Estimating Direct and Spillover Vaccine Effectiveness with Partial Interference under Test-Negative Design Sampling

Cong Jiang[1,2]*, **Fei Fang**[3]*, **Denis Talbot**[4,5], **Mireille E Schnitzer**[2]

[1] *Department of Biostatistics, Harvard Chan School of Public Health , Massachusetts, USA*

[2] *Faculty of Pharmacy, Université de Montréal, Québec, Canada*

[3] *Yale School of Public Health, Yale University, Connecticut, USA*

[4] *Département de médecine sociale et préventive, Université Laval, Québec, Canada*

[5] *Axe santé des populations et pratiques optimales en santé, Québec, Canada*

{cong.jiang, mireille.schnitzer}@umontreal.ca

August 13, 2025

Abstract

The Test-Negative Design (TND), which involves recruiting care-seeking individuals who meet predefined clinical case criteria, offers valid statistical inference for Vaccine Effectiveness (VE) using data collected through passive surveillance, making it cost-efficient and timely. Infectious disease epidemiology often involves interference, where the treatment and/or outcome of one individual can affect the outcomes of others, rendering standard causal estimands ill-defined; ignoring such interference can bias VE evaluation and lead to ineffective vaccination policies. This article addresses the estimation of causal estimands for VE in the presence of partial interference using TND samples. Partial interference means that the vaccination of units within the same group/cluster may influence the outcomes of other members of the cluster. We define the population direct, spillover, total, and overall effects using the geometric risk ratio, which are identifiable under TND sampling. We investigate various stochastic policies for vaccine allocation in a counterfactual scenario, and identify policy-relevant VE causal estimands. We propose inverse-probability weighted (IPW) estimators for estimating the policy-relevant VE causal estimands with partial interference under the TND, and explore the statistical properties of these estimators.

*Cong and Fei contributed equally to this work.

1 Introduction

Vaccination campaigns targeting infectious diseases in populations have multifaceted effects, including boosting individual immunity and strengthening herd immunity Anderson and May (1985); Randolph and Barreiro (2020); Halloran and Hudgens (2018). Herd immunity, the indirect protection against infectious diseases, occurs when a large percentage of a population becomes immune through vaccination or prior infections.

Throughout the COVID-19 pandemic, the protective effects of mRNA vaccines have been evaluated using various methods, with many studies employing the test-negative design (TND) to assess real-world effectiveness (Dean et al., 2021; Schnitzer, 2022; Li et al., 2023). The TND is an observational study design that involves enrolling individuals exhibiting symptoms characteristic of the target infectious disease (such as COVID-19) who seek care for their symptoms and receive a test to determine the pathogen of interest (SARS-CoV-2, the virus that causes COVID-19) (Jackson and Nelson, 2013; Sullivan et al., 2014). The subsequent analysis compares the vaccination statuses between those who test positive (cases) and those who test negative (controls). An adjusted odds-ratio analysis produces an estimate of vaccine effectiveness (VE) for the prevention of outcomes of the target disease requiring care. Utilizing TND offers significant advantages in terms of efficiency and cost-effectiveness, while specifically helping to control for unobserved confounding differences in *healthcare-seeking behavior* between vaccinated and unvaccinated individuals Sullivan et al. (2014, 2016). However, the most causal interpretation of existing TND analytical strategies relies on the strong assumption that one person's vaccination status does not impact another's disease outcome, referred to as the "no interference" or "stable unit-treatment value (SUTVA)" assumption Cox (1958); Rubin (1980, 1986); Schnitzer et al. (2021). While this assumption may hold in certain contexts, it is not reasonable in many real-world applications, particularly in the context of vaccines and infectious diseases. For example, in a densely populated community, an individual's vaccination status could potentially reduce the likelihood of transmission to others, thereby impacting their disease outcomes as well Ogburn and VanderWeele (2014). Due to this methodological limitation, it is necessary to consider the distinct effects of a vaccine on individual immunity versus its role in developing herd immunity.

Partial interference (Sobel, 2006) is the most methodologically studied type of interference, where study individuals are grouped into non-overlapping clusters (e.g., households, schools), with interference assumed to only occur between individuals within the same cluster. The methodological development of IPW estimators under partial interference for direct and spillover effects involves defining causal effect estimands (Hudgens and Halloran (2008)) and constructing corresponding estimators, with derivation of their asymptotic properties as the number of clusters grows (Tchetgen and VanderWeele, 2012; Perez-Heydrich et al., 2014; Liu et al., 2016; Barkley et al., 2020). Unlike traditional settings where the no interference assumption holds, partial interference necessitates knowledge of the entire probability distribution of treatment assignments within each cluster, known as the *cluster-level propensity score* (CPS), rather than just conditional expectations. Building on the foundational work of Tchetgen and VanderWeele (2012) and Liu et al. (2016), these IPW estimators allow

unbiased estimation of average direct and spillover effects, accounting for intra-cluster interference. For inference, large sample (clusters) variance estimators from previous studies such as Perez-Heydrich et al. (2014) are utilized, and for scenarios with a small number of observed clusters, a bootstrap approach is recommended for more reliable inference Papadogeorgou et al. (2019). A notable application of the IPW method under partial interference is the evaluation of cholera vaccine effectiveness in a placebo-controlled, individually randomized trial in Matlab, Bangladesh Ali et al. (2005), which highlighted a significant spillover effect of cholera vaccination.

In observational settings, significant advancements have recently been made in causal estimation under the framework of partial interference. This includes developments such as parametric doubly robust estimators (Liu et al. (2019)), efficient semi-parametric estimation (Park and Kang (2022)), and efficient nonparametric sample-splitting estimators of stochastic policy effects (Lee et al. (2024)). The literature is particularly focused on *policy-relevant causal estimands*, which are defined under a counterfactual policy that modifies the distribution of treatment within the clusters Lee et al. (2024). These policies can be categorized into three types: (1) cluster-level proportions-based policies, including the Type-B policy (Tchetgen and VanderWeele (2012)) and generalized linear mixed model policies (GLMM, Papadogeorgou et al. (2019); Barkley et al. (2020)), (2) shifted propensity score policies, comprising the cluster incremental propensity score policy (CIPS, Kennedy (2019); Lee et al. (2024)) and cluster multiplicative shift policy (CMS, Wen et al. (2023); Lee et al. (2024)), and (3) the treated proportion bound (TPB) policy (Lee et al. (2024)). These hypothetical policies are more closely related to real-world interventions that modify recommendations and environmental factors (e.g., media, access to vaccination) rather than directly intervening on individual vaccination statuses. However, no work has yet proposed methodologies to evaluate direct, spillover, and overall effects of vaccination in the presence of partial interference under the TND, particularly in examining the different policies for these causal effects.

In this work, we focus on the estimation of policy-relevant causal estimands under partial interference from TND-sampled data, in the motivating context of mRNA vaccines to prevent COVID-19 outcomes. Existing partial interference methodological frameworks typically assume the random samples are taken from a super- or finite-population of clusters, with all members of the sampled clusters being subsequently observed. However, given the test-negative design sampling mechanism, we assume a super- or finite-population consisting of clusters of units, where only individuals within each cluster who meet the TND criteria are observed. To allow identifiability under the TND and unlike previous work, we propose defining the population direct effect (DE) and spillover effect (SE) using the geometric mean of risk ratios (RR). The geometric mean of RR not only appropriately reflects the multiplicative nature of risk ratios but also represents the "central tendency" on a logarithmic scale. This is consistent with the log-transformation approach, which stabilizes variance and normalizes the data distribution, thereby making statistical analysis more robust. Additionally, because of the form of the geometric mean of RR, corresponding DE and SE estimators can account for the varying sampling inclusion probabilities across different clusters, influenced by the unique characteristics of each cluster.

Our work considers VE estimation under the test-negative design (TND) in the presence of partial interference, where vaccination within a cluster may influence the outcomes of others. We define policy-relevant VE estimands using the geometric mean of risk ratios and propose inverse-probability weighted (IPW) estimators for their estimation. In Section 2, we introduce the TND sampling framework, define stochastic vaccine allocation policies, and establish the identifiability of direct and spillover effects under TND sampling. Section 3 presents our proposed IPW estimators, deriving their large-sample properties. In addition, building on M-estimation theory, we provide asymptotic results, demonstrating that our estimators are consistent and asymptotically normal under the correct specification of the cluster-level propensity score model. Lastly, Section 4 concludes the paper.

2 Methodology

2.1 Test-Negative Design Sampling and Notation

Suppose there are m observed clusters in the superpopulation according to the distribution F_0. For each cluster i where $i \in \{1, \ldots, m\}$, let N_i denote the total number of individuals in the underlying source population. In this paper, we consider the asymptotic region where the cluster size $N_i = O(1)$ and the number of clusters $m \to \infty$. We consider a test-negative design (TND) study that samples individuals hospitalized for COVID-19-like illness. For subject $j = 1, \ldots, N_i$ in cluster i, V_{ij} (categorical or binary) denotes COVID-19 vaccination status, C_{ij} represents measured confounders such as age, comorbidities, and employment sector, and U_{ij} represents unmeasured variables, which are assumed not to affect V_{ij}. The infection statuses are represented by two indicator variables: I_{ij}^{CoV}, indicating SARS-CoV-2 infection, and I_{ij}^{Other}, which equals 1 for other infections causing COVID-19-like symptoms. W_{ij} indicates COVID-19-like symptoms, occurring only if $I_{ij}^{\text{CoV}} = 1$ or $I_{ij}^{\text{Other}} = 1$, and H_{ij} indicates hospitalization for these symptoms. The TND samples from those with an infection ($I_{ij}^{\text{CoV}} = 1 \cup I_{ij}^{\text{Other}} = 1$), symptoms ($W_{ij} = 1$), and hospitalization ($H_{ij} = 1$). We use the indicator function $\mathbb{I}$ to define $T_{ij} = \mathbb{I}(I_{ij}^{\text{CoV}} = 1 \cup I_{ij}^{\text{Other}} = 1, W_{ij} = 1, H_{ij} = 1)$, representing the inclusion criteria for the TND study. Further, by design, the set of infected and symptomatic individuals for whom $W_{ij} = 1$ and $H_{ij} = 1$ is the same as the set where $H_{ij} = 1$; therefore, eligibility for TND selection (T_{ij}) is determined solely by hospitalization status (H_{ij}), such that $T_{ij} = H_{ij}$. Within each cluster, we observe censored samples with data $(C_{ij}, V_{ij}, I_{ij})\mathbb{I}(H_{ij} = 1)_{j=1}^{N_i}$ for $i = 1, \ldots, m$. The combined outcome of interest is defined as $Y_{ij} = \mathbb{I}(I_{ij}^{\text{CoV}} = 1, H_{ij} = 1)$, indicating a hospitalized (or medically-attended) symptomatic SARS-CoV-2 infection. This outcome involves three consecutive steps: infection with SARS-CoV-2, development of symptoms, and subsequent hospitalization due to these symptoms Schnitzer (2022). The event $Y_{ij} = 1$ identifies cases within the TND sample, while the condition $\{Y_{ij} = 0, H_{ij} = 1\}$, or equivalently $\{I_{ij}^{\text{Other}} = 1, H_{ij} = 1\}$, corresponds to the controls in the sample.

2.2 Stochastic Policies

We define a *stochastic policy* $Q(\sim| \mathbf{C}_i, N_i)$ as a probability distribution on $\{0,1\}^{N_i}$ such that a cluster of size N_i with cluster-level covariate $\mathbf{C}_i$ receives vaccine $\boldsymbol{v}_i$ with probability $Q(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i)$ in the conterfactual scenario. Previously proposed policies include the Type B policy (Tchetgen and VanderWeele (2012)), the generalized linear mixed-effects model (GLMM) shift policy (Papadogeorgou et al. (2019); Barkley et al. (2020)), cluster incremental propensity score policy (Kennedy (2019); Lee et al. (2024)), and cluster multiplicative shift policy (Wen et al. (2023); Lee et al. (2024)). First, the Type B policy (Q_{B}) is defined as:

$$Q_{\mathrm{B}}(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i; \alpha) = \prod_{j=1}^{N_i} \alpha^{v_{ij}}(1-\alpha)^{1-v_{ij}},$$

where α represents the independent and fixed probability of individuals receiving a vaccine, i.e., $\pi_{ij} := \mathbb{P}(V_{ij}=1) = \alpha$ for any individual j within cluster i. We note that under the Type-B policy, α denotes both the individual's propensity and the cluster-average propensity to be vaccinated because the cluster-average propensity of treatment equals $\frac{1}{N_i}\sum_{j=1}^{N_i}\pi_{ij} = \frac{1}{N_i}\sum_{j=1}^{N_i}\alpha = \alpha$.

