CAUSAL INFERENCE FROM OBSERVATIONAL STUDIES WITH CLUSTERED INTERFERENCE, WITH APPLICATION TO A CHOLERA VACCINE STUDY

By Brian G. Barkley[1], Michael G. Hudgens[2], John D. Clemens[3], Mohammad Ali[4] and Michael E. Emch[5]

[1]*Department of Marketing Analytics, Kohl's, Inc., barkleybg@outlook.com*

[2]*Department of Biostatistics, University of North Carolina at Chapel Hill, mhudgens@email.unc.edu*

[3]*Office of the Executive Director, icddr,b, jclemens@icddrb.org*

[4]*Department of International Health, Johns Hopkins University, mali25@jhu.edu*

[5]*Department of Geography, University of North Carolina at Chapel Hill, emch@unc.edu*

Understanding the population-level effects of vaccines has important public health policy implications. Inferring vaccine effects from an observational study is challenging because participants are not randomized to vaccine (i.e., treatment). Observational studies of infectious diseases present the additional challenge that vaccinating one participant may affect another participant's outcome, that is, there may be interference. In this paper recent approaches to defining vaccine effects in the presence of interference are considered, and new causal estimands designed specifically for use with observational studies are proposed. Previously defined estimands target counterfactual scenarios in which individuals independently choose to be vaccinated with equal probability. However, in settings where there is interference between individuals within clusters, it may be unlikely that treatment selection is independent between individuals in the same cluster. The proposed causal estimands instead describe counterfactual scenarios which allow for within-cluster dependence in the individual treatment selections. These estimands may be more relevant for policy-makers or public health officials who desire to quantify the effect of increasing the proportion of vaccinated individuals in a population. Inverse probability-weighted estimators for these estimands are proposed. The large-sample properties of the estimators are derived, and a simulation study demonstrating the finite-sample performance of the estimators is presented. The proposed methods are illustrated by analyzing data from a study of cholera vaccination in over 100,000 individuals in Bangladesh.

1. Introduction. Inferring causal effects from an observational (i.e., nonrandomized or nonexperimental) study is challenging because participants may select their own treatment. Observational studies in many settings, such as infectious disease research, present the additional challenge that one individual's treatment may have an effect on another individual's outcome, that is, there may be interference (Cox (1958)). For example, whether one individual is administered a vaccine may affect whether another individual develops disease from some infectious pathogen. In certain settings it may be reasonable to assume that individuals can be partitioned into clusters such that there may be interference among individuals within a single cluster yet no interference between individuals in distinct clusters. Sobel (2006) described this assumption as "partial interference;" here, this assumption is referred to as "clustered interference." Clusters might entail households, classrooms, geographical areas or other hierarchical structures. For example, in an assessment of the effect of retaining low-achieving children in kindergarten, Hong and Raudenbush (2006) assumed no interference

between students in different schools. Several types of treatment effects (i.e., causal estimands) have been proposed for the setting where there may be clustered interference; for example, see Halloran and Struchiner (1995), Hudgens and Halloran (2008) and Tchetgen Tchetgen and VanderWeele (2012).

Methods have been developed for inference about these causal effects from observational studies (Liu, Hudgens and Becker-Dreps (2016), Perez-Heydrich et al. (2014), Tchetgen Tchetgen and VanderWeele (2012)). One drawback of the treatment effects targeted by these methods is that these causal estimands describe counterfactual scenarios in which individuals select treatment independently and with the same probability. However, in settings where interference within clusters is plausible, it may be unlikely that treatment selection among individuals in the same cluster is independent (Liu, Hudgens and Becker-Dreps (2016)). For instance, suppose a public health policy maker is interested in the effect of seasonal influenza vaccination on risk of influenza-like illness in households. In this case one might expect positive correlation between the vaccination statuses of individuals in the same household. Thus, drawing inference to a counterfactual scenario in which individuals are administered vaccines independently may not be of public health relevance. In this paper new causal estimands are proposed for observational studies where there may be clustered interference; these estimands describe counterfactual scenarios which allow for within-cluster dependence in the individual treatment selections. The proposed estimands may be more relevant for policy makers or public health officials who are interested in quantifying the effect of increasing the proportion of treated individuals in a population. Papadogeorgou, Mealli and Zigler (2019) independently developed similar estimands and methods with motivation from and applications in air pollution epidemiology.

The methods developed here are motivated by a cholera vaccine study in Matlab, Bangladesh, which featured both an experimental and a nonexperimental component (Ali et al. (2005, 2009)). Included in the study were 121,975 women (aged 15 years and older) and children (aged 2–15 years) from 6,415 baris (i.e., households of patrilineally-related individuals). These individuals were eligible to participate in the experimental component of the study, in which each individual was randomized with equal probability to one of three treatment arms: B subunit-killed whole-cell oral cholera vaccine, killed whole-cell-only oral cholera vaccine or placebo. Individuals who did not participate did not receive either version of active treatment. The study collected endpoint data of cholera infection on all individuals, even those who did not participate in the experimental component. Since participation was not controlled by study design and nearly two-fifths of all individuals declined to participate, there was a notable nonexperimental component to the study and potential for confounding exists when analyzing the endpoint data.

The goal of the analysis presented here is to assess the effects of cholera vaccination while allowing for the possibility of within-bari interference, as there is evidence that transmission of cholera often takes place within baris (Ali et al. (2005)). Effects of vaccination due to interference are of particular interest; for instance, does increasing the proportion of individuals vaccinated (i.e., vaccine "coverage") within a bari lead to decreased risk of cholera among individuals not vaccinated? Figure 1 depicts the empirical distributions of the number of individuals and of the vaccine coverage within the baris. In Figure 1 and in the analyses below, any individual who received at least two doses of either of the two cholera vaccines was considered to be treated and, otherwise, was considered to be untreated (Perez-Heydrich et al. (2014)).

The outline of the remainder of this paper is as follows. In Section 2 the potential outcomes framework and interference are discussed. The proposed causal estimands are introduced in Section 3. A set of assumptions sufficient for identifying the target estimands is presented in Section 4. In Section 5 inverse probability-weighted estimators are introduced; the estimators are shown in the Appendix to be consistent and asymptotically normal. Simulations in
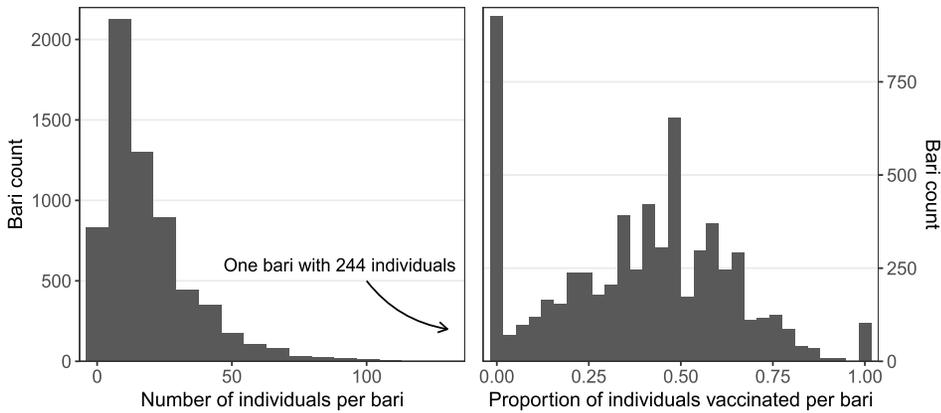
FIG. 1. *Matlab cholera vaccine study. Left: number of individuals per cluster (bari). Right: proportion of individuals vaccinated per cluster.*

Section 6 demonstrate that the proposed estimators are empirically unbiased and that Wald-type confidence intervals attain nominal coverage levels in finite samples. Analysis of the Bangladesh cholera vaccine study is presented in Section 7. Section 8 concludes with a discussion.

2. Counterfactuals and interference. Consider a super-population of clusters of individuals. For each cluster i let N_i equal the number of individuals in the cluster which may vary across clusters in the super-population. For example, in the cholera vaccine study clusters are defined by baris, and N_i is the number of women and children living in the ith bari. For cluster i let $A_i = (A_{i1}, A_{i2}, \ldots, A_{iN_i})$ where A_{ij} denotes the binary treatment (e.g., vaccination) indicator for individual j in the cluster, and $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iN_i})$ where Y_{ij} is the outcome of interest for individual j. For example, Y_{ij} might indicate whether or not individual j in cluster i experienced the outcome after some suitable follow-up period after treatment exposure status was observed. In the data analysis in Section 7, $A_{ij} = 1$ denotes the individual was vaccinated against cholera (and 0 otherwise), and $Y_{ij} = 1$ denotes the individual who became infected with cholera within a one-year follow-up period (and 0 otherwise).

