf{M}^{\cdot}\mathbf{W}\mathbf{M})^{-1} \mathbf{M}_j^{\cdot} \mathbf{W}_j^{1/2}.$$

The individual-robust HC2 variance is given by,

$$\mathrm{var}_{\mathrm{hc2}}^{\mathrm{ind}}(\boldsymbol{\beta}^{\mathrm{w2sls}}) = (\mathbf{M}^{\cdot}\mathbf{W}\mathbf{M})^{-1} \left\{ \sum_{j=1}^{J} \sum_{i=1}^{n_j} w_{ij}^2 \hat{\eta}_{ij}^{*2} (1 - Q_{ij})^{-1} \mathbf{M}_{ij} \mathbf{M}_{ij}^{\cdot} \right\} (\mathbf{M}^{\cdot}\mathbf{W}\mathbf{M})^{-1},$$

where $Q_{ij} = w_{ij}\mathbf{M}_{ij}^{\bullet}(\mathbf{M}_j^{\bullet}\mathbf{W}_j\mathbf{M}_j)^{-1}\mathbf{M}_{ij}$ is the individual leverage and $\hat{\eta}_{ij}^{*} = \hat{\eta}_{ij} - \sum_{i'=1}^{n_j}\hat{\eta}_{i'j}I(Z_{i'j} = z)/n_{jz}$ for $Z_{ij} = z$ is the adjusted residual with $\mathbf{X}_j^{\bullet}\boldsymbol{\eta}_j^{*} = \boldsymbol{0}_4$.

As in the case of ITT analysis, we can show that the weighted average of cluster-robust and individual-robust variance estimators is numerically equivalent to the randomization-based variance estimator.

Theorem 9 (Regression-based Variance Estimator for the CADE). *The randomization-based variance estimator of the average complier direct effect is a weighted average of the cluster-robust and individual-robust HC2 variances,*

$$\text{var}\left\{\text{CADE}(a)\right\} = \left(1 - \frac{J_a}{J}\right)\text{var}_{\text{hc2}}^{\text{cluster}}(\hat{\beta}_{1a}^{\text{w2sls}}) + \frac{J_a}{J}\text{var}_{\text{hc2}}^{\text{ind}}(\hat{\beta}_{1a}^{\text{w2sls}}).$$

Proof is given in Appendix C.5.

4 Empirical Analysis

In this section, we analyze the data introduced in Section 2 by applying the proposed methodology. We focus on the annual household hospital expenditure, which ranges from 0 to INR 500,000 with the median value of 1,000. The outcome is missing for 926 households, which is less than 10% of the sample. For simplicity, we discard the observations with missing data from the current analysis and leave the development of a method for analyzing two-stage randomized experiments with missing data to future research.

As expected, the enrollment rate in the villages assigned to the “High” assignment mechanism is 67.0%, whereas the enrollment rate in the villages under the “Low” assignment mechanism is just 46.2%. Because the encouragement proportion is 80% under the “High” assignment mechanism and 40% under the “Low” assignment mechanism, this implies the existence of two-sided noncompliance, in which some households in the treatment group did not receive the treatment and others in the control group managed to receive it.

Table 2 presents the estimates of ITT effects and complier average direct and spillover effects. We show the results for both the individual/household-weighted and cluster/village-weighted estimands. In the top row, we show the estimated average direct effect on enrollment in RSBY under the "High" treatment mechanism ($\mathrm{DED}(1)$) and under the "Low" treatment mechanism ($\mathrm{DED}(0)$) as well as the estimated average spillover effects under the treatment condition ($\mathrm{SED}(1)$) and the control condition ($\mathrm{SED}(0)$). The treatment assignment is estimated to increase the enrollment rate by more than 40 percentage points. This quantity represents the estimated proportions of compliers.