In real-world settings, the relevance and practicality of the Type B policy are limited; therefore, Papadogeorgou et al. (2019) and Barkley et al. (2020) suggested alternatives wherein vaccination probabilities may vary among individuals. The GLMM shift policy of Papadogeorgou et al. (2019) allowed for the individual's propensity to depend on covariates, while ensuring that the cluster-average propensity of treatment remains at a certain hypothetical value (say β). Specifically, for an individual j within cluster i and with covariate vector C_{ij}, the probability of receiving the vaccine can be represented as $g\left\{\pi(C_{ij};\delta_{0i},\boldsymbol{\delta})\right\} := g\left\{\mathbb{P}(V_{ij}=1 \mid C_{ij};\delta_{0i},\boldsymbol{\delta})\right\} = \delta_{0i} + C_{ij}^{\top}\boldsymbol{\delta}$ using the appropriate link function g, for some fixed, pre-specified value of $\boldsymbol{\delta}$, and for δ_{0i} satisfying $\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{P}\left(V_{ij}=1 \mid C_{ij};\delta_{0i},\boldsymbol{\delta}\right) = \beta$. Then, in the counterfactual world, under the assumption that the variables $V_{ij}'s$ are conditionally independent given $\mathbf{C}_i$, and V_{ij} is conditionally independent of $\mathbf{X}_{i,-j}$ given C_{ij}, the corresponding GLMM policy (denoted as Q_{GLMM}) is defined as

$$\begin{aligned} &Q_{\mathrm{GLMM1}}(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i; \delta_{0i}, \boldsymbol{\delta}) = \prod_{j=1}^{N_i}\left\{g^{-1}\left(\delta_{0i}+C_{ij}^{\top}\boldsymbol{\delta}\right)\right\}^{v_{ij}}\left\{1-g^{-1}\left(\delta_{0i}+C_{ij}^{\top}\boldsymbol{\delta}\right)\right\}^{1-v_{ij}}, \\ &\text{s.t.} \quad \frac{1}{N_i}\sum_{j=1}^{N_i}\left\{g^{-1}\left(\delta_{0i}+C_{ij}^{\top}\boldsymbol{\delta}\right)\right\} = \beta. \end{aligned} \tag{1}$$

Note that, for each cluster, the cluster-average propensity of treatment is set at a certain hypothetical value β. Alternatively, one may consider including cluster-specific random effects in the policy formulation. Incorporating a cluster-specific random effect u_i assumed to follow a distribution, e.g., a Gaussian distribution function Φ with mean zero and variance σ^2, the probability of an individual j within a cluster i receiving the vaccine is $g\left\{\mathbb{P}(V_{ij}=1 \mid C_{ij}, N_i, u_i)\right\} = \gamma_0 + C_{ij}^{\top}\boldsymbol{\gamma} + u_i$ with a constant γ_0; then, the second type GLMM policy can be defined as

$$\begin{aligned} &Q_{\mathrm{GLMM2}}(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i; \gamma_0, \boldsymbol{\gamma}) = \int \prod_{j=1}^{N_i}\left\{g^{-1}\left(\gamma_0+C_{ij}^{\top}\boldsymbol{\gamma}+u\right)\right\}^{v_{ij}}\left\{1-g^{-1}\left(\gamma_0+C_{ij}^{\top}\boldsymbol{\gamma}+u\right)\right\}^{1-v_{ij}} d\Phi(u;\sigma^2), \\ &\text{s.t.} \quad \int\left\{N_i^{-1}\sum_{j=1}^{N_i}\int g^{-1}\left(\gamma_0+C_{ij}^{\top}\boldsymbol{\gamma}+u\right)d\Phi(u;\sigma^2)\right\}dF_{\mathbf{C}} = \beta. \end{aligned} \tag{2}$$

Under a certain assumption discussed in Barkley et al. (2020) Section 4.2, i.e., the conditional odds ratio of treatment remains identical for any two individuals within a cluster, regardless of whether they are in the factual or counterfactual scenario, the parameters $\boldsymbol{\gamma}$ and those in the distribution Φ can be identifiable from the observed treatment GLMM model. Unlike the previously defined Q_{GLMM1} in Eq. (1), where V_{ij}s are conditionally independent, the Q_{GLMM2} in Eq. (2) accounts for correlation among individuals within the same cluster because of the cluster-specific random effects, with the degree of correlation depending on the variance of Φ. Moreover, different from Q_{GLMM1}, which sets the cluster-average propensity of treatment to β, Q_{GLMM2} allows for greater flexibility by ensuring that the marginal probability of treatment selection equals β, while the average individual-level propensity scores within any given cluster may deviate from β.

Building upon Kennedy (2019)'s incremental propensity score interventions for the context of clustered interference, Lee et al. (2024) introduced the Cluster Incremental Propensity Score (CIPS) policy. This policy shifts the observed propensity score distribution by a user-defined function of $\mathbf{C}_i$ and N_i, denoted as $\gamma(\mathbf{C}_i, N_i)$. Specifically, the counterfactual treatment odds are $\gamma(\mathbf{C}_i, N_i)$ times the observed odds. Formally, for the propensity score (π_{ij}) of individual j in cluster i, defined as $\pi_{ij} = \mathbb{P}(V_{ij} = 1 \mid \mathbf{C}_i, N_i)$, the shifted propensity score ($\pi_{ij,\gamma}$) is given by $\pi_{ij,\gamma} := \mathbb{P}_\gamma\left(V_{ij} = 1 \mid \mathbf{C}_i, N_i\right) = \gamma\left(\mathbf{C}_i, N_i\right)\pi_{ij} / \left\{\gamma\left(\mathbf{C}_i, N_i\right)\pi_{ij} + 1 - \pi_{ij}\right\}$, ensuring that the odds ratio $\frac{\pi_{ij,\gamma}}{1-\pi_{ij,\gamma}} / \frac{\pi_{ij}}{1-\pi_{ij}} = \gamma(\mathbf{C}_i, N_i)$. The CIPS policy can then be expressed as

$$Q_{\text{CIPS}}\left(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i; \gamma\right) = \prod_{j=1}^{N_i} \pi_{ij,\gamma}^{v_{ij}} (1 - \pi_{ij,\gamma})^{1-v_{ij}}.$$

Notably, the CIPS policy preserves the within-cluster ranking of unit vaccination probabilities without relying on parametric models. Varying the γ function allows for the generation of different impacts of the policy on the propensity scores. For instance, setting $\gamma(\mathbf{C}_i, N_i)$ to a constant value of 2 results in a scenario where we investigate the relative risk of COVID-19 under the policy that doubles the current odds of vaccination.

A recently defined policy known as the cluster multiplicative shift (CMS) policy, which extends the multiplicative shift policy of Wen et al. (2023) to the partial interference setting, was investigated in Lee et al. (2024). CMS selectively increases the vaccination propensity for individuals with specific covariate values or patterns, such as those indicating individuals at high risk for adverse outcomes. We can use this kind of policy to estimate effects on mean outcomes under counterfactual scenarios where a larger proportion of high-risk individuals receive vaccination. For instance, a CMS policy can be defined based on a shifted propensity score, determined by a user-specified factor λ, and some single-dimensional binary covariate $C_{ij}^* \in C_{ij}$, such that $\pi_{ij,\lambda} := \mathbb{P}_\lambda\left(V_{ij} = 1 \mid \mathbf{C}_i, N_i\right) = \left(1 - \lambda + \lambda\pi_{ij}\right)C_{ij}^* + \pi_{ij}\left(1 - C_{ij}^*\right)$, representing the shifted probability of treatment for individual j in cluster i. Note that the propensity score distribution is adjusted solely for individuals with $C_{ij}^* = 1$, thereby preserving the conditional independence of treatment selection within clusters. The CMS policy can then be expressed as $Q_{\text{CMS}}\left(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i; \lambda\right) = \prod_{j=1}^{N_i} \pi_{ij,\lambda}^{v_{ij}} (1 - \pi_{ij,\lambda})^{1-v_{ij}}$.

2.3 Policy-relevant Causal Estimands and Identifiability

Under partial interference, for m clusters, each containing N_i individuals, where $i \in \{1, \ldots, m\}$, we first define $Y^*_{ij}(\boldsymbol{v}_i)$ as the potential outcome for individual j in cluster i if the vaccination status of cluster i is N_i-vector $\boldsymbol{v}_i$. Then, we define the *individual average potential outcome* for individual j in cluster i under policy Q as:

$$\bar{Y}_{ij}(Q) = \mathbb{E}_Q\left[Y^*_{ij}(\boldsymbol{V}_i)\right] = \sum_{\boldsymbol{v}_i \in \mathcal{V}(N_i)} Y^*_{ij}(\boldsymbol{v}_i) Q(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i),$$

where the sum is taken over all possible configurations of cluster-level vaccine vectors $\boldsymbol{v}_i$ under the policy distribution Q. This expression represents the expected outcome of individuals in the counterfactual scenario where the vaccine is allocated based on policy Q, averaging over all possible vaccination configurations within the cluster given the policy Q. Next, we define the *cluster average marginal potential outcome* for cluster i under policy Q by

$$\mu_i(Q) = N_i^{-1} \sum_{j=1}^{N_i} \bar{Y}_{ij}(Q) = N_i^{-1} \sum_{j=1}^{N_i} \left[\sum_{\boldsymbol{v}_i \in \mathcal{V}(N_i)} Y^*_{ij}(\boldsymbol{v}_i)\, Q(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i) \right], \tag{3}$$

which takes the average over all the units in a cluster, considered as a finite population of a certain size (e.g., N_i for cluster i). Similarly, we define the *cluster average potential outcome* for cluster i under policy Q when a given individual j receives vaccine v by

$$\mu_i(v, Q) = N_i^{-1} \sum_{j=1}^{N_i} \left[\sum_{\boldsymbol{v}_{i(-j)} \in \mathcal{V}(N_i - 1)} Y^*_{ij}\left(v, \boldsymbol{v}_{i(-j)}\right) Q\left(\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right) \right], \tag{4}$$

which represents the cluster i's averaged outcome in the counterfactual world where the vaccine is assigned according to policy Q, but the vaccine of individual j is independently set to v.

To represent causal effects, we define the cluster-level policy-relevant causal estimand by contrasting $\mu_i(v, Q)$ and $\mu_i(Q)$ across different values of v and policies Q (Tchetgen and VanderWeele, 2012). Varying either the value of v or the policy Q while holding the other fixed, we define the *population direct effect* and *population spillover effect* based on the definition of the geometric risk ratio as follows.

Definition 1 *(Population direct and spillover effects of the geometric risk ratio). Assuming that clusters are observed from a super-population according to a distribution F_0, the geometric risk ratio describes the geometric mean of the risk ratios:*

$$DE(Q) := \prod \left(\frac{\mu_i(1, Q)}{\mu_i(0, Q)}\right)^{dF_0}, \quad \textit{and} \quad SE_v(Q', Q) := \prod \left(\frac{\mu_i(v, Q')}{\mu_i(v, Q)}\right)^{dF_0}, \textit{ for } v \in \mathcal{V}, \tag{5}$$

where $\prod$ denotes the product operator if $\mathbf{C}_i$ is discrete, or the product integral if $\mathbf{C}_i$ is continuous.

The above-defined population direct and spillover effects of the geometric risk ratio can also be expressed as follows:

$$DE(Q) = \exp\left(\mathbb{E}_{F_0}\left[\log\left(\frac{\mu_i(1, Q)}{\mu_i(0, Q)}\right)\right]\right), \text{ and } SE_v(Q', Q) = \exp\left(\mathbb{E}_{F_0}\left[\log\left(\frac{\mu_i(v, Q')}{\mu_i(v, Q)}\right)\right]\right), \text{ for } v \in \mathcal{V}; \tag{6}$$

since, for DE, for example, we have

$$DE(Q) = \exp\left(\log\left[\prod \left(\frac{\mu_i(1, Q)}{\mu_i(0, Q)}\right)^{dF_0}\right]\right) = \exp\left(\sum \left[\log\left(\frac{\mu_i(1, Q)}{\mu_i(0, Q)}\right)\right] dF_0\right) = \exp\left(\mathbb{E}_{F_0}\left[\log\left(\frac{\mu_i(1, Q)}{\mu_i(0, Q)}\right)\right]\right),$$

Then, the *VE direct effect* and *VE spillover effect* can be written as $VE^{DE}(Q) = 1 - DE(Q)$, and $VE_v^{SE}(Q) = 1 - SE_v(Q', Q)$, respectively. Following Definition 1, the proposed framework introduces a novel approach to quantifying the population direct effect and spillover effect using the geometric mean of risk ratios. This approach is particularly advantageous within the context of vaccine effectiveness, as it aligns well with the inherent multiplicative nature of risk ratios and accurately reflects the "central tendency" on a logarithmic scale. The geometric mean serves not only to stabilize variance and normalize the data distribution, enhancing the robustness of our statistical analysis, but also to address the challenges linked to varying sampling inclusion probabilities across different clusters under TND. By leveraging the risk ratio at the cluster level, this method effectively cancels out common but unknown or unquantifiable terms in the estimation process, such as those inherent in $\hat{\mu}_i$. The subsequent section 3.2.1 will delve deeper into the mechanics and rationale behind TND sampling, offering comprehensive insights into how the geometric mean framework contributes to more accurate and interpretable measures of the direct and spillover effects within diverse cluster settings.