Assuming clustered interference, the potential outcome for an individual may depend on the individual's own treatment exposure status as well as on the treatment exposures of others in the same cluster. However, any individual's potential outcomes are assumed to be unaffected by the treatment exposures of individuals in different clusters. For instance, in the cholera study analysis women and children in one bari are assumed to be unaffected by the vaccination of individuals in other baris. Let $\mathcal{A}(N_i)$ be the set of all vectors with N_i binary entries such that $a = (a_1, a_2, \ldots, a_{N_i}) \in \mathcal{A}(N_i)$ is a vector whose entries each indicates a potential treatment status for an individual in a cluster of N_i individuals. Let $Y_{ij}(a)$ be the potential outcome for unit j in cluster i if, possibly counter to fact, cluster i had received $a \in \mathcal{A}(N_i)$. In the absence of interference, $Y_{ij}(a) = Y_{ij}(a')$ whenever $a_j = a'_j$ for any $a, a' \in \mathcal{A}(N_i)$. However, assuming no interference when interference is present may result in biased estimates of causal effects. Throughout this paper interference is assumed to be absent between individuals in different clusters, but no additional assumptions are made about the nature of interference within clusters (such as stratified interference (Hudgens and Halloran (2008))).

3. Causal effects.

3.1. *Policies of interest.* Our goal is to draw inference about the difference in expected outcomes arising from population-level policies which change the distribution of treatment.

In particular, we are interested in assessing the effect of changes in the level of cholera vaccine coverage in settings such as Matlab, Bangladesh. In the absence of interference, typical treatment effect estimands compare the policy (or strategy) where all individuals receive treatment (i.e., $A_i = (1, 1, \ldots, 1)$ with probability 1) with the policy where all individuals are not treated (i.e., $A_i = (0, 0, \ldots, 0)$ with probability 1). Here, we consider more general policies where individuals receive treatment according to some probability that, in general, differs between individuals; such policies are sometimes referred to as "stochastic interventions" (Díaz and van der Laan (2012)). This will allow comparison of the effects of different levels of cholera vaccine coverage, for example, 30% vs. 60%.

Tchetgen Tchetgen and VanderWeele (2012) considered the stochastic intervention policy where individuals select treatment with the same, fixed probability (e.g., 0.5). Here, we consider policies where individuals may select treatment with different probabilities, conditional on their pretreatment covariates. For example, older women may be more likely than younger women to receive the cholera vaccine. We let α denote the marginal probability of selecting treatment, averaged over all individuals in the population, for a particular policy. The notation $\Pr_\alpha(\cdot)$ is used to indicate that a probability is with respect to the counterfactual scenario in which the policy α is implemented.

3.2. *Proposed estimands.* For $a \in \mathcal{A}(N_i)$, define $\omega(a, N_i, \alpha) = \Pr_\alpha(A_i = a | N_i)$ to be the probability under policy α that a cluster of N_i individuals experiences treatment status a. Let $\bar{Y}_i(a) = N_i^{-1} \sum_{j=1}^{N_i} Y_{ij}(a)$ denote the average potential outcome in a cluster if the cluster had been exposed to a. The expected potential outcome under α for a single cluster of N_i individuals is defined to be $\bar{Y}_i(\alpha) = \sum_{a \in \mathcal{A}(N_i)} \bar{Y}_i(a) \omega(a, N_i, \alpha)$. In other words, $\bar{Y}_i(\alpha)$ is the expected average potential outcome for the cluster in the counterfactual scenario in which α is implemented.

Define the population mean outcome under α to be $\mu(\alpha) = \mathbb{E}\{\bar{Y}_i(\alpha)\}$, where the expected value is taken over all clusters i in the super-population. That is, each cluster in the super-population is effectively given equal weight in the population mean outcome. An alternative approach would be to define population mean outcomes where each individual is given equal weight, such that larger clusters are given proportionally more weight than smaller clusters (Liu, Hudgens and Becker-Dreps (2016), §2)); in the special case where N_i does not vary across clusters, these two approaches to defining the population mean outcome are equivalent. The overall effect is defined to be $\mathrm{OE}(\alpha, \alpha') = \mu(\alpha) - \mu(\alpha')$ which represents the difference in expected potential outcomes under policy α vs. policy α'. The overall effect is defined here as a difference in mean potential outcomes but could instead be defined as a ratio or some other contrast (Liu, Hudgens and Becker-Dreps (2016)).

In addition, it may also be of interest to consider potential outcomes among only the untreated individuals within a cluster. For instance, in the cholera vaccine study we are interested in the effect of vaccine coverage on women and children who do not receive the vaccine. Let $\bar{Y}_{i,t}(a) = \{\sum_{j=1}^{N_i} I(a_j = t)\}^{-1} \sum_{j=1}^{N_i} Y_{ij}(a) I(a_j = t)$ for $t = 0, 1$. In words, $\bar{Y}_{i,0}(a)$ is the average potential outcome among the untreated individuals within the cluster; likewise $\bar{Y}_{i,1}(a)$ is the average potential outcome among the treated individuals within the cluster. In the special case when $a = (1 - t, 1 - t, \ldots, 1 - t)$, define $\bar{Y}_{i,t}(a) = 0$ for each of $t = 0, 1$. Denote the population mean potential outcomes when untreated to be $\mu_0(\alpha) = \mathbb{E}\{\sum_{a \in \mathcal{A}(N_i)} \bar{Y}_{i,0}(a) \omega(a, N_i, \alpha)\}$. The spillover effect when untreated is defined to be the difference in population mean potential outcomes when untreated under policy α vs. α', that is, $\mathrm{SE}_0(\alpha, \alpha') = \mu_0(\alpha) - \mu_0(\alpha')$. Similarly, let $\mu_1(\alpha) = \mathbb{E}\{\sum_{a \in \mathcal{A}(N_i)} \bar{Y}_{i,1}(a) \omega(a, N_i, \alpha)\}$, and define $\mathrm{SE}_1(\alpha, \alpha') = \mu_1(\alpha) - \mu_1(\alpha')$ to be the spillover effect when treated.

3.3. *Relation to previous estimands.* Consider a policy in which all individuals in a cluster are exposed to treatment independently with the same probability; Tchetgen Tchetgen and VanderWeele (2012) refer to this as a "type B parameterisation." For $\alpha \in [0, 1]$, let $\omega_{\mathrm{B}}(a, N_i, \alpha) = \prod_{j=1}^{N_i} \alpha^{a_j}(1 - \alpha)^{1-a_j}$ denote the counterfactual probabilities under such a type B policy. Likewise, let $\mu_{\mathrm{B}}(\alpha) = \mathbb{E}\{\sum_{a \in \mathcal{A}(N_i)} \bar{Y}(a)\omega_{\mathrm{B}}(a, N_i, \alpha)\}$ be the population mean potential outcome for a type B policy, and define the overall effect with respect to two type B policies to be $\mathrm{OE}_{\mathrm{B}}(\alpha, \alpha') = \mu_{\mathrm{B}}(\alpha) - \mu_{\mathrm{B}}(\alpha')$.

The policies of interest in this paper include as a special case type B policies where treatment exposure is uncorrelated. The estimands proposed in this paper can thus be seen as a generalization of the type B estimands, as the type B policies describe only the limiting counterfactual scenarios in which there is no within-cluster dependence of individual treatment selections. In general, $\omega(a, n, \alpha) \neq \omega_{\mathrm{B}}(a, n, \alpha)$ for almost any triplet (a, n, α), and the corresponding policies, estimands and interpretations differ. In the data analysis of the cholera vaccine study in Section 7, estimates of the type B estimands are presented for comparison to the estimates of the proposed estimands.

The estimands in Section 3.2 are similar to those independently proposed by Papadogeorgou, Mealli and Zigler (2019). However, Papadogeorgou et al. consider counterfactual scenarios where each cluster's average individual-level propensity score equals exactly α. In some settings such a scenario may be unrealistic. For instance, it is unlikely that each cluster in the cholera vaccine study would share the same average individual-level propensity score. The likelihood that individuals would be vaccinated could depend on a myriad of individual level factors (such as age, sex, etc) and one would not expect, in general, the distribution of these individual level variables to be the same across baris. In contrast, the estimands defined in Section 3.2 consider the less restrictive counterfactual setting where the marginal probability of selecting treatment equals α, but any cluster's average individual-level propensity scores may not equal α exactly. Another difference between the estimands in Section 3.2 and those in Papadogeorgou, Mealli and Zigler (2019) relates to how average potential outcomes are defined. Papadogeorgou et al. first construct individual average potential outcomes by computing a unit's expected outcome under policy α, conditional on that unit's treatment (i.e., conditional on $A_{ij} = a$). These individual average potential outcomes are then averaged first within and then across groups. The estimands defined in Section 3.2 are not formulated in this manner and thus avoid avoid difficulties in interpretation that can arise when computing expected outcomes conditional on $A_{ij} = a$ (Eck, Morozova and Crawford (2018), VanderWeele and Tchetgen Tchetgen (2011)).

4. Identifiability. In this section we describe a set of assumptions sufficient for identifiability of the estimands defined in Section 3.2. The assumptions include a cluster level version of the usual no unmeasured confounders assumption and a correctly specified parametric model of the conditional distribution of treatment given covariates.