Interestingly, we find that the average spillover effects are estimated to be positive and, sometimes, statistically significant. In particular, the village-weighted average spillover effect on enrollment is 4.4 percentage points with the standard error of 1.8 under the treatment condition. The finding implies that assigning a greater proportion of households to the treatment condition makes another household of the same village more likely to enroll in RSBY especially if the latter is also encouraged to enroll.[3]

The middle row of Table 2 presents the estimated ITT effects on the outcome. The estimated average direct effect under the "Low" assignment mechanism ($\mathrm{DEY}(0)$) tends to be positive where the village-weighted estimate is statistically significant. In contrast, the estimated average direct effect under the "High" assignment mechanism ($\mathrm{DEY}(1)$) is negative although not statistically significant.

One possible explanation for this difference is that the assignment to the treatment condition makes people visit hospitals more often and spend more on healthcare so long as fewer households within the same village are assigned to the treatment. When a large number of households within the same village are assigned to the treatment condition, the overcrowding of hospitals may reduce hospital visits of each treated household.

We examine the plausibility of this explanation by estimating the direct effect of the treatment assignment on the number of hospital visits. The estimated direct effect under the “High” treatment assignment mechanism is –0.157, whereas that under the “Low” treatment assignment mechanism is 0.132. Although these estimates are not statistically significant, they provide suggestive evidence consistent with the overcrowding hypothesis.

The bottom row of Table 2 presents the estimates of the complier average direct and spillover effects. The village-weighted complier average direct effect under the “Low” assignment mechanism ($\mathrm{CADE}(0)$) is positive and statistically significant, implying that enrollment in RSBY directly increases the household hospital expenditure when few households are assigned to the treatment condition. In contrast, the complier average direct effect under the “High” assignment mechanism ($\mathrm{CADE}(1)$) is negative. This difference is consistent with the overcrowding hypothesis discussed above.

In addition, we also estimate the complier average spillover effects. Unfortunately, they are imprecisely estimated, making it difficult to draw a definite conclusion about whether or not the proportion of treated households in a village directly affects one’s outcome among those who enroll in RSBY only when a greater proportion of households is encouraged to sign up for the insurance program.

Because most of the estimates are not statistically significant, it is difficult to draw a definitive conclusion. However, our analysis provides some suggestive policy recommendations. First, the estimated positive spillover effect of encouragement on enrollment suggests that the government could increase the enrollment rate by leveraging existing social networks among households within each village. Second, the estimated negative CADE under the “High” treatment assignment mechanism condition suggests that there might be overcrowding of local

hospitals when many households newly enroll in the RSBY. The government can address this issue by increasing the capacity of local hospitals.

In addition to the quantities in our analysis above, we may also be interested in other quantities, e.g., the average spillover effect of the treatment assignment when all households are assigned to the treatment condition versus households are assigned to the treatment condition, and the direct effect of one's own treatment receipt when the treatment receipts of other households are fixed at some constant levels. Unfortunately, without modeling assumptions, we are unable to identify these quantities. In Appendix E, we propose a model-based approach and estimate these quantities using our application data.

5 Concluding Remarks

In this paper, we consider two-stage randomized experiments with noncompliance and interference. We merge two strands of the causal inference literature, one on experiments with noncompliance and the other on experiments with interference. We introduce new causal quantities of interest, propose nonparametric identification results and consistent estimators, and derive their variances. We connect these randomization-based estimators to two-stage least squares regressions that are commonly used by applied researchers. We apply the proposed methodology to evaluate the efficacy of the India's National Health Insurance program (RSBY) and find some evidence of spillover effects. We believe that the proposed methodology can help applied researchers make best use of this effective experimental design for studying interference problems. In future research, it is of interest to relax the assumption of partial interference and allow for interference between units of different clusters.

References

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association* **91**, 434, 444–455.

Aronow, P. and Samii, C. (2017). Estimating average causal effects under general interference. *Annals of Applied Statistics* **11**, 4, 1912–1947.

Aronow, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research* **41**, 1, 3–16.

Athey, S., Eckles, D., and Imbens, G. W. (2017). Exact p-values for network interference. *Journal of the American Statistical Association* Forthcoming.

Basse, G. and Feller, A. (2018). Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association* **113**, 521, 41–55.

Bell, R. M. and McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology* **28**, 2, 169–181.