Comparing to the previous work defined, e.g., the direct effect, using a ratio scale as the form of

$$DE^{HH}(Q) = \frac{\mathbb{E}_{F_0}[\mu_i(1, Q)]}{\mathbb{E}_{F_0}[\mu_i(0, Q)]} = \exp\left(\log\left[\frac{\mathbb{E}_{F_0}[\mu_i(1, Q)]}{\mathbb{E}_{F_0}[\mu_i(0, Q)]}\right]\right) \tag{7}$$

where the superscript HH in DE^{HH} represents the notions of direct, indirect (spillover), total, and overall causal effects as defined by Hudgens and Halloran (2008). The difference between our estimand and their estimand can be characterized as follows.

Lemma 1 (Difference between $DE(Q)$ and $DE^{HH}(Q)$) *Let $DE(Q)$ and $DE^{HH}(Q)$ be defined as in equations* (6) *and* (7)*, respectively. Then*

$$|DE(Q) - DE^{HH}(Q)| \leq DE^{HH}(Q) \cdot \left| \frac{\mathbb{E}(\mu_i^2(1, Q))}{2[\mathbb{E}(\mu_i(1, Q))]^2} + \frac{\mathbb{E}(\mu_i^2(0, Q))}{2[\mathbb{E}(\mu_i(0, Q))]^2} + o(1) \right| \tag{8}$$

Similar bounds apply to $SE^{HH}(Q, Q')$, $TE^{HH}(Q)$, and $OE^{HH}(Q)$ by replacing $\mu_i(\cdot))$ with the corresponding quantities in each estimand. Lemma 1 shows that the difference between $DE(Q)$ and $DE^{HH}(Q)$ is of order one and should therefore be expected. Moreover, if both $DE(Q)$ and the terms on the right-hand side of the inequality can be consistently estimated, then this result provides a basis for constructing upper and lower bounds for $DE^{HH}(Q)$ using an estimator of $DE(Q)$.

Because $\log(x)$ is a strictly convex function of x, we have, using Jensen's inequality,

$$\mathbb{E}_{F_0}\left\{\log \frac{\mu_i(1, Q)}{\mu_i(0, Q)}\right\} \leq \log \mathbb{E}_{F_0}\left\{\frac{\mu_i(1, Q)}{\mu_i(0, Q)}\right\} \tag{9}$$

and the proposed $DE(Q) \leq DE^{HH}(Q)$.

Based on the above definitions of direct and spillover effects, we can further define the *population total effect* and *population overall effect* using the geometric risk ratio as follows:

$$TE(Q', Q) = \exp\left(\mathbb{E}_{F_0}\left[\log\left(\frac{\mu_i(1, Q')}{\mu_i(0, Q)}\right)\right]\right), \text{ and } OE(Q', Q) = \exp\left(\mathbb{E}_{F_0}\left[\log\left(\frac{\mu_i(Q')}{\mu_i(Q)}\right)\right]\right). \tag{10}$$

do we need this paragraph? It follows immediately from their definitions that the total effect can be decomposed as the product of direct and indirect causal effects, such that $TE(Q', Q) = DE(Q')SE_0(Q', Q)$. The VE total effect $VE^{TE}(Q', Q) = 1 - TE(Q', Q)$ measures the reduction in the risk of infection for a vaccinated individual in a cluster assigned treatment based on one policy Q' compared to a counterfactual scenario where the unvaccinated individual is in the cluster from the same study population assigned treatment based on another policy Q. The VE overall effect $VE^{OE}(Q', Q) = 1 - OE(Q', Q)$ measures the reduction in the risk of infection for an individual, whether vaccinated or unvaccinated, in a cluster with treatment assignment based on one policy Q' compared to what would have happened to an individual in the cluster from the same study population assigned treatment based on another policy Q.

The standard identification of direct, spillover, total and overall effects requires three assumptions: (1) *Consistency:* $Y_{ij} = \sum_{\boldsymbol{v}_i \in \mathcal{V}(N_i)} Y^*_{ij}(\boldsymbol{v}_i)\mathbb{I}(\boldsymbol{V}_i = \boldsymbol{v}_i)$; (2) *Conditional Exchangeability:* For F_0, $\mathbf{Y}^*_i(\boldsymbol{v}_i) \perp\!\!\!\perp \boldsymbol{V}_i \mid \mathbf{C}_i, N_i$ for all $\boldsymbol{v}_i \in \mathcal{V}(N_i)$; (3) *Positivity:* for some $\epsilon \in (0, 1)$, $\epsilon < \mathbb{P}(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{C}_i, N_i) < 1 - \epsilon$ for all $\boldsymbol{v}_i \in \mathcal{V}(N_i)$, and applying iterated expectations, the causal estimands defined in Equaitons 4 and 3 can be identified as

$$\begin{aligned}\mathbb{E}_{F_0}(\log[\mu_i(v, Q)]) &= \mathbb{E}_{F_0}\left(\log[N_i^{-1}\sum_{j=1}^{N_i}\bar{Y}_{ij}(v, Q)]\right)\\ &= \mathbb{E}_{F_0}\left(\log\left\{N_i^{-1}\sum_{j=1}^{N_i}\left[\sum_{\boldsymbol{v}_i \in \mathcal{V}(N_i)}\mathbb{P}\left(Y_{ij} = 1 \mid \boldsymbol{V}_i = \boldsymbol{v}_i, \mathbf{C}_i, N_i\right)\mathbb{I}(v_{ij} = v)Q\left(\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\right]\right\}\right);\end{aligned} \tag{11}$$

and

$$\mathbb{E}_{F_0}(\log[\mu_i(Q)]) = \mathbb{E}_{F_0}\left(\log[N_i^{-1}\sum_{j=1}^{N_i}\bar{Y}_{ij}(Q)]\right) = \mathbb{E}_{F_0}\left(\log\left\{N_i^{-1}\sum_{j=1}^{N_i}\left[\sum_{\boldsymbol{v}_i \in \mathcal{V}(N_i)}\mathbb{P}\left(Y_{ij} = 1 \mid \boldsymbol{V}_i = \boldsymbol{v}_i, \mathbf{C}_i, N_i\right)Q(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i)\right]\right\}\right). \tag{12}$$

See detailed proof in Appendix A. Building on the above two identified policy-relevant causal estimands, the defined population direct, spillover, total, and overall effects of the geometric risk ratio can be correspondingly identified; for example, for the population direct effects, we have:

$$\log[DE(Q)] = \mathbb{E}_{F_0}\left[\log\left(\frac{N_i^{-1}\sum_{j=1}^{N_i}\left[\sum_{\boldsymbol{v}_i \in \mathcal{V}(N_i)}\mathbb{P}\left(Y_{ij} = 1 \mid \boldsymbol{V}_i = \boldsymbol{v}_i, \mathbf{C}_i, N_i\right)\mathbb{I}(v_{ij} = 1)Q\left(\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\right]}{N_i^{-1}\sum_{j=1}^{N_i}\left[\sum_{\boldsymbol{v}_i \in \mathcal{V}(N_i)}\mathbb{P}\left(Y_{ij} = 1 \mid \boldsymbol{V}_i = \boldsymbol{v}_i, \mathbf{C}_i, N_i\right)\mathbb{I}(v_{ij} = 0)Q\left(\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\right]}\right)\right]. \tag{13}$$

3 Estimation of Policy-relevant Estimands

3.1 Estimation under simple random sampling (SRS)

For each cluster, consider observed data under simple random sampling (SRS) from a finite population (with size N_i) of cluster i denoted by $\{Y_{ik}, C_{ik}, V_{ik}, k \in \mathcal{S}_i, i = 1, 2, \cdots, M\}$. For example, the cluster sample mean of the outcome of cluster i is written $\bar{y}_i = \frac{1}{|\mathcal{S}_i|}\sum_{k \in \mathcal{S}_i} Y_{ik}(Q)$, where $|\mathcal{S}_i|$ is the cluster sample size of the simple random sample $\mathcal{S}_i$.

In causal inference, the propensity score (PS) (Rosenbaum and Rubin, 1983) is used to balance the distributions of confounders between the treatment and the control group. In the presence of partial interference, the cluster-level PS, $\tau(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i) = \mathbb{P}(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{C}_i, N_i)$, is the joint probability of the treatment status of all individuals within the cluster given all individuals' covariates (Zhang et al. (2023); Kang et al. (2023)).

Under SRS within each cluster, for $v \in \mathcal{V}$, inverse probability weighting (IPW) estimators of $\mu_i(v, Q)$ and $\mu_i(Q)$ are

$$\hat{\mu}_i^{\text{ipw,SRS}}(v, Q) = (|\mathcal{S}_i|)^{-1} \sum_{k \in \mathcal{S}_i} \frac{Y_{ik} \mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\hat{\tau}_i}, \text{ and } \hat{\mu}_i^{\text{ipw,SRS}}(Q) = (|\mathcal{S}_i|)^{-1} \sum_{k \in \mathcal{S}_i} \frac{Y_{ik} Q(\boldsymbol{V}_i \mid \mathbf{C}_i)}{\hat{\tau}_i} \tag{14}$$

where, with parameters $\boldsymbol{\alpha}$ and σ_B^2, the cluster-level PS, specifically $\hat{\tau}_i = \hat{\mathbb{P}}_{|\mathcal{S}_i|}\left(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{C}_i, \hat{\boldsymbol{\alpha}}, \hat{\sigma}_B^2\right)$, is modelled as follows. Building upon the previous work of Tchetgen and VanderWeele (2012), Perez-Heydrich et al. (2014), Liu et al. (2016), Barkley et al. (2020), and Kang et al. (2023), we employ a mixed-effects model for the treatment to model the cluster-level PS, with the link function g (e.g., the logit link):

$$\begin{aligned} &g\left(\mathbb{P}(V_{ik} = 1 \mid C_{ik}, B_i, \boldsymbol{\alpha})\right) = \boldsymbol{\alpha}^\top C_{ik} + B_i, \\ &\tau_i = \mathbb{P}\left(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) = \int_{-\infty}^{\infty} \prod_{k \in \mathcal{S}_i} \mathbb{P}(V_{ik} = v_{ik} \mid C_{ik}, B_i, \boldsymbol{\alpha})\, \mathbb{P}(B_i = b)\, db,\ C_{ik} \perp B_i,\ B_i \overset{i.i.d.}{\sim} N\left(0, \sigma_B^2\right), \end{aligned} \tag{15}$$

where the Normal random effect B_i with mean 0 and variance σ_B^2 accounts for the correlation in treatment selection among individuals within the same cluster. Then, the model parameters for the cluster-level PS, which includes both $\boldsymbol{\alpha}$ and σ_B^2, are estimated by $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}_B^2$ using maximum likelihood estimation or a restricted maximum likelihood estimation (Bates et al., 2015).

To demonstrate the unbiasedness of the estimators for the cluster average marginal potential outcome for $v \in \mathcal{V}$ and the cluster average potential outcome, we consider the randomness arising from the simple random sampling and, conditional on the sampling, the randomness from the treatment assignments. Specifically, we have:

$$\mathbb{E}\{\hat{\mu}_i^{\text{ipw,SRS}}(v, Q)] = \mathbb{E}_{\mathcal{S}_i}\left[\mathbb{E}\left\{\hat{\mu}_i^{\text{ipw,SRS}}(v, Q) \mid \mathcal{S}_i\right\}\right],$$

where the outer expectation ($\mathbb{E}_{\mathcal{S}_i}$) is with respect to the probability sampling design specified by the probability measure $\mathscr{P}$ over the set of all possible candidate samples. For simple random sampling without replacement of cluster i, $\mathscr{P}(\mathcal{S}_i)$ is given by $\mathscr{P}(\mathcal{S}_i) = \binom{N_i}{n_i}$ if $|\mathcal{S}_i| = n_i$; otherwise, $\mathscr{P}(\mathcal{S}_i) = 0$, where, again, N_i is the total number of individuals in the underlying source population for cluster i, and n_i is the SRS sample size for cluster i.