Let there be a random sample of $i = 1, \ldots, M$ clusters, and denote by $O_i = \{N_i, L_i, A_i, Y_i\}$ the observed values of the random variables for cluster i, where $L_i = (L_{i1}, L_{i2}, \ldots, L_{iN_i})$ is an N_i-vector of pretreatment variables. In general, L_{ij} may include individual-level variables associated with individual j, such as their age, or cluster-level variables, such as the average age of individuals in cluster i. The ordering of individuals in each cluster is assumed to be uninformative. As in Tchetgen Tchetgen and VanderWeele (2012) and Perez-Heydrich et al. (2014), assume exchangeability at the cluster level conditional on the baseline variables

$$Y_i(a) \perp A_i | L_i, N_i \quad \text{for any } i \text{ and any } a \in \mathcal{A}(N_i).$$

In addition, assume positivity at the cluster level

$$\Pr(A_i = a | L_i, N_i) > 0 \quad \text{for any } a \in \mathcal{A}(N_i).$$

Plausibility of the exchangeability and positivity assumptions in the context of the cholera vaccine study is discussed in Section 8.

4.1. *Model for observed treatment.* Following Tchetgen Tchetgen and VanderWeele (2012), Perez-Heydrich et al. (2014) and Liu, Hudgens and Becker-Dreps (2016), assume the following mixed effects model for treatment:

$$\mathcal{L}\big(\text{Pr}(A_{ij} = 1 | L_i, N_i, b_i)\big) = \beta_0 + \beta_1 L_{ij} + b_i, \tag{1}$$

where $\mathcal{L}$ is some suitable link function and b_i denotes a random intercept for cluster i which is assumed to follow distribution Φ with mean zero and parameter(s) σ. The random effect b_i allows for correlation in treatment selection between individuals within the same cluster. In the cholera vaccine study analysis presented below, $\mathcal{L}$ is the logit link, and b_i is assumed to have a Gaussian distribution with mean zero and standard deviation σ; this model can be fit using the R package `lme4` (Bates et al. (2015)) with `glmer(data = data, formula = Treatment ~ Covariate + (1|ClusterID), family = "binomial")`.

Under the assumed mixed model the conditional probability that a cluster's treatment vector equals $a \in \mathcal{A}(N_i)$ is

$$
\begin{aligned}
&\text{Pr}(A_i = a | L_i, N_i) \\
&= \int \prod_{j=1}^{N_i} \mathcal{L}^{-1}(\beta_0 + \beta_1 L_{ij} + b)^{a_j} \{1 - \mathcal{L}^{-1}(\beta_0 + \beta_1 L_{ij} + b)\}^{(1-a_j)} \, d\Phi(b; \sigma).
\end{aligned}
\tag{2}
$$

This conditional probability describes the relationship between the N_i individual observed treatments and covariates for individuals within the cluster. Unlike in the case where no interference is assumed, the discrete treatment space includes 2^{N_i} possible unique treatment vectors for a cluster, and so (2) is a type of generalized propensity score, which we refer to as a cluster propensity score.

4.2. *Model for treatment under counterfactual policy α.* In addition, assume under counterfactual policy α that

$$\mathcal{L}\big(\text{Pr}_\alpha(A_{ij} = 1 | L_i, N_i, u_i)\big) = \gamma_{0\alpha} + \gamma_{1\alpha} L_{ij} + u_i,$$

where u_i is mean zero with distribution Φ and parameter(s) ϕ_α. Analogous to (1), it follows that the counterfactual cluster propensity score equals

$$
\begin{aligned}
&\text{Pr}_\alpha(A_i = a | L_i, N_i) \\
&= \int \prod_{j=1}^{N_i} \mathcal{L}^{-1}(\gamma_{0\alpha} + \gamma_{1\alpha} L_{ij} + u)^{a_j} \{1 - \mathcal{L}^{-1}(\gamma_{0\alpha} + \gamma_{1\alpha} L_{ij} + u)\}^{(1-a_j)} \, d\Phi(u; \phi_\alpha).
\end{aligned}
$$

The parameters $(\beta_0, \beta_1, \sigma)$ in (2) are identifiable from the observable random variables. However, the parameters $(\gamma_{0\alpha}, \gamma_{1\alpha}, \phi_\alpha)$, counterfactual cluster propensity scores $\text{Pr}_\alpha(A_i = a | L_i, N_i)$ and counterfactual probabilities $\omega(a, n, \alpha)$ are not identifiable without additional assumptions. In some settings it may be reasonable to assume that particular features of the distribution of covariates and treatment would remain the same under counterfactual policy α. For example, we assume $\text{Pr}(L_i) = \text{Pr}_\alpha(L_i)$, that is, the covariate distribution is unaffected by the policy.

Similarly, assume $\beta_1 = \gamma_{1\alpha}$, that is, the conditional odds ratio of treatment for any two individuals within a cluster is the same across the factual and counterfactual scenarios. This assumption implies that the ranking of individuals within clusters by probability of treatment is preserved across policies. For example, if, according to the observed treatment model, older

women are more likely to receive cholera vaccination; then, under policy α, older women will also be more likely to receive the cholera vaccine.

Finally, suppose $\sigma = \phi_\alpha$. Under this assumption certain aspects of the within-cluster correlation remain the same under policy α. For example, suppose $\mathcal{L}$ is the logit link and Φ is a Gaussian distribution function. Then, model (1) is implied by the latent variable model $A_{ij}^* = \beta_0 + \beta_1 L_{ij} + b_i + \epsilon_{ij}$, where ϵ_{ij} is assumed to have a logistic distribution with mean zero and variance $\pi^2/3$ and $A_{ij} = I(A_{ij}^* > 0)$. The latent variable A_{ij}^* can be interpreted as an unobserved continuous variable reflecting an individual's propensity to receive treatment. Under this model $\text{Corr}(A_{ij}^*, A_{ik}^* | L_{ij}, L_{ik}) = (\sigma^2 + \pi^2/3)^{-1}\sigma^2$ (Fitzmaurice, Laird and Ware (2011, page 417)). Likewise, the counterfactual treatment model above is implied by the latent variable model $A_{ij}^* = \gamma_{0\alpha} + \gamma_{1\alpha} L_{ij} + u_i + \epsilon_{\alpha ij}$, where $\epsilon_{\alpha ij}$ is also assumed to have a logistic distribution with mean zero and variance $\pi^2/3$, so that $\text{Corr}_\alpha(A_{ij}^*, A_{ik}^* | L_{ij}, L_{ik}) = (\phi_\alpha^2 + \pi^2/3)^{-1}\phi_\alpha^2$. That is, within-cluster correlation between the latent A_{ij}^*'s is the same in the factual and counterfactual scenarios.

Under the above assumptions,

$$
(3) \qquad \alpha = \int \left\{ N_i^{-1} \sum_{j=1}^{N_i} \int \mathcal{L}^{-1}(\gamma_{0\alpha} + \beta_1 L_{ij} + u)\, d\Phi(u; \sigma) \right\} dF_L,
$$

so the counterfactual model's intercept parameter $\gamma_{0\alpha}$ and, thus, the counterfactual cluster propensity scores are identifiable. It follows that the counterfactual probabilities $\omega(a, n, \alpha)$ are also identifiable from the observable data.

5. Inference. Following Tchetgen Tchetgen and VanderWeele (2012) and Perez-Heydrich et al. (2014), consider the following inverse probability-weighted (IPW) estimator of $\mu(\alpha)$:

$$
(4) \qquad \hat{\mu}(\alpha) = M^{-1} \sum_{i=1}^{M} \frac{\bar{Y}_i \omega(A_i, N_i, \alpha)}{\text{Pr}(A_i | L_i, N_i)},
$$

where $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} Y_{ij}$. The inverse probability-weight for cluster i is the reciprocal of the cluster propensity score; these and the counterfactual probabilities are unknown in an observational study and must be estimated from data.

Under the assumptions in Section 4, a logistic mixed effects model is fit to the data, and the model parameters $(\beta_0, \beta_1, \sigma)$ can be estimated by maximum likelihood. Then, the fitted parameters $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})$ are substituted into (2) to obtain an estimate of each cluster's propensity score. For each policy α, $\hat{\gamma}_{0\alpha}$ solves equation (3), with F_L replaced by its empirical distribution; that is, $\alpha = M^{-1} \sum_{i=1}^{M} N_i^{-1} \sum_{j=1}^{N_i} \int \mathcal{L}^{-1}(\gamma_{0\alpha} + \hat{\beta}_1 L_{ij} + u)\, d\Phi(u; \hat{\sigma})$ is solved to obtain $\hat{\gamma}_{0\alpha}$. The counterfactual cluster propensity scores for cluster i and treatments $a \in \mathcal{A}(N_i)$ are estimated by substitution, i.e.,

$$
\hat{\text{Pr}}_\alpha(A_i = a | L_i, N_i)
$$

$$
= \int \prod_{j=1}^{N_i} \mathcal{L}^{-1}(\hat{\gamma}_{0\alpha} + \hat{\beta}_1 L_{ij} + u)^{a_j} \{1 - \mathcal{L}^{-1}(\hat{\gamma}_{0\alpha} + \hat{\beta}_1 L_{ij} + u)\}^{(1-a_j)}\, d\Phi(u; \hat{\sigma}).
$$