Berman, P., Ahuja, R., and Bhandari, L. (2010). The impoverishing effect of healthcare payments in india: new methodology and findings. *Economic and Political Weekly* 65–71.

Chin, A. (2018). Central limit theorems via stein's method for randomized experiments under interference. *arXiv preprint arXiv:1804.03105*.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, London.

Forastiere, L., Airoldi, E. M., and Mealli, F. (2016). Identification and estimation of treatment and interference effects in observational studies on networks. *arXiv preprint arXiv:1609.06245*.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 361–374.

Halloran, M. E. and Struchiner, C. J. (1995). Causal inference in infectious diseases. *Epidemiology* **6**, 2, 142–151.

Holland, P. W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association* **81**, 945–960.

Hong, G. (2015). *Causality in a social world: Moderation, mediation and spill-over*. John Wiley & Sons.

Hong, G. and 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**, 475, 901–910.

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association* **103**, 482, 832–842.

Imai, K. and Jiang, Z. (2018). experiment: R package for designing and analyzing randomized experiments. available at the Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=experiment.

Jowett, M. (2003). Do informal risk sharing networks crowd out public voluntary health insurance? evidence from Vietnam. *Applied Economics* **35**, 10, 1153–1161.

Kang, H. and Imbens, G. (2016). Peer encouragement designs in causal inference with partial interference and identification of local average network effects. *arXiv preprint arXiv:1609.04464*.

Lehmann, E. L. (2004). *Elements of large-sample theory*. Springer Science & Business Media.

Li, X. and Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association* **112**, 520, 1759–1769.

Lin, W., Liu, Y., and Meng, J. (2014). The crowding-out effect of formal insurance on informal risk sharing: An experimental study. *Games and Economic Behavior*.

Liu, L. and Hudgens, M. G. (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association* **109**, 505, 288–301.

Neyman, J. (1923). On the application of probability theory to agricultural experiments: Essay on principles, section 9. (translated in 1990). *Statistical Science* **5**, 465–480.

Ohlsson, E. (1989). Asymptotic normality for two-stage sampling from a finite population. *Probability theory and related fields* **81**, 3, 341–352.

Rosenbaum, P. R. (2007). Interference beteween units in randomized experiments. *Journal of the American Statistical Association* **102**, 477, 191–200.

Rubin, D. B. (1990). Comments on “On the application of probability theory to agricultural experiments. Essay on principles. Section 9” by J. Splawa-Neyman translated from the Polish and edited by D. M. Dabrowska and T. P. Speed. *Statistical Science* **5**, 472–480.

Sävje, F., Aronow, P. M., and Hudgens, M. G. (2017). Average treatment effects in the presence of unknown interference. *arXiv preprint arXiv:1711.06399*.

Shahrawat, R. and Rao, K. D. (2011). Insured yet vulnerable: out-of-pocket payments and india's poor. *Health policy and planning* **27**, 3, 213–221.

Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association* **101**, 476, 1398–1407.

Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2010). On causal inference in the presence of interference. *Statistical Methods in Medical Research* **21**, 1, 55–75.

Vanderweele, T. J., Hong, G., Jones, S. M., and Brown, J. L. (2013). Mediation and spillover effects in group-randomized trials: A case study of the 4rs educational intervention. *Journal of the American Statistical Association* **108**, 502, 469–482.

Notes

[1] For a more detailed description of the design, see the pre-analysis plan posted on the American Economic Association's Registry at https://www.socialscienceregistry.org/trials/1793

[2] For the sake of simplicity, we analyze the dichotomized assignment. In the original experiment, households could be assigned to any of four groups; group A was given RSBY for free, group B was given RSBY for free and a cash transfer equal to the premium on insurance, group C was sold RSBY for the same premium as the government paid for RSBY coverage, and group D had no intervention. Here, the High assignment is defined as villages where 80% of households were assigned to groups A and B, whereas the Low assignment is defined as villages with 60% of households assigned to groups C and D.

[3] This result should be viewed as an illustration that spillovers are possible in this study. We leave a more complete examination of spillover effects to subsequent work based on the pre-specified analysis.