For the inner expectation, following the theories outlined in existing IPW estimators for partial interference by Tchetgen and VanderWeele (2012), Perez-Heydrich et al. (2014), or Papadogeorgou et al. (2019), we can demonstrate that the estimators in Equation (14) are unbiased for the cluster average potential outcome if the cluster-level propensity score model (15) is correctly specified. Consequently, the corresponding population total effect estimator (as well as those of the other defined effects), which relies on Equation (14), is consistent and asymptotically normal, and the sandwich-type variance estimators are similar to those in Perez-Heydrich et al. (2014). The detailed results under SRS are given in Appendix D for our proposed population direct and spillover effects of the geometric risk ratio. Besides the

above results regarding the SRS of each cluster in the partial interference analysis, the goal of this subsection is to introduce sampling randomness into our analysis, which leads to our focus on TND sampling within clusters.

3.2 Estimation under the test-negative design sampling

3.2.1 TND sampling mechanism in non-probability sampling and assumptions

The TND is a type of non-probability sampling. Unlike probability sampling, where each element of the population has a known chance of being selected, non-probability sampling includes methods where some population elements have no chance of selection or their selection probability is not known in advance (Wu and Thompson (2020)). Thus, one typically selects and fits a model for the unknown selection mechanism, where the outcome is the inclusion indicator variable based on the unit's characteristics. However, in TND sampling, the selection probability cannot be estimated because we only have information on the selected individuals and not on the source population, which is nevertheless our target of inference. Fortunately, the risk ratio form of the target estimand helps us circumvent the issue of the inestimable selection probability (Schnitzer et al., 2021).

Consider the inclusion criterion for the test-negative design. Let the TND-sampled data for cluster i be denoted as $k \in \mathcal{S}_{\text{TND},i}$, where $i = 1, 2, \ldots, M$. Define the TND sample indicator variable H_{ij} as $\mathbb{I}(j \in \mathcal{S}_{\text{TND},i})$ for $j = 1, 2, \ldots, N_i$, with the indicator function $\mathbb{I}$. This is a Bernoulli random variable under the TND sampling design and is defined for every j in the cluster i, with probability $\rho_{ij} := \mathscr{P}(\mathcal{S}_{TND,i}) = \mathbb{E}_{\mathcal{S}_{TND,i}}(H_{ij}) = \mathbb{P}_i(j \in \mathcal{S}_{TND,i}) = \mathbb{P}_i(H_{ij} = 1)$, where $\mathbb{E}_{\mathcal{S}_{TND,i}}$ is the expectation with respect to the TND sampling design over the set of all possible candidate TND samples. Then, building on the Horvitz-Thompson principle, for $j = 1, 2, \cdots, N_i$, and any fixed function f we have the TND sampling unbiasedness equality, which we refer to as the *Standard TND Sampling Debiased Principle*, as follows:

$$\begin{aligned}
\mathbb{E}_{\mathcal{S}_{TND,i}}\left(\frac{1}{|\mathcal{S}_{TND,i}|}\sum_{j\in\mathcal{S}_{TND,i}} f(y_{ij}, v_{ij}, \boldsymbol{v}_{i(-j)}, \mathbf{C}_i)\right) &= \mathbb{E}_{\mathcal{S}_{TND,i}}\left(\frac{1}{|\mathcal{S}_{TND,i}|}\sum_{j=1}^{N_i} H_{ij} f(y_{ij}, v_{ij}, \boldsymbol{v}_{i(-j)}, \mathbf{C}_i)\right) \\
&= \frac{1}{|\mathcal{S}_{TND,i}|}\sum_{j=1}^{N_i} f(y_{ij}, v_{ij}, \boldsymbol{v}_{i(-j)}, \mathbf{C}_i)\mathbb{E}_{\mathcal{S}_{TND,i}}\left(H_{ij}\right) \\
&= \frac{\rho_{ij}}{|\mathcal{S}_{TND,i}|}\sum_{j=1}^{N_i} f(y_{ij}, v_{ij}, \boldsymbol{v}_{i(-j)}, \mathbf{C}_i) = \frac{\rho_{ij}N_i}{|\mathcal{S}_{TND,i}|}(N_i)^{-1}\sum_{j=1}^{N_i} f(y_{ij}, v_{ij}, \boldsymbol{v}_{i(-j)}, \mathbf{C}_i).
\end{aligned}$$

It is important to note that the factor term in the TND Sampling Debiased Principle, i.e., $\rho_{ij}N_i \Big/ |\mathcal{S}_{TND,i}|$, includes elements ρ_{ij} and N_i, which is unknown and cannot be estimated using TND samples. In addition, in the presence of interference, dependence between individuals' H is another layer of the complicity of the problem, due the the casual path such that $\boldsymbol{v}_i \to I_{ij}^{\text{CoV}}(\boldsymbol{v}_i) \to W_{ij} \to H_{ij}$.

However, in the presence of partial interference, the standard identifiability condition based on the inverse of individual-level selection probabilities (e.g., ρ_{ij}^{-1}) in test-negative design (TND) studies is not necessary. Instead, it becomes essential to adopt a more appropriate

quantity that captures the bias induced by the TND selection mechanism under interference. To address this, we introduce the *joint TND inclusion probability* for each cluster, defined as, with $\boldsymbol{H}_i = (H_{i1}, ..., H_{ij}, ..., H_{iN_i})^\top$ the TND inclusion indicator vector,

$$\rho_i^0 := \mathbb{P}(\boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i),$$

where $\mathcal{H}_i$ denotes the set of valid inclusion patterns for the TND sample in cluster i, i.e.,

$$\mathcal{H}_i = \left\{\boldsymbol{h}_i \in \{0,1\}^{N_i} \mid \boldsymbol{h}_i \text{ satisfies all TND inclusion criteria}\right\}.$$

Unlike the marginal inclusion probability $q_0 = \mathbb{P}(H = 1)$ assumed in non-interference settings, ρ_i^0 captures within-cluster dependencies, including potential correlations between individual outcomes and their peers' selection statuses. This refinement allows for a more accurate representation of the distribution shift between the observed TND sample and the target population, supporting valid inference under partial interference. Our target estimand is expressed on the risk ratio scale, which has been advocated in recent work (e.g., Schnitzer (2022), Jiang et al. (2023)) as a robust choice under such settings.

Moreover, our newly defined population-level direct and spillover effects based on the geometric risk ratio, which aggregate cluster-specific risk ratios, allow for the cancellation of inestimable terms. In particular, the joint TND selection probabilities ρ_i^0 may vary across clusters due to differences in realized values of $\boldsymbol{h}_i$, resulting in heterogeneity in ρ_i^0. As a consequence, the form of $DE^{HH}(Q)$ is only *partially identifiable*, since these heterogeneous selection probabilities cannot be canceled out across clusters. In contrast, the geometric formulation defined in Definition 1 enables cancellation of the selection probabilities within each cluster, rendering the estimands fully identifiable. See Lemma 5 for details.

Finally, we present three assumptions under TND sampling to ensure valid statistical estimation. For the cluster i, with $\boldsymbol{H}_i$ the vector of TND selection indicators for individuals within cluster i and $\mathcal{H}_i$ denotes the set of inclusion patterns that satisfy all TND criteria, we assume (A1) $\mathbb{P}_i(\mathbf{Y}_i = \mathbf{y}_i \mid \boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i) = 0$ for $\mathbf{y}_i \in \{\mathbf{0}, \mathbf{1}\}$; (A2) $0 < \mathbb{P}_i(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{C}_i, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i) < 1$; (A3) $0 < \mathbb{P}_i(\mathbf{C}_i = \mathbf{c}_i \mid \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i) < 1$; (A4) cluster-level control exchangeability assumption: in each cluster, being hospitalized for symptoms of *another infection* is independent of vaccination conditional on covariates, that is $\{\mathbf{Y}_i = \mathbf{0}, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i\} \perp\!\!\!\perp \boldsymbol{V}_i \mid \mathbf{C}_i$.

(A1) - (A3) are essentially positivity-type cluster-level assumptions for the TND samples in each cluster. Specifically, (A1) requires that each cluster contains both cases and controls. (A2) requires positivity for the treatment assignment and (A3) pertains positivity for the cluster-level covariates. Lastly, (A4) is the cluster-level control exchangeability assumption that was discussed in Jiang et al. (2023), which is sufficient but not necessary for the typical TND assumption that the vaccine does not have an impact on non-target infections (Feng et al., 2017).

3.2.2 Proposed estimation under the TND sampling

Under cluster-level TND sampling, with the sample data for cluster i and $k \in \mathcal{S}_{TND,i}$ for $i = 1, 2, \cdots, m$, we have the following inverse probability weighting estimators:

$$\hat{\mu}_i^{\text{ipw,TND}}(v, Q) = \sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik} \mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\, |\mathcal{S}_{TND,i}|} \tilde{\rho}_i^0 \quad (v \in \mathcal{V}), \text{ and } \hat{\mu}_i^{\text{ipw,TND}}(Q) = \sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik} Q(\boldsymbol{V}_i \mid \mathbf{C}_i)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\, |\mathcal{S}_{TND,i}|} \tilde{\rho}_i^0 \tag{16}$$

where $|\mathcal{S}_{TND,i}|$ is the cluster sample size of the TND sample $\mathcal{S}_{TND,i}$, and $\tilde{\rho}_i^0 := \frac{|\mathcal{S}_{TND,i}|}{N_i} \rho_i^0$ where ρ_i^0 is the joint TND inclusion probability of cluster i, and

$$\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) = \hat{\mathbb{P}}_{TND,|\mathcal{S}_{TND,i}|}(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i),$$

which is the cluster-level propensity score, estimated from control data in TND samples (i.e., $\{\mathbf{Y}_i = \mathbf{0}, \boldsymbol{T}_i = \mathbf{1}\}$). Further, similar to the estimation of the cluster propensity score under the SRS sampling in Equation (15), we have model that

$$\begin{aligned} \tau_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) &= \mathbb{P}_{TND}\left(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) \\ &= \int_{-\infty}^{\infty} \prod_{k \in \{\mathcal{S}_{TND,i}, \mathbf{Y}_i = \mathbf{0}\}} \mathbb{P}(V_{ik} = v_{ik} \mid C_{ik}, B_i, \boldsymbol{\alpha})\, \mathbb{P}(B_i = b)\, db, \end{aligned} \tag{17}$$

where $C_{ik} \perp B_i,\ B_i \overset{i.i.d.}{\sim} N\left(0, \sigma_B^2\right)$, and $\mathbb{P}(V_{ik} = 1 \mid C_{ik}, B_i, \boldsymbol{\alpha}) = g^{-1}\left(\boldsymbol{\alpha}^\top C_{ik} + B_i\right)$. To estimate $\boldsymbol{\alpha}$ and σ_B^2, we maximize the log-likelihood for the mixed effects model

$$\sum_{i=1}^{m} l(\boldsymbol{V}_i, \mathbf{C}_i; \boldsymbol{\alpha}, \sigma_B^2) := \log\left(\prod_{i=1}^{m} \mathbb{P}_{TND}\left(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)\right) = \sum_{i=1}^{m} \log\left\{\mathbb{P}_{TND}\left(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)\right\}, \tag{18}$$

with respect to $\boldsymbol{\alpha}$ and σ_B^2, yielding $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}_B^2$.

Lemma 2 *(Identifiability for cluster-level PS) Under cluster-level control exchangeability assumption, i.e.,* $\{\mathbf{Y}_i = \mathbf{0}, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i\} \perp\!\!\!\perp \boldsymbol{V}_i \mid \mathbf{C}_i$, *then the cluster-level propensity score* $\tau_i = \mathbb{P}(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{C}_i)$ *can be identified as* $\tau_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) = \mathbb{P}_{TND}(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)$.

Proof: See Appendix B, Section 1.

Proposition 1 *(Consistent estimation for cluster-level PS) Assume that the cluster-level propensity model* $\mathbb{P}_{TND}\left(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)$ *is correctly specified such as in Equation (17). If the parameters* $\boldsymbol{\alpha}, \sigma_B^2$ *are estimated by the MLE estimators for the mixed effects model in Equation (18), then, under cluster-level control exchangeability assumption, the cluster-level PS* $\tau_i = \mathbb{P}\left(V_{ij} = v, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)$ *is consistently estimated by*

$$\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) = \hat{\mathbb{P}}_{TND,|\mathcal{S}_{TND,i}|}\left(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i, \hat{\boldsymbol{\alpha}}, \hat{\sigma}_B^2\right).$$

In addition, $\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\|_2 = O_p\left(\frac{1}{\sqrt{|\mathcal{Y}_0|}}\right)$*, where* $|\mathcal{Y}_0|$ *is the total number of TND observed controls among all m clusters, and* $|\hat{\sigma}_B^2-\sigma_B^2| = O_p\left(\frac{1}{\sqrt{m}}\right)$*. Jiang (2007)*

Note that, unlike previous cluster-level propensity score (PS) estimators for inverse probability weighting (IPW) under partial interference—such as those proposed in Tchetgen Tchetgen et al. (2009); Liu et al. (2016); Papadogeorgou et al. (2019)—our consistency result for the cluster-level PS estimator holds under both repeated test-negative design sampling within a cluster and repeated treatment assignments. In contrast, the consistency of earlier estimators applies only to repeated treatment assignments, assuming fixed cluster sizes.