Since the ordering of individuals in clusters is assumed to be uninformative, $\omega(a, n, \alpha) = \omega(a', n, \alpha)$ whenever $f(a) = f(a')$ for any two $a, a' \in \mathcal{A}(n)$ where $f(a) = \sum_{j=1}^{n} a_j$. For example, under this assumption $a = (1, 0)$, and $a = (0, 1)$ will occur with equal probability under policy α in clusters of size two. Note this (mild) assumption only pertains to the

treatment assignment mechanism under policy α, and it implies no restrictions on the interference structure within the cluster. Let $\mathcal{A}(n, s) = \{a \in \mathcal{A}(n) | f(a) = s\}$ where $|\mathcal{A}(n, s)| = \binom{n}{s}$, and define $\omega(s, n, \alpha) = \sum_{a \in \mathcal{A}(n,s)} \omega(a, n, \alpha)$ for $s = 0, 1, \ldots, n$. Estimate the counterfactual probabilities for any cluster i by $\widehat{\omega}(A_i, N_i, \alpha) = \binom{N_i}{f(A_i)}^{-1} \widehat{\omega}(f(A_i), N_i, \alpha)$, where for any triplet (s, n, α),

$$
\begin{aligned}
&\hat{\omega}(s, n, \alpha) \\
(5) \quad &= \left\{ \sum_{i=1}^{M} I(N_i = n) \right\}^{-1} \sum_{a \in \mathcal{A}(n,s)} \left\{ \sum_{i=1}^{M} \hat{\mathrm{Pr}}_\alpha(A_i = a | L_i, N_i) I(N_i = n) \right\}.
\end{aligned}
$$

These estimates, along with the estimated cluster propensity scores, are substituted into (4) to calculate $\hat{\mu}(\alpha)$. The estimators $\hat{\mathrm{OE}}(\alpha, \alpha') = \hat{\mu}(\alpha) - \hat{\mu}(\alpha')$ can be obtained in a similar manner. For $t = 0, 1$, the estimators $\hat{\mu}_t(\alpha)$ and $\hat{\mathrm{SE}}_t(\alpha, \alpha')$ are defined similarly using the outcomes $\bar{Y}_{t,i} = \{\sum_{j=1}^{N_i} I(A_{ij} = t)\}^{-1} \sum_{j=1}^{N_i} Y_{ij} I(A_{ij} = t)$, where $\bar{Y}_{t,i} = 0$ in the case when $A_{ij} = 1 - t$ for all $j = 1, \ldots, N_i$.

In the Appendix these estimators are shown to be consistent and asymptotically normal using standard large-sample estimating equation theory (Stefanski and Boos (2002)). Wald-type confidence intervals (CIs) can be constructed using the empirical sandwich estimators of the asymptotic variances.

The estimators described above may be computationally challenging in practice, as the estimator $\widehat{\omega}(a, n, \alpha)$ requires a numerical integration technique for each of the $\binom{n}{f(a)}$ vectors in $\mathcal{A}(n, f(a))$. Therefore, the following approximation is proposed to decrease computation time. The intuition behind this approximation is to replace the summation $\sum_{a \in \mathcal{A}(n,s)} \{\cdot\}$ in (5) with an estimate by randomly sampling the summands and then computing a weighted sum of the sampled summands. For each $s = 0, 1, \ldots, n$, define $\mathcal{A}(n, s, k)$ to be a subset of exactly $k_{s,n} = \min\{k, \binom{n}{s}\}$ vectors selected in a simple random sample from $\mathcal{A}(n, s)$, where $k > 1$ is chosen by the investigator. Now, estimate the counterfactual probabilities by $\widehat{\omega}(a, n, \alpha, k) = \binom{n}{f(a)}^{-1} \widehat{\omega}(f(a), n, \alpha, k)$, where for any triplet (s, n, α),

$$
\begin{aligned}
&\hat{\omega}(s, n, \alpha, k) \\
&= \left\{ \sum_{i=1}^{M} I(N_i = n) \right\}^{-1} k_{s,n}^{-1} \binom{n}{s} \sum_{a \in \mathcal{A}(n,s,k)} \sum_{i=1}^{M} \hat{\mathrm{Pr}}_\alpha(A_i = a | L_i, N_i) I(N_i = n).
\end{aligned}
$$

Replacing $\widehat{\omega}(a, n, \alpha)$ in $\hat{\mu}(\alpha)$ with $\widehat{\omega}(a, n, \alpha, k)$ results in an estimator which we denote $\hat{\mu}(\alpha, k)$. With analogous replacements define $\hat{\mathrm{OE}}(\alpha, \alpha', k)$ as well as $\hat{\mu}_t(\alpha, k)$ and $\hat{\mathrm{SE}}_t(\alpha, \alpha', k)$ for $t = 0, 1$. These estimators are evaluated in a simulation study in Section 6 and are employed in the data analysis of the cholera vaccine study in Section 7. All of the above estimators are implemented in the R package clusteredinterference (Barkley (2019)) which is freely provided on CRAN with an example vignette. In practice, specification of the value of k may be a compromise between better approximation (larger k) and faster computation (smaller k). This method may be extended by specifying different values of k to estimate distinct counterfactual probabilities which is outlined in the Appendix. A short discussion on estimating counterfactual probabilities under the assumption of uninformative ordering of individuals within clusters is additionally provided in the Supplementary Material Appendix S.1 (Barkley et al. (2020)).

Papadogeorgou, Mealli and Zigler (2019) consider IPW estimators similar in form to $\hat{\mu}(\alpha)$ and $\hat{\mu}_t(\alpha)$ which are also based on mixed effect propensity score models as in Sections 4.1–4.2. Their IPW estimators of the average potential outcome, when an individual is untreated

under policy α, differ in that the numerator includes an estimator of $\Pr_\alpha(A_i = a | A_{ij} = 0, L_i, N_i)$, whereas the numerator of $\hat{\mu}_t(\alpha)$ includes an estimator of $\Pr_\alpha(A_i = a | L_i, N_i)$ (namely, $\hat{\omega}(A_i, N_i, \alpha)$). This dissimiliarty is due to the differences in the target estimands, as discussed in Section 3.3. Papadogeorgou et al. also use large-sample estimating equation theory to establish consistency and asymptotic normality of their estimators.

6. Simulations. A simulation study was carried out on 1000 datasets to demonstrate the finite-sample performance of the proposed estimators. To generate each dataset, the following steps were carried out for each of $i = 1, \ldots, M = 125$ clusters:

I The number of individuals in the cluster N_i was simulated such that $\Pr(N_i = 8) = 0.4$, $\Pr(N_i = 22) = 0.35$, and $\Pr(N_i = 40) = 0.25$.

II Covariates for each individual $j = 1, \ldots, N_i$ in cluster i were simulated to be $L_{ij1} \sim N(40, 5)$ and $L_{ij2} \sim N(X_i, 0.2)$, where $X_i \sim N(6, 1)$ was a cluster-level random variable.

III Treatment status A_{ij} for each individual j in cluster i was simulated from a Bernoulli distribution with mean $\Pr(A_{ij} = 1 | L_{ij}, b_i) = \mathcal{L}^{-1}(\beta_0 + \beta_1 L_{ij1} + \beta_2 L_{ij2} + b_i)$ where $b_i \sim N(0, \sigma)$ was a cluster-level random intercept and $(\beta_0, \beta_1, \beta_2, \sigma) = (0.75, -0.015, -0.025, 0.75)$.

IV The outcome Y_{ij} for each individual j in cluster i was simulated from a Bernoulli distribution with mean $\Pr(Y_{ij} = 1 | A_i, L_{ij}) = \mathcal{L}^{-1}(0.1 - 0.05 L_{ij1} + 0.5 L_{ij2} - 0.5 A_{ij} + 0.2 g(A_{i,-j}) - 0.25 A_{ij} g(A_{i,-j}))$, where the function $g(A_{i,-j}) = (N_i - 1)^{-1} \sum_{j' \neq j} A_{ij'}$.

A logistic mixed effects model was fit with a random intercept for cluster and main effects for L_1 and L_2, that is, the propensity score models were correctly specified. Such a model could be fit with `lme4::glmer()` (Bates et al. (2015)) using arguments `formula = A ~ L1 + L2 + (1|Cluster)` and `family = "binomial"`. To determine the performance of the estimators that use the greatest degree of subsampling approximation, $k = 1$ was chosen. The asymptotic variance of the estimators was estimated with the empirical sandwich variance estimator as described in the Appendix, from which Wald-type 95% CIs were constructed.

True values of the estimands for policies $\alpha \in \{0.4, 0.5, 0.55\}$ were determined empirically using the same data generating process outlined above in steps I–II and analogues to steps III–IV. The process is described here briefly, with more details provided in the Supplementary Material Appendix S.2 (Barkley et al. (2020)). For each α, $\gamma_{0\alpha}$ was determined by solving (3) with F_L approximated by its empirical distribution over 10^7 clusters. Then, the counterfactual probabilities $\omega(a, n, \alpha)$ were determined by generating treatment vectors under policy α for 10^8 clusters, replacing β_0 in step III with $\gamma_{0\alpha}$. An empirical comparison of true values of $\omega(a, n, \alpha)$ arising from this simulation study and the true values of $\omega_B(a, n, \alpha)$ for the type B policies is provided in Figure S.1 in the Supplementary Material Appendix S.2 (Barkley et al. (2020)). Next, potential outcomes were generated for 10^8 clusters via the causal model analogous to the regression model specified in step IV. These potential outcomes were combined with the counterfactual probabilities to determine the true values of $\mu(\alpha)$, $\text{OE}(\alpha, \alpha')$, and $\mu_t(\alpha)$ and $\text{SE}_t(\alpha, \alpha')$ for $t = 0, 1$.