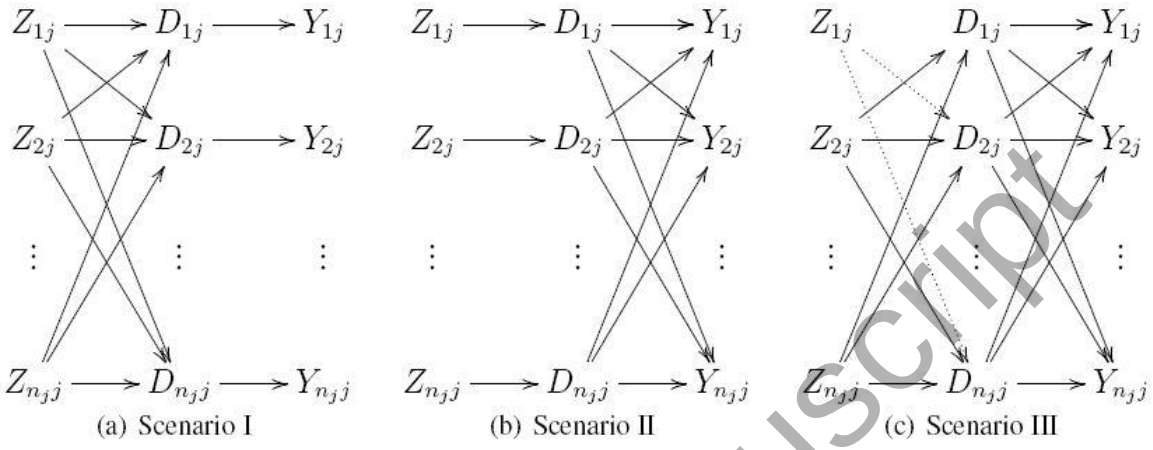

Fig. 1 Three Scenarios that Imply Assumption 5: (a) no spillover effect of the treatment receipt on the outcome; (b) no spillover effect of the treatment assignment on the treatment receipt; (c) if treatment assignment of a unit does not affect his own treatment receipt, then it should not affect other units' treatment receipts, i.e., the dotted edges do not exist.

Table 1 Two-Stage Randomization Design.

Village-level arms		Household-level arms			
Mechanisms	Number of villages	Treatment	Control	Number of households	Enrollment rates
High	219	80%	20%	5,714	67.0%
Low	216	40%	60%	5,373	46.2%

Table 2 Estimated Intention-to-Treat (ITT) and Complier Average Direct and Spillover Effects. For the "household-weighted" estimates, we equally weight households, whereas each village is equally weighted for the "village-weighted" estimates. The top row presents the average direct effects on enrollment in RSBY under the high treatment mechanism ($DED(1)$) and under the low treatment mechanism ($DED(0)$) as well as the average spillover effects under the treatment ($SED(1)$) and control ($SED(0)$) conditions. The middle row presents the same set of ITT estimates for hospital expenditure. Finally, the bottom row presents the complier average direct effect under the "High" ($CADE(1)$) and "Low" ($CADE(0)$) treatment assignment mechanisms as well as the complier average spillover effect under the treatment ($CASE(1)$) and control ($CASE(0)$) conditions.

Enrollment in RSBY	$DED(1)$	$DED(0)$	$SED(1)$	$SED(0)$
household-weighted	0.482 (0.023)	0.441 (0.021)	0.086 (0.053)	0.045 (0.028)
village-weighted	0.457 (0.019)	0.445 (0.017)	0.044 (0.018)	0.031 (0.021)
Hospital expenditure	$DEY(1)$	$DEY(0)$	$SEY(1)$	$SEY(0)$
household-weighted	–795 (514)	875 (530)	–1374 (823)	297 (858)
village-weighted	–222 (575)	1666 (734)	–1677 (972)	211 (761)
Hospital expenditure	$CADE(1)$	$CADE(0)$	$CASE(1)$	$CASE(0)$
household-weighted	–1649 (1061)	1984 (1215)	–15900 (15342)	6568 (18305)
village-weighted	–485 (1258)	3752 (1652)	–38341 (26845)	6846 (25042)