Corollary 1 *(Consistent estimation for* $\tau_{TND,i}$*) Let* $\nabla_{\boldsymbol{\alpha},\sigma_B^2}$ *denote the gradient of* $\tau_{TND,i}(\mathbf{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)$ *with respect to the* $p+1$ *parameter vector* $(\boldsymbol{\alpha}^T, \sigma_B^2)^T$*. Suppose there exists a constant* $C_0 > 0$ *such that* $\|\nabla_{\boldsymbol{\alpha},\sigma_B^2}\|_\infty \leq C_0$*. Then, under the convergence rates for* $\hat{\boldsymbol{\alpha}}$ *and* $\hat{\sigma}_B^2$ *established in Proposition 1, it holds with probability* $1 - o_p(1)$ *that*

$$\left|\hat{\tau}_{TND,i}(\mathbf{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) - \tau_{TND,i}(\mathbf{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\right| \leq Cm^{-1/2},$$

for some constant $C > 0$ *and* $i \in [m]$*.*

Proof 1 *From the formula in equation* (17) *and the differentiability of the inverse logistic function, we apply first-order Taylor expansion to obtain*

$$\begin{aligned}
&\left|\hat{\tau}_{TND,i}(\mathbf{V}_i|\mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) - \tau_{TND,i}(\mathbf{V}_i|\mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\right| \\
&\leq \left|\nabla_{\boldsymbol{\alpha},\sigma_B^2}\tau_{TND,i}(\mathbf{V}_i|\mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\left((\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T, \hat{\sigma}_B^2 - \sigma_B^2\right)^T\right| + O(\|\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T, \hat{\sigma}_B^2 - \sigma_B^2\right)^T\|_2^2) \\
&\leq \|\nabla_{\boldsymbol{\alpha},\sigma_B^2}\tau_{TND,i}(\mathbf{V}_i|\mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\|_2 \cdot (\|\hat{\boldsymbol{\alpha}} - \alpha\|_2 + \|\hat{\sigma}_B^2 - \sigma_B^2\|) + O(\|\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T, \hat{\sigma}_B^2 - \sigma_B^2\right)^T\|_2^2) \\
&\leq_{(1)} (p+1)C_0 \cdot 2\frac{1}{\sqrt{m}} + C \cdot 2\frac{1}{m} \leq C\frac{1}{\sqrt{m}}
\end{aligned}$$

where (1) follows from the Cauchy–Schwarz inequality, along with the boundedness of the gradient and the convergence rates established in Proposition 1.

Theorem 1 *(Unbiasedness of IPW Estimators Based on TND Sampling) Let* $\hat{\mu}_i^{\text{ipw,TND}}(v, Q)$ *and* $\hat{\mu}_i^{\text{ipw,TND}}(Q)$ *be the estimators defined in Equations (16) based on TND samples and if the cluster-level propensity score* $\tau_i = \mathbb{P}(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{C}_i)$ *is known, then* $\hat{\mu}_i^{\text{ipw,TND}}(v, Q)$ *and* $\hat{\mu}_i^{\text{ipw,TND}}(Q)$ *are unbiased estimators of* $\mu_i(v, Q)$ *and* $\mu_i(Q)$*, respectively, as defined in Equations (11).*

Proof: See Appendix B.

In addition, replacing $|\mathcal{S}_{TND,i}|$ with its consistent estimator $\hat{n}_{TND} := \sum_{k\in\mathcal{S}_{TND,i}} \frac{\mathbb{I}(V_{ik}=v)Q(\boldsymbol{V}_{i(-k)}|\mathbf{C}_i)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i|\mathbf{Y}_i=\mathbf{0},\mathbf{C}_i)}$ yields a Hájek-type estimator Hájek (1971); Liu et al. (2016). It is well established that the Hájek-type inverse probability weighting (IPW) estimator exhibits superior efficiency relative to the traditional IPW estimator in finite samples Särndal et al. (2003). Details regarding the construction of the Hájek-type IPW estimator are provided in the Appendix C.

Regarding the target direct and spillover effects in Equations (6), we propose the following estimators:

$$\widehat{DE}(Q) = \exp\left(m^{-1}\sum_{i=1}^{m}\left[\log\left\{\frac{\hat{\mu}_i^{\text{ipw,TND}}(1,Q)}{\hat{\mu}_i^{\text{ipw,TND}}(0,Q)}\right\}\right]\right), \text{ and } \widehat{SE}_v(Q',Q) = \exp\left(m^{-1}\sum_{i=1}^{m}\left[\log\left\{\frac{\hat{\mu}_i^{\text{ipw,TND}}(v,Q')}{\hat{\mu}_i^{\text{ipw,TND}}(v,Q)}\right\}\right]\right), \text{ for } v \in \mathcal{V}. \tag{19}$$

It is important to note that $\log\{\widehat{DE}(Q)\}$ is a solution for $\log\{DE(Q)\}$ to the estimating equation

$$\sum_{i=1}^{m}\psi_{v,Q}\left[\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log\{DE(Q)\}\right] = 0,$$

where

$$\psi_{v,Q}\left[\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log\{DE(Q)\}\right] = \log\left\{\frac{\hat{\mu}_i^{\text{ipw,TND}}(1,Q)}{\hat{\mu}_i^{\text{ipw,TND}}(0,Q)}\right\} - \log\{DE(Q)\}. \tag{20}$$

Further, for the score functions system of cluster-level propensity score Equation (18), we have the estimating equation that $\sum_i^m \boldsymbol{\psi}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) = \mathbf{0}$, where

$$\boldsymbol{\psi}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) = \left(\partial l\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)/\partial\boldsymbol{\alpha}^\top, \partial l\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)/\partial\sigma_B^2\right)^\top. \tag{21}$$

Building on these estimating equations, the following theorem addresses the asymptotic normality of $\log(\widehat{DE}(Q))$:

Theorem 2 *(Asymptotic Normality of the Proposed IPW-type Estimators Based on TND Sampling) Assume that the parametric propensity score model is correctly specified and that the regularity assumptions in the Appendix hold. Then, the following results hold:*

1. $\sqrt{m}\left(\log(\widehat{DE}(Q)) - \log(DE(Q))\right)$ *converges in distribution to* $N(0, \Sigma_{log(DE)})$*, as* $m \to \infty$*, where* $\Sigma_{log(DE)}$ *is expressed as*

$$(\mathbb{B}_{21} - \mathbb{F}_{21})\,\mathbb{B}_{11}^{-1}\mathbb{B}_{21}^{\top} - \mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} + F_{22}, \tag{22}$$

where

$$\mathbb{B}_{11}(\boldsymbol{\alpha}, \sigma_B^2) = \mathbb{E}_{F_0}\left\{-\partial\boldsymbol{\psi}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)/\partial(\boldsymbol{\alpha}, \sigma_B^2)^\top\right\}; \mathbb{B}_{21}(\boldsymbol{\alpha}, \sigma_B^2) = \mathbb{E}_{F_0}\left\{-\partial\psi_{v,Q}\left(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(DE(Q))\right)/\partial(\boldsymbol{\alpha}, \sigma_B^2)^\top\right\};$$

$$\mathbb{F}_{21}(\boldsymbol{\alpha}, \sigma_B^2, \log(DE(Q))) = \mathbb{E}_{F_0}\left\{\psi_{v,Q}\left(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(DE(Q))\right)\boldsymbol{\psi}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)^\top\right\};$$

$$F_{22}(\log(DE(Q))) = \mathbb{E}_{F_0}\left\{\psi_{v,Q}\left(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(DE(Q))\right)^2\right\},$$

regarding estimation equations in Equations 20 and 21.

Proof: See Appendix D. Similar properties related to spillover, total, and overall effects are also elaborated in the same appendix.

In addition, by applying the Delta Method with the exponential function such that $\exp[\log(DE(Q))] = DE(Q)$, we have that $\sqrt{m}\left(\widehat{DE}(Q) - DE(Q)\right)$ converges in distribution to $N(0, \Sigma_{DE})$, as $m \to \infty$, where $\Sigma_{DE} = (DE(Q))^2\Sigma_{log(DE)}$. The regularity conditions and detailed steps are verified in Appendix D, ensuring the application of the central limit theorem to our estimators. The asymptotic variances Σ_0 and Σ_{TND} are derived based on the specifics of the IPW-type estimators and the TND sampling framework.

Algorithm 1: Partial Interference Test-Negative Design (PI-TND) direct and indirect VE IPW Algorithm

- **Step 1:** Identify the appropriate clusters among the TND samples based on criteria such as hospital units, zip codes or cities, resulting in $i = 1, 2, ..., M$ clusters.

- **Step 2:** Estimate the cluster-level propensity score model using only control data:

$$\tau_i\left(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i\right) = \mathbb{P}_{TND}\left(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right),$$

  and identify and compute the vaccination stochastic policies of interest (Q), e.g., examining the cluster multiplicative shift (CMS) policy.

- **Step 3:** Construct the population direct, spillover, total, and overall effects estimator for the proposed policy-relevant causal estimand (i.e., Definition (1)), such as, for the VE direct effects:

$$\widehat{VE}^{DE}(Q) = 1 - \widehat{DE}(Q), \text{ and } \widehat{VE}^{SE}(Q) = 1 - \widehat{SE}_v(Q', Q)$$

  where $\widehat{DE}(Q)$ and $\widehat{SE}_v(Q', Q)$ are solution of the estimating equations (20) and (D.11), and approximate $(1-\alpha) \times 100\%$ confidence intervals are

$$\mathrm{CI}_m(VE^{DE}) := 1 - \left[\widehat{DE}(Q) \mp \frac{1}{\sqrt{m}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{(\widehat{DE}(Q))^2\widehat{\Sigma}_{\log(DE)}}\right],$$
$$\mathrm{CI}_m(VE^{SE}) := 1 - \left[\widehat{SE}(Q) \mp \frac{1}{\sqrt{m}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{(\widehat{SE}(Q))^2\widehat{\Sigma}_{\log(SE)}}\right],$$

  where both $\widehat{\Sigma}_{\log(DE)}$ and $\widehat{\Sigma}_{\log(SE)}$ are expressed in the form of $m^{-1}\left\{\left(\hat{\mathbb{B}}_{21} - 2\hat{\mathbb{F}}_{21}\right)\hat{\mathbb{B}}_{11}^{-1}\hat{\mathbb{B}}_{21}^{\top} - \hat{\mathbb{B}}_{21}\hat{\mathbb{B}}_{11}^{-1}\hat{\mathbb{F}}_{21}^{\top} + \hat{F}_{22}\right\}$, with each term detailed in the proof of Theorem (2).

Thus, building on the above results and using cluster-level TND samples, our defined population direct, spillover, total, and overall effects, as well as the corresponding VE metrics, can be consistently estimated. We acknowledge that not only the biased term (the joint TND inclusion probability) but also the estimator $\hat{\tau}_{\mathrm{TND},i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \boldsymbol{C}_i)$ cancels out across clusters in the ratio $\hat{\mu}_i^{\mathrm{ipw,TND}}(1, Q)/\hat{\mu}_i^{\mathrm{ipw,TND}}(0, Q)$. Nevertheless, its estimation remains essential, as it plays a critical role in assessing the overlap between observed and hypothetical treatment assignments, which in turn informs the scale of variance for the estimators (e.g., $\Sigma_{log(DE)}$ in Equation (22)).

Lemma 3 *Suppose there exist constants $\Gamma_1, \Gamma_2 > 0$ such that $\Gamma_1 \leq \rho_i^0 \leq \Gamma_2$ for all $i \in [m]$. Then, with probability $1 - o_p(1)$,*

$$\widehat{DE}(Q) - c(\Gamma_1, \Gamma_2) \leq DE^{HH}(Q) \leq \widehat{DE}(Q) + c(\Gamma_1, \Gamma_2),$$

where

$$c(\Gamma_1, \Gamma_2) = \frac{\Gamma_2^2}{\Gamma_1^2} \cdot \max\left\{\widehat{DE}(Q), \frac{\Gamma_2}{\Gamma_1} \cdot \frac{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1, Q)}{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0, Q)}\right\} \cdot \left(\frac{m\sum_{i=1}^m\left(\frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1, Q)\right)^2}{2\left[\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1, Q)\right]^2} + \frac{m\sum_{i=1}^m\left(\frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0, Q)\right)^2}{2\left[\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0, Q)\right]^2}\right).$$

Note that the quantity $\frac{1}{\rho_i}\hat{\mu}_i^{\text{ipw,TND}}(\cdot, Q)$ is fully identifiable from the TND sample. While the realized values of H_i may vary substantially across clusters, the identical distribution across clusters ensures that the lower and upper bounds Γ_1 and Γ_2 for ρ_i^0 are the same across $i \in [m]$.