The IPW estimates from each dataset were compared to the true estimand values determined above; a summary of these results is presented in Table 1. The average bias of the estimators was negligible. The average of the estimated asymptotic standard errors was approximately equal to the empirical Monte Carlo standard error. The Wald-type 95% CIs contained the true parameter values for approximately 95% of the simulated datasets. Thus, the estimators performed well in this simulation study.

Summary of results from simulation study described in Section 6. "Truth" denotes the true value of the estimand targeted by the estimator; "Bias" denotes the average bias of the IPW estimates over the 1000 datasets; "Cov%" denotes the empirical coverage of Wald-type 95% CIs; "ASE" denotes the average of the estimated sandwich standard errors times 100; "ESE" denotes the empirical standard error times 100; "SER" denotes the ratio of ASE divided by ESE; $\alpha_1 = 0.4$, $\alpha_2 = 0.5$ and $\alpha_3 = 0.55$

Estimator	Truth	Bias	Cov%	ASE	ESE	SER
$\hat{\mu}(\alpha_1, k=1)$	0.662	−0.003	94.3%	1.88	1.84	1.02
$\hat{\mu}(\alpha_2, k=1)$	0.651	0.000	95.5%	1.63	1.53	1.06
$\hat{\mu}(\alpha_3, k=1)$	0.645	0.001	96.4%	1.65	1.55	1.07
$\hat{OE}(\alpha_2, \alpha_1, k=1)$	−0.011	0.003	97.2%	1.08	0.96	1.13
$\hat{OE}(\alpha_3, \alpha_1, k=1)$	−0.017	0.004	97.4%	1.44	1.34	1.08
$\hat{OE}(\alpha_3, \alpha_2, k=1)$	−0.006	0.001	97.4%	0.53	0.44	1.21
$\hat{\mu}_0(\alpha_1, k=1)$	0.712	−0.002	95.2%	2.10	2.02	1.04
$\hat{\mu}_0(\alpha_2, k=1)$	0.711	−0.001	95.7%	2.15	2.02	1.07
$\hat{\mu}_0(\alpha_3, k=1)$	0.709	−0.001	95.3%	2.46	2.35	1.05
$\hat{SE}_0(\alpha_2, \alpha_1, k=1)$	−0.001	0.001	95.8%	1.33	1.20	1.11
$\hat{SE}_0(\alpha_3, \alpha_1, k=1)$	−0.003	0.001	94.7%	1.93	1.86	1.04
$\hat{SE}_0(\alpha_3, \alpha_2, k=1)$	−0.002	0.000	94.8%	0.79	0.72	1.10
$\hat{\mu}_1(\alpha_1, k=1)$	0.573	0.007	94.2%	3.04	3.09	0.99
$\hat{\mu}_1(\alpha_2, k=1)$	0.581	0.004	95.0%	2.25	2.24	1.01
$\hat{\mu}_1(\alpha_3, k=1)$	0.582	0.001	95.3%	2.10	2.07	1.01
$\hat{SE}_1(\alpha_2, \alpha_1, k=1)$	0.008	0.003	94.9%	1.51	1.46	1.04
$\hat{SE}_1(\alpha_3, \alpha_1, k=1)$	0.009	0.005	95.2%	2.02	1.98	1.02
$\hat{SE}_1(\alpha_3, \alpha_2, k=1)$	0.002	0.002	96.4%	0.65	0.57	1.13

7. Analysis of cholera vaccine study in Matlab, Bangladesh. The Matlab cholera vaccine study was analyzed using the proposed methods. The identifiability assumptions discussed in Section 4 were assumed. The IPW estimators were computed with $k = 3$, and Wald-type CIs were constructed from the empirical sandwich variance estimator. Following VanderWeele and Shpitser (2011) and Perez-Heydrich et al. (2014), pretreatment variables were considered for inclusion in the treatment model if they possibly caused cholera infection, study participation or both. These variables included individual age (centered, in decades) and distance from bari to the nearest river (in kilometers). Average age within the bari (centered, in decades) was also considered, thus potentially allowing the treatment of individuals to depend on baseline covariates of other individuals in the same bari. Candidate models were compared by evaluating predictive accuracy using leave-cluster-out cross-validation (Chen, Zeng and Wang (2015)). Each candidate logistic mixed effects model was fit using some combination of linear, quadratic or interaction terms for the pretreatment variables. The model selected for this analysis performed as well as or better than other candidates and included a linear term for distance, linear and quadratic terms for age, and an interaction term between linear distance and linear age; see Supplementary Material Appendix S.3 (Barkley et al. (2020)). The Tchetgen and Coull (2006) test yielded $p = 0.051$, suggesting the random effect distribution might be non-Gaussian. However, removal of five small baris (each with 15 individuals or less and no treated individuals) yielded $p = 0.15$, indicating adequate fit when excluding these few baris. Similarly, removing 10 of the baris (out of the 6415 total) yielded $p = 0.40$. Inferences about the overall and spillover effects of the cholera vaccine were nearly identical when these baris were excluded; see Figures S.4, S.5 and S.6 in the Supplementary Material Appendix (Barkley et al. (2020)). The variance component of the random intercept was estimated to be $\hat{\sigma} = 0.93$ with 95% CI (0.90, 0.95) calculated
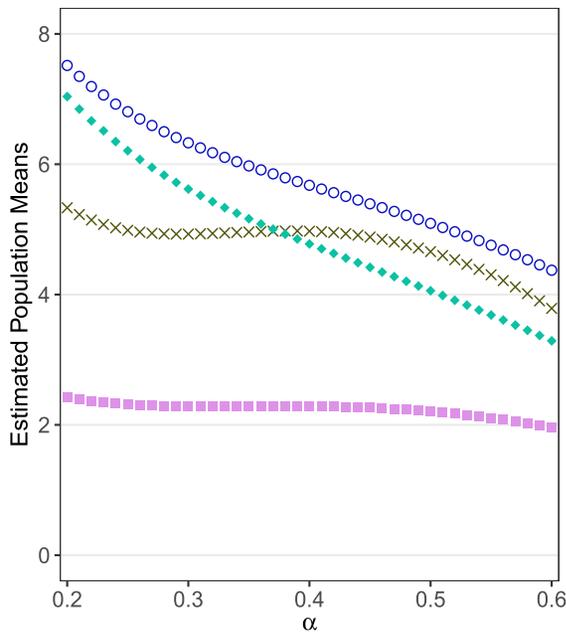
FIG. 2.   *Estimates of the population mean estimands from the analysis of the Matlab cholera vaccine study. The light green diamonds indicate* $\hat{\mu}(\alpha, k = 3)$. *The dark blue circles indicate* $\hat{\mu}_0(\alpha, k = 3)$, *and the light pink squares indicate* $\hat{\mu}_1(\alpha, k = 3)$. *The dark brown* ×*'s indicate* $\hat{\mu}_B(\alpha)$ *which target the type B estimands from Tchetgen Tchetgen and VanderWeele* (2012). *All estimates are multiplied by* 1000.

via profile likelihood using `lme4::confint.merMod()` (Bates et al. (2015)), indicating correlation between individual treatment statuses within clusters.

Figure 2 depicts point estimates of the population mean estimands over policies ranging from $\alpha = 0.2$ to $\alpha = 0.6$. Estimates are presented in units of one case of cholera infection per 1000 individuals per year. Estimates of $\mu_1(\alpha)$ were relatively invariant to α, suggesting minimal spillover effects when an individual is vaccinated. In contrast, estimates of $\mu_0(\alpha)$ decreased noticeably as α increased, suggesting a protective spillover effect when an individual is not vaccinated. The estimates of $\mu(\alpha)$ similarly suggest lower risk of cholera infection at the population level for policies with greater levels of vaccine coverage.

Overall effect estimates and corresponding 95% CIs are depicted in Figure 3. Negative effects are favorable, corresponding to a reduction in cholera infections. For example, $\hat{\text{OE}}(0.45, 0.3, k = 3) = -1.2$ (95% CI $-1.8, -0.7$), indicating a protective effect of policy $\alpha = 0.45$ compared to $\alpha = 0.3$. In particular, we expect 1.2 fewer cases of cholera per 1000 person-years if there is 45% vaccine coverage compared to 30% vaccine coverage.

Estimated spillover effects are depicted in Figure 4. The estimates of $\hat{\text{SE}}_1(\alpha, \alpha', k = 3)$ were approximately zero; the CIs included zero for almost all contrasts shown, indicating mostly negligible spillover effect, if any, among treated individuals within clusters. However, $\hat{\text{SE}}_0(\alpha, \alpha', k = 3)$ was negative for $\alpha > \alpha'$ and positive for $\alpha < \alpha'$, and all of the CIs excluded zero. Thus, there is evidence of a protective effect of policies with higher probability of treatment exposure conferred to individuals who did not themselves obtain treatment.