Lemma 4 demonstrates that if prior knowledge about Γ_1 and Γ_2 is available, one can obtain an informative range for the magnitude of $DE^{HH}(Q)$. In practice, by considering multiple plausible values for (Γ_1, Γ_2), investigators can assess the sensitivity of $DE^{HH}(Q)$ to variation in these bounds. It is important to emphasize that the purpose of Lemma 4 is to elucidate the relationship between our proposed estimand and the corresponding estimand in the existing literature, rather than to facilitate inference for the estimator of $DE^{HH}(Q)$. Lemma 4 can be extended analogously to the spillover, total, and overall effects—namely $SE^{HH}(Q, Q')$, $TE^{HH}(Q', Q)$, and $OE^{HH}(Q', Q)$—by replacing $\hat{\mu}_i(\cdot)$ with the corresponding terms from the respective estimators of $SE(Q, Q')$, $TE(Q', Q)$, and $OE(Q', Q)$ respectively.

Lemma 4 (Difference between $DE(Q)$ and $DE^{HH}(Q)$) *Assume there exist constants $\Gamma_1, \Gamma_2 > 0$ such that $\Gamma_1 \le \rho_i \le \Gamma_2$ for all $i \in [m]$. Then, there exists a constant $C > 0$ such that, with probability $1 - o_p(1)$,*

$$|DE(Q) - DE^{HH}(Q)| \le C\frac{1}{\sqrt{m}} + \max\left(\widehat{DE}(Q) - \frac{\Gamma_1}{\Gamma_2}\frac{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q)}, \frac{\Gamma_2}{\Gamma_1}\frac{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q)} - \widehat{DE}(Q)\right)$$

The bound on the right-hand side is estimable and facilitates sensitivity analysis through varying the constants (Γ_1, Γ_2). Notably, the estimands $DE(Q)$ and $DE^{HH}(Q)$ do not preserve order under policy shifts in general. That is, for two policies Q and Q', it is generally not the case that $DE(Q) > DE(Q')$ implies $DE^{HH}(Q) > DE^{HH}(Q')$.

Proof 2 *The difference between $DE(Q)$ and $DE^{HH}(Q)$ can be decomposed as*

$$\left|DE(Q) - DE^{HH}(Q)\right| \le \left|DE(Q) - \widehat{DE}(Q)\right| + \left|\widehat{DE}(Q) - \widehat{DE}^{HH}(Q)\right| + \left|\widehat{DE}^{HH}(Q) - DE^{HH}(Q)\right| := (I) + (II) + (III),$$

where $\widehat{DE}^{HH}(Q) = \frac{\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(0,Q)}$ is a partially identifiable plug-in estimator that depends on the unknown parameters ρ_i. For term (I), applying the delta method to $\log DE(Q)$, and using the result in Theorem 2, we obtain that with probability $1 - o_p(1)$,

$$(I) \le Cm^{-1/2}.$$

For term (II), we rely on the following bounds:

$$\frac{\Gamma_1}{\Gamma_2} \cdot \frac{\sum_{i=1}^m \rho_i^{-1}\hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \rho_i^{-1}\hat{\mu}_i^{ipw,TND}(0,Q)} \le \widehat{DE}^{HH}(Q) \le \frac{\Gamma_2}{\Gamma_1} \cdot \frac{\sum_{i=1}^m \rho_i^{-1}\hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \rho_i^{-1}\hat{\mu}_i^{ipw,TND}(0,Q)},$$

where the right-hand side is fully estimable given (Γ_1, Γ_2).

For term (III), we analyze the difference between sample averages and population expectations. Fixing $v \in \{0,1\}$, define the oracle version of the IPW estimator as $\hat{\mu}_i^{ipw,TND,orac}(v,Q)$ by replacing $\hat{\tau}_{TND,i}$ with the true $\tau_{TND,i}$. We then decompose

$$\left|\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i^{ipw,TND}(v,Q) - \mathbb{E}_{F_0}[\mu_i(v,Q)]\right|$$
$$\leq \left|\frac{1}{m}\sum_{i=1}^{m}\left(\hat{\mu}_i^{ipw,TND}(v,Q) - \hat{\mu}_i^{ipw,TND,orac}(v,Q)\right)\right| + \left|\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i^{ipw,TND,orac}(v,Q) - \mathbb{E}_{F_0}[\mu_i(v,Q)]\right| := (A) + (B).$$

For term (A), we have:

$$(A) \leq \frac{1}{m}\sum_{i=1}^{m}|\hat{\tau}_{TND,i}(\mathbf{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) - \tau_{TND,i}(\mathbf{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)| \sum_{k \in S_{TND,i}} \frac{Y_{ik} I(V_{ik} = v) Q(\mathbf{V}_{i(-k)} \mid \mathbf{C}_i)\rho_i^0}{\hat{\tau}_{TND,i}(\mathbf{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) \cdot \tau_{TND,i}(\mathbf{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) \cdot |S_{TND,i}|}$$
$$\leq_{(1)} C \cdot \frac{1}{m}\sum_{i=1}^{m} \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) - \tau_{TND,i}(\mathbf{V}_i \mid = O(m^{-1/2}),$$

where (1) uses Corollary 1 and the boundedness of all terms in the summand. For term (B), we apply Chebyshev's inequality:

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i^{ipw,TND,orac}(v,Q) - \mathbb{E}_{F_0}[\mu_i(v,Q)]\right| \geq \sqrt{\frac{\log m}{m}}\right) \leq \frac{m}{\log m} \cdot \mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i^{ipw,TND,orac}(v,Q)\right)$$
$$\leq_{(1)} \frac{1}{\log m} \cdot \frac{1}{m}\sum_{i=1}^{m}\mathrm{Var}(\hat{\mu}_i^{ipw,TND,orac}(v,Q)) \leq_{(2)} \frac{1}{\log m} \cdot \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}\left[\hat{\mu}_i^{ipw,TND,orac}(v,Q)^2\right] \leq \frac{C}{\log m},$$

where (1) uses the independence of clusters, and (2) follows from boundedness of the outcomes, policies Q and $\tau_{TND,i}$.

Combining bounds for terms (I), (II), and (III) completes the proof.

4 Conclusion

In this work, we developed a methodological framework for estimating vaccine effectiveness under the test-negative design while accounting for partial interference. By defining policy-relevant direct and spillover effects using the geometric mean of risk ratios, we introduced inverse-probability weighted (IPW) estimators that remain valid under TND sampling. Our theoretical results establish the identifiability and large-sample properties of these estimators, ensuring their consistency and asymptomatic normality.

Acknowledgments

The authors gratefully acknowledge a Canadian Institutes of Health Research Project Grant, and CJ is grateful for the postdoctoral fellowship awarded by the StatLab at the Centre de Recherches Mathématiques. DT is supported by a research-career award from the Fonds de recherche du Québec – Santé. MES holds a Canada Research Chair in Causal Inference and Machine Learning in Health Science.

Bibliography

Ali, M., Emch, M., 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. *The Lancet*, 366(9479):44–49.

Anderson, R. M. and May, R. M. (1985). Vaccination and herd immunity to infectious diseases. *Nature*, 318(6044):323–329.

Barkley, B. G., Hudgens, M. G., Clemens, J. D., Ali, M., and Emch, M. E. (2020). Causal inference from observational studies with clustered interference, with application to a cholera vaccine study. *The Annals of Applied Statistics*, 14(3):1432–1448.

Bates, D., Mäechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Boos, D. D., Stefanski, L., Boos, D. D., and Stefanski, L. (2013). M-estimation (estimating equations). *Essential Statistical Inference: Theory and Methods*, pages 297–337.

Cox, D. R. (1958). Planning of experiments.

Dean, N. E., Hogan, J. W., and Schnitzer, M. E. (2021). Covid-19 vaccine effectiveness and the test-negative design. *N Engl J Med*, 385(15):1431–1433.

Feng, S., Fowlkes, A. L., Steffens, A., Finelli, L., and Cowling, B. J. (2017). Assessment of virus interference in a test-negative study of influenza vaccine effectiveness. *Epidemiology (Cambridge, Mass.)*, 28(4):514.

Hájek, J. (1971). Comment on a paper by d. basu. *Foundations of statistical inference*, 236.

Halloran, M. E. and Hudgens, M. G. (2018). Estimating population effects of vaccination using large, routinely collected data. *Statistics in medicine*, 37(2):294–301.

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

Jackson, M. L. and Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. *Vaccine*, 31(17):2165–2168.

Jiang, C., Talbot, D., Carazo, S., and Schnitzer, M. E. (2023). Tnddr: Efficient and doubly robust estimation of covid-19 vaccine effectiveness under the test-negative design. *arXiv preprint arXiv:2310.04578*.

Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. Springer.

Kang, H., Park, C., and Trane, R. (2023). Propensity score modeling: Key challenges when moving beyond the no-interference assumption. *Observational Studies*, 9(1):43–53.

Kennedy, E. H. (2019). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, 114(526):645–656.

Lee, C., Zeng, D., and Hudgens, M. G. (2024). Efficient nonparametric estimation of stochastic policy effects with clustered interference. *Journal of the American Statistical Association*, 102(just-accepted):1–23.

Li, K. Q., Shi, X., Miao, W., and Tchetgen Tchetgen, E. (2023). Double negative control inference in test-negative design studies of vaccine effectiveness. *Journal of the American Statistical Association*, pages 1–12.

Liu, L., Hudgens, M. G., and Becker-Dreps, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika*, 103(4):829–842.

Liu, L., Hudgens, M. G., Saul, B., Clemens, J. D., Ali, M., and Emch, M. E. (2019). Doubly robust estimation in observational studies with partial interference. *Stat*, 8(1):e214.

Ogburn, E. L. and VanderWeele, T. J. (2014). Causal Diagrams for Interference. *Statistical Science*, 29(4):559 – 578.

Papadogeorgou, G., Mealli, F., and Zigler, C. M. (2019). Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics*, 75(3):778–787.

Park, C. and Kang, H. (2022). Efficient semiparametric estimation of network treatment effects under partial interference. *Biometrika*, 109(4):1015–1031.

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(3):731–741.

Randolph, H. E. and Barreiro, L. B. (2020). Herd immunity: understanding covid-19. *Immunity*, 52(5):737–741.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593.

Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American statistical association*, 81(396):961–962.

Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.

Schnitzer, M. E. (2022). Estimands and estimation of covid-19 vaccine effectiveness under the test-negative design: connections to causal inference. *Epidemiology (Cambridge, Mass.)*, 33(3):325.

Schnitzer, M. E., Harel, D., Ho, V., Koushik, A., and Merckx, J. (2021). Identifiability and estimation under the test-negative design with population controls with the goal of identifying risk and preventive factors for sars-cov-2 infection. *Epidemiology*, 32(5):690–697.

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.

Stefanski, L. A. and Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician*, 56(1):29–38.

Sullivan, S. G., Feng, S., and Cowling, B. J. (2014). Influenza vaccine effectiveness: potential of the test-negative design. a systematic review. *Expert review of vaccines*, 13(12):1571–1591.

Sullivan, S. G., Tchetgen Tchetgen, E. J., and Cowling, B. J. (2016). Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. *Am J Epidemiol*, 184(5):345–353.

Tchetgen, E. J. T. and VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75.

Tchetgen Tchetgen, E. J., Robins, J. M., and Rotnitzky, A. (2009). On doubly robust estimation in a semiparametric odds ratio model. *Biometrika*, 97(1):171–180.

Wen, L., Marcus, J. L., and Young, J. G. (2023). Intervention treatment distributions that depend on the observed treatment process and model double robustness in causal survival analysis. *Statistical Methods in Medical Research*, 32(3):509–523.

Wu, C. and Thompson, M. E. (2020). Non-probability survey samples. *Sampling Theory and Practice*, pages 319–331.

Zhang, B., Hudgens, M. G., and Halloran, M. E. (2023). Propensity score in the face of interference: Discussion of rosenbaum and rubin (1983). *Observational studies*, 9(1):125.

SUPPLEMENTARY MATERIAL FOR

Estimating COVID-19 Vaccine Effectiveness with Partial Interference under Test-Negative Design Sampling

Description: This supplementary material includes the proofs of the theoretical results and additional simulation studies. All the code for conducting simulation studies and analyzing real data is accessible in the online repository `https://github.com/CONGJIANG/TNDDR`.

A Proof of Identification:

With a length N_i row vector of ones $\mathbf{1}_{1\times N_i}$ and column vector $\mathbf{Y}_i^*(\boldsymbol{v}_i) = \left(Y_{i1}^*(\boldsymbol{v}_i), ..., Y_{iN_i}^*(\boldsymbol{v}_i)\right)^\top$, we have

$$\begin{aligned}
\mathbb{E}_{F_0}\left(\log[\mu_i(Q)]\right) = \mathbb{E}_{F_0}\left(\log[N_i^{-1}\sum_{j=1}^{N_i}\bar{Y}_{ij}(Q)]\right) &= \mathbb{E}_{F_0}\left(\log\left\{N_i^{-1}\sum_{j=1}^{N_i}\left[\sum_{\boldsymbol{v}_i\in\mathcal{V}(N_i)} Y_{ij}^*(\boldsymbol{v}_i)\, Q\left(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i\right)\right]\right\}\right) \\
&= \mathbb{E}_{F_0}\left(\log\left\{N_i^{-1}\left[\sum_{\boldsymbol{v}_i\in\mathcal{V}(N_i)} \mathbf{1}_{1\times N_i}\mathbf{Y}_i^*(\boldsymbol{v}_i)\, Q\left(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i\right)\right]\right\}\right) \\
&= \mathbb{E}_{F_0}\left(\log\left\{N_i^{-1}\left[\sum_{\boldsymbol{v}_i\in\mathcal{V}(N_i)} \mathbf{1}_{1\times N_i}\mathbb{E}(\mathbf{Y}_i^*(\boldsymbol{v}_i) \mid \mathbf{C}_i) Q\left(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i\right)\right]\right\}\right) \\
&= \mathbb{E}_{F_0}\left(\log\left\{N_i^{-1}\left[\sum_{\boldsymbol{v}_i\in\mathcal{V}(N_i)} \mathbf{1}_{1\times N_i}\mathbb{E}(\mathbf{Y}_i \mid \boldsymbol{V}_i = \boldsymbol{v}_i, \mathbf{C}_i, N_i) Q\left(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i\right)\right]\right\}\right) \\
&= \mathbb{E}_{F_0}\left(\log\left\{N_i^{-1}\sum_{j=1}^{N_i}\left[\sum_{\boldsymbol{v}_i\in\mathcal{V}(N_i)} \mathbb{P}\left(Y_{ij} = 1 \mid \boldsymbol{V}_i = \boldsymbol{v}_i, \mathbf{C}_i, N_i\right) Q\left(\boldsymbol{v}_i \mid \mathbf{C}_i, N_i\right)\right]\right\}\right).
\end{aligned}$$

where the fourth equation follows from iterated expectation, and the fifth equation from conditional exchangeability and causal consistency.

B Proof of Lemma 2 and Theorem 1:

B.1 Proof of Lemma 2:

We want to show that the cluster-level propensity score $\tau_i = \mathbb{P}\left(V_{ij} = v, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)$ can be identified as $\mathbb{P}_{TND}\left(V_{ij} = v, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{Y}_i = 0, \mathbf{C}_i\right)$, under the cluster-level control exchangeability assumption, i.e., $\{\mathbf{Y}_i = 0, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i\} \perp\!\!\!\perp \boldsymbol{V}_i \mid \mathbf{C}_i$.

Cluster-level control exchangeability assumption: being hospitalized for symptoms of *another infection* is independent of vaccination conditional on covariates, that is $\{\mathbf{Y}_i = 0, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i\} \perp\!\!\!\perp \boldsymbol{V}_i \mid \mathbf{C}_i$, or $\{\mathbf{I}_i^{\text{Other}} = 1, \mathbf{W}_i = 1, \mathbf{W}_i = 1\} \perp\!\!\!\perp \boldsymbol{V}_i \mid \mathbf{C}_i$.

$$
\begin{aligned}
&\mathbb{P}_{TND}(V_{ij}=v_{ij}, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{Y}_i=0, \mathbf{C}_i) \\
=&\mathbb{P}(V_{ij}=v_{ij}, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{Y}_i=0, \boldsymbol{H}_i=\boldsymbol{h}_i, \boldsymbol{h}_i\in\mathcal{H}_i, \mathbf{C}_i) \\
=&\frac{\mathbb{P}(\mathbf{Y}_i=0, \boldsymbol{H}_i=\boldsymbol{h}_i, \boldsymbol{h}_i\in\mathcal{H}_i \mid V_{ij}=v_{ij}, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}, \mathbf{C}_i)\mathbb{P}(V_{ij}=v_{ij}, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i)}{\mathbb{P}(\mathbf{Y}_i=0, \boldsymbol{H}_i=\boldsymbol{h}_i, \boldsymbol{h}_i\in\mathcal{H}_i \mid \mathbf{C}_i)} \\
=&\mathbb{P}(V_{ij}=v_{ij}, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i),
\end{aligned}
$$

where the cluster-level control exchangeability assumption implies the last equation based on the fact that $\mathbb{P}(\mathbf{Y}_i=0, \boldsymbol{H}_i=\boldsymbol{h}_i, \boldsymbol{h}_i\in\mathcal{H}_i \mid V_{ij}=v_{ij}, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}, \mathbf{C}_i)=\mathbb{P}(\mathbf{Y}_i=0, \boldsymbol{H}_i=\boldsymbol{h}_i, \boldsymbol{h}_i\in\mathcal{H}_i \mid \mathbf{C}_i)$.

B.2 Proof of Theorem 1

In this section of the appendix, we prove Theorem 1 by focusing primarily on establishing *the TND identification equality*, which involves Lemma 2 and another lemma (i.e., Lemma 5) which is presented in the following proof.

Proof: Under the assumptions that (1) $\mathbb{P}_i(\mathbf{Y}_i=\mathbf{y}_i \mid \boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{H}_i=\boldsymbol{h}_i, \boldsymbol{h}_i\in\mathcal{H}_i)=0$ for $\mathbf{y}_i\in\{\mathbf{0},\mathbf{1}\}$; (2) $0<\mathbb{P}_i\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, \boldsymbol{H}_i=\boldsymbol{h}_i, \boldsymbol{h}_i\in\mathcal{H}_i\right)<1$; (3) $0<\mathbb{P}_i(\mathbf{C}_i=\mathbf{c}_i \mid \boldsymbol{H}_i=\boldsymbol{h}_i, \boldsymbol{h}_i\in\mathcal{H}_i)<1$, we want to show the *TND identification equality*, that is,

$$
\mathbb{E}_{P_{TND,i}}\left[\frac{Y_{ij}\mathbb{I}(V_{ij}=v)Q\left(\boldsymbol{V}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}_{TND}\left(V_{ij}=v, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{Y}_i=\mathbf{0}, \mathbf{C}_i\right)}\right]\rho_i^0=\mathbb{E}_{P_i}\left[\sum_{\boldsymbol{v}_i\in\mathcal{V}(N_i)}\mathbb{P}\left(Y_{ij}=1 \mid \boldsymbol{V}_i=\boldsymbol{v}_i, \mathbf{C}_i, N_i\right)\mathbb{I}(V_{ij}=v)Q\left(\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\right]. \tag{B.1}
$$

First, in Lemma 5, we show that

$$
\mathbb{E}_{P_{TND,i}}\left[\frac{Y_{ij}\mathbb{I}(V_{ij}=v)Q\left(\boldsymbol{V}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}\left(V_{ij}=v, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)}\right]\rho_i^0=\mathbb{E}_{P_i}\left[\sum_{\boldsymbol{v}_i\in\mathcal{V}(N_i)}\mathbb{P}\left(Y_{ij}=1 \mid \boldsymbol{V}_i=\boldsymbol{v}_i, \mathbf{C}_i, N_i\right)\mathbb{I}(V_{ij}=v)Q\left(\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\right],
$$

to demonstrate the unbiasedness of $\hat{\mu}_i^{\text{ipw,TND}}(v, Q)$ when the cluster-level propensity score $\tau_i=\mathbb{P}\left(V_{ij}=v, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)$ is known.

Lemma 5 *If the cluster-level propensity score $\tau_i=\mathbb{P}\left(V_{ij}=v, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)$ is known, then we have*

$$
\mathbb{E}_{P_{TND,i}}\left[\frac{Y_{ij}\mathbb{I}(V_{ij}=v_{ij})Q\left(\boldsymbol{V}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}\left(V_{ij}=v_{ij}, \boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)}\right]\rho_i^0=\mathbb{E}_{P_i}\left[\sum_{\boldsymbol{v}_i\in\mathcal{V}(N_i)}\mathbb{P}\left(Y_{ij}=1 \mid \boldsymbol{V}_i=\boldsymbol{v}_i, \mathbf{C}_i, N_i\right)Q\left(\boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\right].
$$

Proof: Note that for cluster i, we let $\boldsymbol{H}_i=(H_{i1},\ldots,H_{ij},\ldots,H_{iN_i})^\top$ be the TND inclusion indicator vector for cluster i, and denote the TND inclusion probability for cluster i as $\rho_i^0:=\mathbb{P}(\boldsymbol{H}_i=\boldsymbol{h}_i, \boldsymbol{h}_i\in\mathcal{H}_i)$, where $\mathcal{H}_i$ denotes the set of valid inclusion patterns for the TND sample in cluster i, i.e.,

$$
\mathcal{H}_i=\left\{\boldsymbol{h}_i\in\{0,1\}^{N_i} \mid \boldsymbol{h}_i \text{ satisfies all TND inclusion criteria}\right\}.
$$

Let $S_{TND_i} = \{j \in [N_i] : H_{ij} = 1\}$. $P_{TND,i}$ to be the distribution of $(\mathbf{Y}_{S_{TND,i}}(\mathbf{v}_i), \mathbf{V}_{S_{TND,i}}, \mathbf{C}_{S_{TND,i}})$. P_i is the distribution of $(Y_i(\mathbf{v}_i), \mathbf{V}_i, \mathbf{C}_i)$ and $\rho_{ij} := \mathbb{P}(H_{ij} = 1)$. Then for each unit j in cluster i, we have