Figures 2 and 3 also depict point estimates of the type B estimands and corresponding 95% CIs, computed using the R package `inferference` (Saul and Hudgens (2017)) based on the same logistic mixed effects propensity score model employed with the proposed estimators. Relative to the estimates of the proposed estimands, the estimates of the type B estimands were smaller with corresponding 95% CIs that often included zero. For example, $\hat{\text{OE}}_B(0.2, 0.5) = 0.7$ (95% CI $-0.3, 1.7$), while $\hat{\text{OE}}(0.2, 0.5, k = 3) = 3.0$ (95% CI 1.6, 4.3).
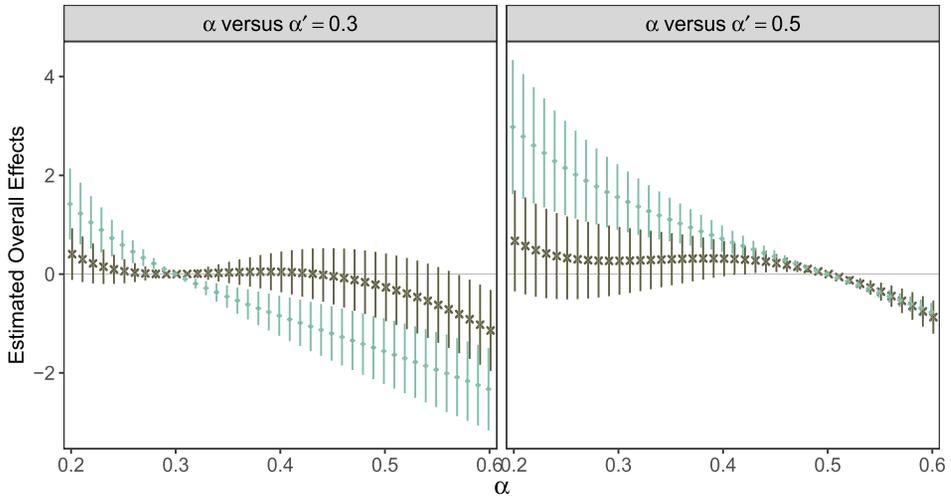
FIG. 3. *Estimated overall effects from the analysis of the Matlab cholera vaccine study for selected contrasts. The diamonds and light green lines indicate the point estimates and 95% CIs from $\hat{\text{OE}}(\alpha, \alpha', k = 3)$. The ×'s and dark brown lines indicate the point estimates and 95% CIs from $\hat{\text{OE}}_B(\alpha, \alpha')$ which target the type B estimands from Tchetgen Tchetgen and VanderWeele (2012). All estimates are multiplied by 1000.*

Thus, inferences based on the type B estimands tended to underestimate the population-level utility of cholera vaccination compared to results based on the proposed estimands.

8. Discussion. Drawing causal inference from observational data when interference may be present poses several challenges, including defining the causal effects of interest. Proposed in this paper are causal estimands for use in observational studies when clustered interference is plausible. The proposed causal effects are contrasts in mean potential outcomes arising from different policies that change the distribution of treatment. IPW estimators were proposed and shown to be consistent and asymptotically normal under certain identifying assumptions, and empirical sandwich estimators were derived for the asymptotic variance of
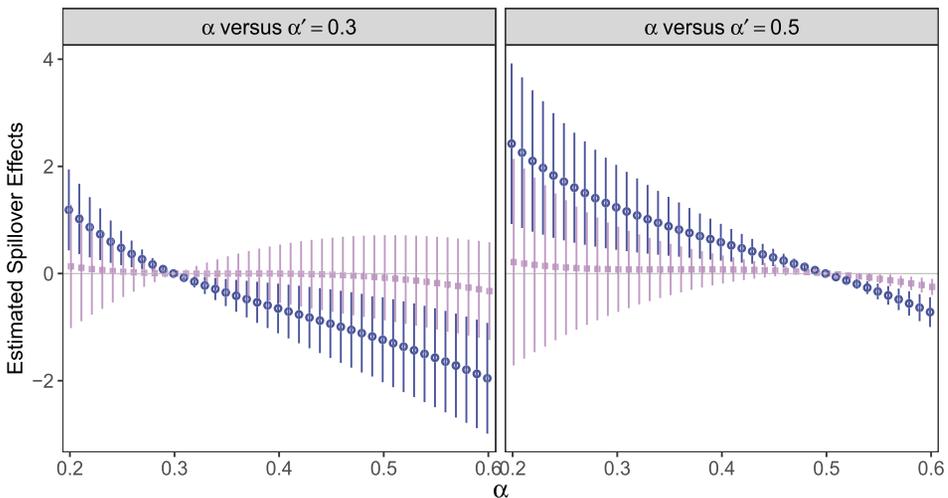

FIG. 4. *Estimated spillover effects from the analysis of the Matlab cholera vaccine study for selected contrasts. The circles and dark blue lines indicate the point estimates and 95% CIs from $\hat{\text{SE}}_0(\alpha, \alpha', k = 3)$. The squares and light pink lines indicate the point estimates and 95% CIs from $\hat{\text{SE}}_1(\alpha, \alpha', k = 3)$. All estimates are multiplied by 1000.*

the estimators. The IPW estimators performed well in finite samples with minimal bias, and the Wald-type confidence intervals attained nominal coverage levels. These methods were illustrated in an analysis of a large cholera vaccine study, providing evidence that increasing the proportion of individuals vaccinated reduces cholera infections.

The policy effects considered here may be more relevant in public health settings such as infectious disease research because within-cluster characteristics are incorporated into the proposed estimands. To reduce the burden of infectious diseases through vaccination programs, it is important to consider the "ecological circumstances" of the disease (Ali et al. (2009)). Previously proposed type B estimands define treatment effects in the counterfactual scenario where individuals are independently exposed to treatment. However, scenarios in which treatment exposures are correlated may represent more relevant ecological circumstances. Aside from controlled trials, in general one might expect treatment correlation in settings where interference is present. Indeed, the cholera vaccine study analysis in Section 7 indicates strong evidence of treatment correlation within clusters. Unlike the type B estimands, the proposed causal estimands instead describe counterfactual scenarios which allow for within-cluster dependence in the individual treatment selections. Likewise, the proposed estimands preserve the conditional odds ratio of treatment for any two individuals within the same cluster. By incorporating these within-cluster features, inferences targeting the proposed estimands may be of greater relevance to public health investigators and policy-makers concerned with controlling the spread of infectious disease in a particular population.

An appealing aspect of IPW estimators is that no outcome modeling is required and no assumptions are stipulated regarding the interference structure within clusters. On the other hand, consistency of IPW estimators is, in general, dependent on the correct specification of the treatment model. The estimators presented here also require that the model for the counterfactual distribution of treatment is correctly specified. Therefore, in applications it will be important to evaluate model fit. For the cholera vaccine analysis presented in Section 7, model fit was assessed via predictive accuracy using leave-cluster-out cross-validation and the Tchetgen and Coull (2006) test of the Gaussian random effect assumption. Fortunately, inferences from generalized linear mixed effects models tend to be robust to misspecification of the random effect distributional assumption (McCulloch and Neuhaus (2011)).

An alternative approach, which could potentially be pursued in future research, would be to use nonparametric methods to improve robustness to misspecification of the treatment model. However, such an approach may impede identifiability of the target causal estimands without further untestable identifying assumptions. In addition, that the treatment vector A_i can take on a large number of discrete possible values may pose difficulty if a nonparametric approach is considered. For example, if $N_i = 20$, then there are greater than 10^6 possible realizations of A_i. Thus, applying existing multi-class classifiers in this setting may be challenging.

Another limitation of the proposed methods is the computational difficulty due to a large number of nuisance parameters, depending on the joint distribution of (A_i, N_i). Future work may consider reducing the number of nuisance parameters, perhaps through approximating the counterfactual treatment distribution. Future research may also consider assuming different structures of interference to better align with the epidemiology of cholera; for example, see Ali et al. (2018).

Typical of methods for drawing causal inference from observational data, the approach here relies on exchangeability and positivity assumptions. Within the context of the cholera vaccine study analysis, exchangeability assumes the cholera potential outcomes to be independent of vaccination status conditional on the baseline covariates age and distance from the bari to the nearest river. Because participants in the experimental component of the study were randomized to vaccine or not, the exchangeability assumption should hold if the potential outcomes are conditionally independent of individuals' decisions to participate in the

randomized trial given age and distance to the nearest river. These covariates were selected on the basis of earlier analyses of these data by Perez-Heydrich et al. (2014) suggesting age and distance may both be common causes of study participation and cholera infection.

The positivity assumption might be considered reasonable in the cholera vaccine study, as all women and children in Matlab were invited to participate in the vaccine trial. However, if certain subgroups of women or children would under no circumstances be willing to receive the cholera vaccine, then the positivity assumption would be violated. The high number of baris with no individuals vaccinated displayed in the right panel of Figure 1 is suggestive of this possibility, although most of these baris tended to be small with no or few individuals participating in the trial. In the absence of interference, possible positivity violations can give rise to extreme weights. The IPW estimators considered here can be viewed as weighted averages of the observed outcomes with weights $\hat{\omega}(A_i, N_i, \alpha)/\hat{\Pr}(A_i|L_i, N_i)$. For the cholera vaccine study analysis, extreme weights were not observed. For example, for $\alpha = 0.5$, the median and maximum weights were 0.9 and 4.6, respectively; by comparison, estimates of the type B estimands had median and maximum weights 1.1 and 8.5. Similar results were obtained for $\alpha = 0.4$ and $\alpha = 0.6$. See Figure S.3 in the Supplementary Material Appendix (Barkley et al. (2020)).

While this work is motivated by infectious disease research, it is applicable in many other areas in which interference may be present. By defining causal effects of population-level interventions (Westreich (2017)) in the presence of interference, the proposed estimands may have greater practical utility and be more relevant to investigators and policy-makers.

APPENDIX

The IPW estimators introduced in Section 5 of the main paper are shown to be consistent and asymptotically normal using standard large-sample estimating equation theory or "M-estimation" (Stefanski and Boos (2002)). Presented for illustration below is a simple example where each cluster has exactly n individuals, and at least one cluster $i \leq M$ is observed to experience treatment $f(A_i) = s$ for each $s = 0, 1, \ldots, n$. Let $\omega_\alpha = (\omega(0, n, \alpha), \ldots, \omega(n-1, n, \alpha))$ be the ordered vector of the possibly unique counterfactual probabilities excepting $\omega(n, n, \alpha)$; the law of total probability implies that $\omega(n, n, \alpha) = 1 - \sum_{s=0}^{n-1} \omega(s, n, \alpha)$. Let $\theta_\alpha = (\beta_0, \beta_1, \sigma, \gamma_{0\alpha}, \omega_\alpha, \mu(\alpha))$ be the ordered vector of all parameters to estimate. Next, estimating functions corresponding to each element of θ_α are introduced.

Estimating functions for the parameters $\nu = (\beta_0, \beta_1, \sigma)$ in the mixed treatment model are the score functions of the log likelihood. Let $\psi_\nu = (\psi_{\beta_0}, \psi_{\beta_1}, \psi_\sigma)^\mathsf{T}$ be a column vector denoting these estimating functions. For β_1, the estimating function is

$$\psi_{\beta_1}(O_i; \theta_\alpha) = \frac{\partial}{\partial \beta_1} \log\{\Pr(A_i|L_i, N_i)\},$$

where $\Pr(A_i|L_i, N_i)$ is given in (2) of the main text. For $\gamma_{0\alpha}$, define the estimating function

$$\psi_{\gamma_{0\alpha}}(O_i; \theta_\alpha) = \left\{ N_i^{-1} \sum_{j=1}^{N_i} \int \mathcal{L}^{-1}(\gamma_{0\alpha} + \beta_1 L_{ij} + b_i)\, d\Phi(b_i; \sigma) \right\} - \alpha.$$

For each $\omega(s, n, \alpha) \in \omega_\alpha$, define the estimating function

$$\psi_{\omega(s,n,\alpha)}(O_i; \theta_\alpha) = \left\{ \sum_{a \in \mathcal{A}(n,s)} \Pr_\alpha(A_i = a|L_i, N_i) - \omega(s, n, \alpha) \right\} I(N_i = n),$$

and let $\psi_{\omega_\alpha} = (\psi_{\omega(0,n,\alpha)}, \psi_{\omega(1,n,\alpha)}, \ldots, \psi_{\omega(n-1,n,\alpha)})^\mathsf{T}$. For the target estimand, define

$$\psi_{\mu(\alpha)}(O_i; \theta_\alpha) = \frac{\bar{Y}_i \omega(A_i, N_i, \alpha)}{\Pr(A_i|L_i, N_i)} - \mu(\alpha),$$

where $\omega(A_i, N_i, \alpha) = \binom{N_i}{f(A_i)}^{-1} \omega(f(A_i), N_i, \alpha)$ and where $\Pr(A_i|L_i, N_i)$ is the propensity score for the cluster as in (2) of the main text.

Let $\psi_{\theta_\alpha} = (\psi_\nu, \psi_{\gamma_{0\alpha}}, \psi_{\omega_\alpha}, \psi_{\mu(\alpha)})^\mathsf{T}$, and let $q = |\theta_\alpha|$ be the number of parameters to estimate. The estimator $\hat{\theta}_\alpha$ can be expressed as a solution to the following system of estimating equations:

$$\sum_{i=1}^{M} \psi_{\theta_\alpha}(O_i; \theta_\alpha) = \sum_{i=1}^{M} \begin{bmatrix} \psi_\nu(O_i; \theta_\alpha) \\ \psi_{\gamma_{0\alpha}}(O_i; \theta_\alpha) \\ \psi_{\omega_\alpha}(O_i; \theta_\alpha) \\ \psi_{\mu(\alpha)}(O_i; \theta_\alpha) \end{bmatrix} = 0_{q \times 1}.$$

To show that $\mu(\alpha)$ is the solution to $\int \psi_{\mu(\alpha)}(O|\theta_\alpha) \, dF_O(O) = 0$, note

$$\int \psi_{\mu(\alpha)}(O|\theta_\alpha) \, dF_O(O) = \mathbb{E}\left\{ \sum_{a \in \mathcal{A}(N)} \frac{\bar{Y}(a)\omega(a, N, \alpha)}{\Pr(A = a|L, N)} I(A = a) \right\} - \mu(\alpha)$$

$$= \mathbb{E}_{L,N}\left[\sum_{a \in \mathcal{A}(N)} \left\{ \mathbb{E}_{A,\{Y(a)\}|L,N}\left(\bar{Y}(a)\omega(a, N, \alpha)\right) \right. \right.$$

$$\left. \left. \times \mathbb{E}_{A,\{Y(a)\}|L,N}\left(\frac{I(A = a)}{\Pr(A = a|L, N)}\right) \right\} \right] - \mu(\alpha)$$

$$= \mathbb{E}\left\{ \sum_{a \in \mathcal{A}(N)} \bar{Y}(a)\omega(a, N, \alpha) \right\} - \mu(\alpha);$$

the first term in the last statement equals $\mu(\alpha)$ by definition, and so $\mu(\alpha)$ solves the equation $\int \psi_{\mu(\alpha)}(O; \theta_\alpha) \, dF_O(O) = 0$. Since ψ_ν are simply the score functions, $\int \psi_\nu(O; \theta_\alpha) \, dF_O(O) = 0_{|\nu| \times 1}$. Note that the right side of (3) in the main text equals $\alpha + \int \psi_{\gamma_{0\alpha}}(O; \theta_\alpha) \, dF_O(O)$, so $\gamma_{0\alpha}$ solves $\int \psi_{\gamma_{0\alpha}}(O; \theta_\alpha) \, dF_O(O) = 0$. Finally, $\int \psi_{\omega(s,n,\alpha)}(O; \theta_\alpha) \, dF_O(O) = 0$ follows from $\omega(a, n, \alpha) = \mathbb{E}_L\{\Pr_\alpha(A = a|L, N = n)\}$. Combining these results shows that $\int \psi_{\theta_\alpha}(O; \theta_\alpha) \, dF_O(O) = 0_{q \times 1}$.

From Stefanski and Boos (2002), $\hat{\theta}_\alpha \xrightarrow{p} \theta_\alpha$ and $\sqrt{M}(\hat{\theta}_\alpha - \theta_\alpha) \xrightarrow{d} N(0, \Sigma_\alpha)$, where $\Sigma_\alpha = U_\alpha^{-1} W_\alpha (U_\alpha^{-1})^\mathsf{T}$ for $U_\alpha = \mathbb{E}\{-\dot{\psi}_{\theta_\alpha}(O; \theta_\alpha)\}$ and $W_\alpha = \mathbb{E}\{\psi_{\theta_\alpha}(O; \theta_\alpha)^{\otimes 2}\}$. Consistent estimators for U_α and W_α are $\hat{U}_\alpha = M^{-1} \sum_{i=1}^{M}\{-\dot{\psi}_{\theta_\alpha}(O_i; \theta_\alpha)|_{\theta_\alpha = \hat{\theta}_\alpha}\}$ and $\hat{W}_\alpha = M^{-1} \sum_{i=1}^{M}\{\psi_{\theta_\alpha}(O_i; \hat{\theta}_\alpha)^{\otimes 2}\}$. The empirical sandwich variance estimator $\hat{\Sigma}_\alpha = \hat{U}_\alpha^{-1} \times \hat{W}_\alpha (\hat{U}_\alpha^{-1})^\mathsf{T}$ is consistent for Σ_α, and so $\hat{\text{Var}}(\hat{\mu}(\alpha)) = M^{-1}[\hat{\Sigma}_\alpha]_{[q,q]}$ approximates the variance of $\hat{\mu}(\alpha)$ for large M, where $[\hat{\Sigma}_\alpha]_{[q,q]}$ is the bottom-right element of $\hat{\Sigma}_\alpha$.

An analogous approach is described for $\hat{\text{OE}}(\alpha, \alpha', k)$, where it is now necessary to estimate $\gamma_{0\alpha'}$ and $\omega_{\alpha'}$ as well. Let $\theta_{\alpha,\alpha'} = (\nu, \gamma_{0\alpha}, \gamma_{0\alpha'}, \omega_\alpha, \omega_{\alpha'}, \text{OE}(\alpha, \alpha'))$ be the ordered vector of all parameters to estimate. For each $\omega(s, n, \alpha) \in \omega_\alpha$, define the estimating function

$$\psi_{k,\omega(s,n,\alpha)}(O_i; \theta_{\alpha,\alpha'})$$

$$= \left\{ k_{s,n}^{-1} \binom{n}{s} \sum_{a \in \mathcal{A}(n,s,k)} \Pr_\alpha(A_i = a|L_i, N_i) - \omega(s, n, \alpha) \right\} I(N_i = n),$$

and let $\psi_{k,\omega_\alpha} = (\psi_{k,\omega(0,n,\alpha)}, \psi_{k,\omega(1,n,\alpha)}, \ldots, \psi_{k,\omega(n-1,n,\alpha)})^\mathsf{T}$. For the target estimand, define

$$\psi_{\text{OE}(\alpha,\alpha')}(O_i; \theta_{\alpha,\alpha'}) = \frac{\bar{Y}_i\{\omega(A_i, N_i, \alpha) - \omega(A_i, N_i, \alpha')\}}{\Pr(A_i|L_i, N_i)} - \text{OE}(\alpha, \alpha').$$

It is easily shown that $\int \psi_{\text{OE}(\alpha,\alpha')}(O; \theta_{\alpha,\alpha'}) \, dF_O(O) = 0$ using a proof analogous to the one for $\psi_{\mu(\alpha)}$ presented above. In a similar manner, that $\int \psi_{k,\omega(s,n,\alpha)}(O; \theta_{\alpha,\alpha'}) \, dF_O(O) = 0$ follows directly from $\int \psi_{\omega(s,n,\alpha)}(O; \theta_\alpha) \, dF_O(O) = 0$. Finally, let $\psi_{k,\theta_{\alpha,\alpha'}} = (\psi_\nu, \psi_{\gamma_{0\alpha}}, \psi_{\gamma_{0\alpha'}},$

$\psi_{k,\omega_\alpha}, \psi_{k,\omega_{\alpha'}}, \psi_{\mathrm{OE}(\alpha,\alpha')})^\mathsf{T}$. Then, $\theta_{\alpha,\alpha'}$ solves $\int \psi_{k,\theta_{\alpha,\alpha'}}(O;\theta_{\alpha,\alpha'})\,dF_O(O) = 0_{q'\times 1}$, $\hat{\theta}_{\alpha,\alpha'}$ solves $\sum_{i=1}^M \psi_{k,\theta_{\alpha,\alpha'}}(O_i;\theta_{\alpha,\alpha'}) = 0_{q'\times 1}$ for $q' = |\theta_{\alpha,\alpha'}|$ and the above results follow.

The difference in $\hat{\mathrm{OE}}(\alpha,\alpha')$ and $\hat{\mathrm{OE}}(\alpha,\alpha',k)$ arises solely from the estimating functions used for the counterfactual probabilities, that is, ψ_{ω_α} and ψ_{k,ω_α}, respectively. When $\binom{n}{\lfloor n/2 \rfloor} \leq k$, then $\mathcal{A}(n,s) = \mathcal{A}(n,s,k)$ for all s, and ψ_{ω_α} is equivalent to ψ_{k,ω_α}. As mentioned in the main text, one could use different values of k for distinct estimating equations. For example, one could estimate $\omega(s,n,\alpha)$ with $\psi_{k,\omega(s,n,\alpha)}$ and $\omega(s',n',\alpha)$ with $\psi_{k',\omega(s',n',\alpha)}$, where $\omega(s,n,\alpha) \neq \omega(s',n',\alpha)$ and $k \neq k'$, and the above results would still apply.

SUPPLEMENTARY MATERIAL

Supplement to "Causal inference from observational studies with clustered interference, with application to a cholera vaccine study" (DOI: 10.1214/19-AOAS1314SUPP; .pdf). This document contains: a short discussion on estimating counterfactual probabilities, detailed instructions for empirically determining the true values of the estimands in the simulation study presented in the main paper, model fit statistics for several models considered for the main data analysis, a comparison of the IPW weights for the proposed and the type B estimators, and comparisons of results for the full dataset vs. those after removing several baris.

REFERENCES

ALI, M., EMCH, M. E., VON SEIDLEIN, L., YUNUS, M., SACK, D. A., RAO, M., HOLMGREN, J. and CLEMENS, J. D. (2005). Herd immunity conferred by killed oral cholera vaccines in Bangladesh: A reanalysis. *Lancet* **366** 44–49.

ALI, M., EMCH, M. E., YUNUS, M. and CLEMENS, J. D. (2009). Modeling spatial heterogeneity of disease risk and evaluation of the impact of vaccination. *Vaccine* **27** 3724–3729.

ALI, M., KIM, D. R., KANUNGO, S., SUR, D., MANNA, B., DIGILIO, L., DUTTA, S., MARKS, F., BHAT-TACHARYA, S. K. et al. (2018). Use of oral cholera vaccine as a vaccine probe to define the geographical dimensions of person-to-person transmission of cholera. *Int. J. Infect. Dis.* **66** 90–95. https://doi.org/10.1016/j.ijid.2017.11.020

BARKLEY, B. G. (2019). clusteredinterference: Causal effects from observational stud- ies with clustered interference. R package version 1.0.1. Available at https://CRAN.R-project.org/package=clusteredinterference.

BARKLEY, B. G., HUDGENS, M. G., CLEMENS, J. D., ALI, M. and EMCH, M. E. (2020). Supplement to "Causal inference from observational studies with clustered interference, with application to a cholera vaccine study." https://doi.org/10.1214/19-AOAS1314SUPP

BATES, D., MÄCHLER, M., BOLKER, B. and WALKER, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67** 1–48. https://doi.org/10.18637/jss.v067.i01

CHEN, T., ZENG, D. and WANG, Y. (2015). Multiple kernel learning with random effects for predicting longitudinal outcomes and data integration. *Biometrics* **71** 918–928. MR3436717 https://doi.org/10.1111/biom.12343

COX, D. R. (1958). *Planning of Experiments. A Wiley Publication in Applied Statistics*. Wiley, New York. MR0095561

DÍAZ, I. D. and VAN DER LAAN, M. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics* **68** 541–549. MR2959621 https://doi.org/10.1111/j.1541-0420.2011.01685.x

ECK, D. J., MOROZOVA, O. and CRAWFORD, F. W. (2018). Randomization for the direct effect of an infectious disease intervention in a clustered study population. Preprint. Available at arXiv:1808.05593.

FITZMAURICE, G. M., LAIRD, N. M. and WARE, J. H. (2011). *Applied Longitudinal Analysis*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2830137

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

HONG, G. and RAUDENBUSH, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *J. Amer. Statist. Assoc.* **101** 901–910. MR2324091 https://doi.org/10.1198/016214506000000447

HUDGENS, M. G. and HALLORAN, M. E. (2008). Toward causal inference with interference. *J. Amer. Statist. Assoc.* **103** 832–842. MR2435472 https://doi.org/10.1198/016214508000000292

LIU, L., HUDGENS, M. G. and BECKER-DREPS, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika* **103** 829–842. MR3620442 https://doi.org/10.1093/biomet/asw047

MCCULLOCH, C. E. and NEUHAUS, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statist. Sci.* **26** 388–402. MR2917962 https://doi.org/10.1214/11-STS361

PAPADOGEORGOU, G., MEALLI, F. and ZIGLER, C. M. (2019). Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics* **75** 778–787. MR4012083 https://doi.org/10.1111/biom.13049

PEREZ-HEYDRICH, C., HUDGENS, M. G., HALLORAN, M. E., CLEMENS, J. D., ALI, M. and EMCH, M. E. (2014). Assessing effects of cholera vaccination in the presence of interference. *Biometrics* **70** 731–741. MR3261790 https://doi.org/10.1111/biom.12184

SAUL, B. C. and HUDGENS, M. G. (2017). A recipe for inferference: Start with causal inference. Add interference. Mix well with R. *J. Stat. Softw.* **82** 1–21. https://doi.org/10.18637/jss.v082.i02

SOBEL, M. E. (2006). What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *J. Amer. Statist. Assoc.* **101** 1398–1407. MR2307573 https://doi.org/10.1198/016214506000000636

STEFANSKI, L. A. and BOOS, D. D. (2002). The calculus of *M*-estimation. *Amer. Statist.* **56** 29–38. MR1939394 https://doi.org/10.1198/000313002753631330

TCHETGEN, E. J. and COULL, B. A. (2006). A diagnostic test for the mixing distribution in a generalised linear mixed model. *Biometrika* **93** 1003–1010. MR2285086 https://doi.org/10.1093/biomet/93.4.1003

TCHETGEN TCHETGEN, E. J. and VANDERWEELE, T. J. (2012). On causal inference in the presence of interference. *Stat. Methods Med. Res.* **21** 55–75. MR2867538 https://doi.org/10.1177/0962280210386779

VANDERWEELE, T. J. and SHPITSER, I. (2011). A new criterion for confounder selection. *Biometrics* **67** 1406–1413. MR2872391 https://doi.org/10.1111/j.1541-0420.2011.01619.x

VANDERWEELE, T. J. and TCHETGEN TCHETGEN, E. J. (2011). Effect partitioning under interference in two-stage randomized vaccine trials. *Statist. Probab. Lett.* **81** 861–869. MR2793754 https://doi.org/10.1016/j.spl.2011.02.019

WESTREICH, D. (2017). From patients to policy: Population intervention effects in epidemiology. *Epidemiology* **28** 525–528. https://doi.org/10.1097/EDE.0000000000000648