$$
\begin{aligned}
&\mathbb{E}_{P_{TND,i}}\left[\frac{Y_{ij}\mathbb{I}(V_{ij} = v_{ij})Q\left(\boldsymbol{V}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}\left(V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)}\right]\rho_{ij} \\
&= \int_{\mathscr{C}_i} \sum_{\boldsymbol{v}_{i(-j)}} \frac{Y^*_{ij}(v_{ij}, \boldsymbol{v}_{i(-j)})\mathbb{I}(V_{ij} = v_{ij})Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}\left(V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)}\rho_i^0\mathbb{P}\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i\right)\mathbb{P}(\mathbf{C}_i = \mathbf{c}_i \mid \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i)d\mathbf{C}_i \\
&= \int_{\mathscr{C}_i} \sum_{\boldsymbol{v}_{i(-j)}} \frac{\mathbb{E}\left(Y^*_{ij}(v_{ij}, \boldsymbol{v}_{i(-j)})\mathbb{I}(V_{ij} = v_{ij}) \mid \mathbf{C}_i, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i\right)Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}\left(V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)}\rho_i^0\mathbb{P}\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i\right)\mathbb{P}(\mathbf{C}_i = \mathbf{c}_i \mid \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i)d\mathbf{C}_i \\
&= \int_{\mathscr{C}_i} \sum_{\boldsymbol{v}_{i(-j)}} \frac{\mathbb{P}(Y^*_{ij}(v_{ij}, \boldsymbol{v}_{i(-j)}) = 1, V_{ij} = v_{ij} \mid \mathbf{C}_i, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i)Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}\left(V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)}\mathbb{P}\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i\right)\rho_i^0\mathbb{P}(\mathbf{C}_i = \mathbf{c}_i \mid \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i)d\mathbf{C}_i \\
&= \int_{\mathscr{C}_i} \sum_{\boldsymbol{v}_{i(-j)}} \frac{\mathbb{P}(Y^*_{ij}(v_{ij}, \boldsymbol{v}_{i(-j)}) = 1, V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i)Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}\left(V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)}\rho_i^0\mathbb{P}(\mathbf{C}_i = \mathbf{c}_i \mid \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i)d\mathbf{C}_i \\
&= \int_{\mathscr{C}_i} \sum_{\boldsymbol{v}_{i(-j)}} \frac{\mathbb{P}(Y^*_{ij}(v_{ij}, \boldsymbol{v}_{i(-j)}) = 1, V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i)Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}\left(V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)}\mathbb{P}(\boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i \mid \mathbf{C}_i = \mathbf{c}_i)\mathbb{P}(\mathbf{C}_i = \mathbf{c}_i)d\mathbf{C}_i \\
&= \int_{\mathscr{C}_i} \sum_{\boldsymbol{v}_{i(-j)}} \frac{\mathbb{P}(Y^*_{ij}(v_{ij}, \boldsymbol{v}_{i(-j)}) = 1, V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)}, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i \mid \mathbf{C}_i)Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)}{\mathbb{P}\left(V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i\right)}\mathbb{P}(\mathbf{C}_i = \mathbf{c}_i)d\mathbf{C}_i \\
&= \int_{\mathscr{C}_i} \sum_{\boldsymbol{v}_{i(-j)}} \mathbb{P}(Y^*_{ij}(v_{ij}, \boldsymbol{v}_{i(-j)}) = 1, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i \mid V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)}, \mathbf{C}_i)Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\mathbb{P}(\mathbf{C}_i = \mathbf{c}_i)d\mathbf{C}_i \\
&= \int_{\mathscr{C}_i} \sum_{\boldsymbol{v}_{i(-j)}} \mathbb{P}(Y_{ij} = 1, \boldsymbol{H}_i = \boldsymbol{h}_i, \boldsymbol{h}_i \in \mathcal{H}_i \mid V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)}, \mathbf{C}_i)Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\mathbb{P}(\mathbf{C}_i = \mathbf{c}_i)d\mathbf{C}_i \\
&= \int_{\mathscr{C}_i} \sum_{\boldsymbol{v}_{i(-j)}} \mathbb{P}(Y_{ij} = 1 \mid V_{ij} = v_{ij}, \boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)}, \mathbf{C}_i)Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\mathbb{P}(\mathbf{C}_i = \mathbf{c}_i)d\mathbf{C}_i \\
&= \mathbb{E}_{P_i}\left\{\sum_{\boldsymbol{v}_{i(-j)}} Q\left(\boldsymbol{V}_{i(-j)} = \boldsymbol{v}_{i(-j)} \mid \mathbf{C}_i, N_i\right)\mathbb{P}(Y_{ij} = 1 \mid v_{ij}, \boldsymbol{v}_{ij}, \mathbf{C}_i)\right\} =: \mathbb{E}_{P_i;Q}(Y_{ij}(\mathbf{V}_i)),
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}_{P_{TND,i}}\left[\frac{Y_{ij}\mathbb{I}(V_{ij}=v_{ij})Q\left(\boldsymbol{V}_{i(-j)}\mid \mathbf{C}_i,N_i\right)}{\mathbb{P}\left(V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i\right)}\right]\rho_i^0\\
&=\int_{\mathscr{C}_i}\sum_{\boldsymbol{v}_{i(-j)}}\frac{Y^*_{ij}(v_{ij},\boldsymbol{v}_{i(-j)})\mathbb{I}(V_{ij}=v_{ij})Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)}{\mathbb{P}\left(V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i\right)}\rho_i^0\mathbb{P}\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\right)\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i\mid \boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i)d\mathbf{C}_i\\
&=\int_{\mathscr{C}_i}\sum_{\boldsymbol{v}_{i(-j)}}\frac{\mathbb{E}\left(Y^*_{ij}(v_{ij},\boldsymbol{v}_{i(-j)})\mathbb{I}(V_{ij}=v_{ij})\mid \mathbf{C}_i,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\right)Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)}{\mathbb{P}\left(V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i\right)}\rho_i^0\mathbb{P}\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\right)\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i\mid \boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i)d\mathbf{C}_i\\
&=\int_{\mathscr{C}_i}\sum_{\boldsymbol{v}_{i(-j)}}\frac{\mathbb{P}(Y^*_{ij}(v_{ij},\boldsymbol{v}_{i(-j)})=1,V_{ij}=v_{ij}\mid \mathbf{C}_i,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i)Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)}{\mathbb{P}\left(V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i\right)}\mathbb{P}\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\right)\rho_i^0\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i\mid \boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i)d\mathbf{C}_i\\
&=\int_{\mathscr{C}_i}\sum_{\boldsymbol{v}_{i(-j)}}\frac{\mathbb{P}(Y^*_{ij}(v_{ij},\boldsymbol{v}_{i(-j)})=1,V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i)Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)}{\mathbb{P}\left(V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i\right)}\rho_i^0\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i\mid \boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i)d\mathbf{C}_i\\
&=\int_{\mathscr{C}_i}\sum_{\boldsymbol{v}_{i(-j)}}\frac{\mathbb{P}(Y^*_{ij}(v_{ij},\boldsymbol{v}_{i(-j)})=1,V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i)Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)}{\mathbb{P}\left(V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i\right)}\mathbb{P}(\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\mid \mathbf{C}_i=\mathbf{c}_i)\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i)d\mathbf{C}_i\\
&=\int_{\mathscr{C}_i}\sum_{\boldsymbol{v}_{i(-j)}}\frac{\mathbb{P}(Y^*_{ij}(v_{ij},\boldsymbol{v}_{i(-j)})=1,V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)},\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\mid \mathbf{C}_i)Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)}{\mathbb{P}\left(V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i\right)}\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i)d\mathbf{C}_i\\
&=\int_{\mathscr{C}_i}\sum_{\boldsymbol{v}_{i(-j)}}\mathbb{P}(Y^*_{ij}(v_{ij},\boldsymbol{v}_{i(-j)})=1,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\mid V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)},\mathbf{C}_i)Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i)d\mathbf{C}_i\\
&=\int_{\mathscr{C}_i}\sum_{\boldsymbol{v}_{i(-j)}}\mathbb{P}(Y_{ij}=1,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\mid V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)},\mathbf{C}_i)Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i)d\mathbf{C}_i\\
&=\int_{\mathscr{C}_i}\sum_{\boldsymbol{v}_{i(-j)}}\mathbb{P}(Y_{ij}=1\mid V_{ij}=v_{ij},\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)},\mathbf{C}_i)Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i)d\mathbf{C}_i\\
&=\mathbb{E}_{P_i}\left\{\sum_{\boldsymbol{v}_{i(-j)}}Q\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,N_i\right)\mathbb{P}(Y_{ij}=1\mid v_{ij},\boldsymbol{v}_{ij},\mathbf{C}_i)\right\}=:\mathbb{E}_{P_i;Q}(Y_{ij}(\mathbf{V}_i)),
\end{aligned}
$$

where the first equality follows by the definition of $\mathbb{E}_{P_{TND,i}}$ (w.r.t. distribution $\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i\mid \boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i)$), the law of total expectations (w.r.t. distribution $\mathbb{P}\left(\boldsymbol{V}_{i(-j)}=\boldsymbol{v}_{i(-j)}\mid \mathbf{C}_i,\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\right)$) and the causal consistency assumption that $Y_{ij}=Y^*_{ij}(v_{ij},\boldsymbol{v}_{i(-j)})$; the second equality is by the law of total expectation. The fourth equality follows by the conditional exchangeability or ignorability assumption that $Y^*_i(\boldsymbol{v}_i)\perp \boldsymbol{V}_i\mid \mathbf{C}_i,N_i$; the fifth equality follows by the Bayes rule that $\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i\mid \boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i)=\mathbb{P}(\boldsymbol{H}_i=\boldsymbol{h}_i,\boldsymbol{h}_i\in\mathcal{H}_i\mid \mathbf{C}_i=\mathbf{c}_i)\mathbb{P}(\mathbf{C}_i=\mathbf{c}_i)/\rho_i^0$.

Therefore, we have

$$
\mathbf{E}_{P_{TND,i}}\left[\sum_{j\in\mathcal{S}_{TND,i}}\frac{Y_{ij}\mathbb{I}\left(V_{ij}=v\right)Q\left(\boldsymbol{V}_{i(-j)}\mid \mathbf{C}_i\right)}{\tau_{TND,i}\left(\boldsymbol{V}_i\mid \mathbf{Y}_i=\mathbf{0},\mathbf{C}_i\right)|\mathcal{S}_{TND,i}|}\tilde{\rho}_i^0\right]=_{(1)}\frac{1}{N_i}\mathbb{E}_{P_{TND,i}}\left[\sum_{j\in\mathcal{S}_{TND,i}}\frac{Y_{ij}\mathbb{I}\left(V_{ij}=v\right)Q\left(\boldsymbol{V}_{i(-j)}\mid \mathbf{C}_i\right)}{\tau_{TND,i}\left(\boldsymbol{V}_i\mid \mathbf{Y}_i=\mathbf{0},\mathbf{C}_i\right)}\rho_i^0\right]
$$

$$
=
$$

- If bias sampling is on cluster level, then ρ is identical across different clusters. All asymptotics works. The problem is for bias cancellation. Assume $\bar{Y}_i=\frac{1}{N_i}\sum_{j=1}^{N_i}Y_{ij}$, then the bias cancelling can happen is we assume $\bar{Y}_i\to H_i=1$. But notice that this is a sufficient condition, not the necessary one, meaning it is still possible that $\bar{Y}_i=0$ and $H_i=1$.

- Problem is on how average of TND sample converges to the expectation of its expectation. If using law of large number, the independent should appear in some extent. If assume neighborhood interference, we don't know full network information. If assume the hypothetical and realized treatmtent the same, more plausible on experiment design, but not plausible on the observational studies where the realized one always is not the one we want to learn.

Regarding the inverse probability weighting estimators under TND (i.e., Equation 16):

$$\hat{\mu}_i^{\text{ipw,TND}}(v, Q) = \sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik} \mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\, |\mathcal{S}_{TND,i}|} \rho_i^0, \text{ and } \hat{\mu}_i^{\text{ipw,TND}}(Q) = \sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik} Q(\boldsymbol{V}_i \mid \mathbf{C}_i)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\, |\mathcal{S}_{TND,i}|} \rho_i^0,$$

we have the unbiasedness result in the theorem, which follows directly from the result of Lemma 5.

C Hájek-type IPW esitmators

In this section, we are presenting the Hájek-type IPW estimators for $\mu_i(v, Q)$ and $\mu_i(Q)$. With correctly-specified estimated cluster-level propensity score under TND sampling,

$$\mathbb{E}_{TND}\left[\sum_{k \in \mathcal{S}_{TND,i}} \frac{\mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)}\right] \rho_i^0 = |\mathcal{S}_{TND,i}|,$$

so that replacing $|\mathcal{S}_{TND,i}|$ by its consistent estimator $\hat{n}_{TND} := \sum_{k \in \mathcal{S}_{TND,i}} \frac{\mathbb{I}(V_{ik}=v) Q\left(\boldsymbol{V}_{i(-k)} | \mathbf{C}_i\right)}{\hat{\tau}_{TND,i}\left(\boldsymbol{V}_i | \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i\right)}$ yields the Hájek-type (Hájek (1971); Liu et al. (2016)) IPW esitmator for $\mu_i(v, Q)$:

$$\hat{\mu}_i^{\text{ipw,TND}}(v, Q) = \sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik} \mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\, \hat{n}_{TND}} = \sum_{k \in \mathcal{S}_{TND,i}} \tilde{A}_{ik} Y_{ik}, \tag{C.1}$$

where

$$\tilde{A}_{ik} = \frac{\mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)} / \hat{n}_{TND} = \frac{\mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)} \Big/ \sum_{k \in \mathcal{S}_{TND,i}} \frac{\mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)}$$

and $\sum_{k \in \mathcal{S}_{TND,i}} \tilde{A}_{ik} = 1$. Similarly, the Hájek-type IPW estimator for $\mu_i(Q)$ takes the form

$$\hat{\mu}_i^{\text{ipw,TND}}(Q) = \sum_{k \in \mathcal{S}_{TND,i}} \tilde{B}_{ik} Y_{ik}, \text{ with } \tilde{B}_{ik} = \frac{Q(\boldsymbol{V}_i \mid \mathbf{C}_i)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)} \Big/ \sum_{k \in \mathcal{S}_{TND,i}} \frac{Q(\boldsymbol{V}_i \mid \mathbf{C}_i)}{\hat{\tau}_{TND,i}(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)}.$$

Notably, the Hájek-type IPW estimator demonstrates superior efficiency compared to the traditional IPW estimator when applied to finite samples Särndal et al. (2003).

D Proof of Theorem 2:

This section presents the proof of Theorem 2. We begin by examining the estimation equations derived from both the cluster-level propensity score (referenced in Equation 17) and those used to solve the target effects (e.g., Equation 20).

Firstly, we focus on the estimation equations associated with the cluster-level propensity score:

$$\begin{aligned}\tau_i\left(\boldsymbol{V}_i \mid \mathbf{Y}_i=\mathbf{0}, \mathbf{C}_i\right) &= \mathbb{P}_{TND}\left(\boldsymbol{V}_i=\boldsymbol{v}_i \mid \mathbf{Y}_i=\mathbf{0}, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) \\ &= \int_{-\infty}^{\infty} \prod_{k \in\{\mathcal{S}_{TND,i}, \mathbf{Y}_i=\mathbf{0}\}} \mathbb{P}\left(V_{ik}=v_{ik} \mid C_{ik}, B_i, \boldsymbol{\alpha}\right) \mathbb{P}\left(B_i=b\right) db,\end{aligned} \tag{D.1}$$

where $C_{ik} \perp B_i,\ B_i \overset{i.i.d.}{\sim} N\left(0, \sigma_B^2\right)$, and $\mathbb{P}\left(V_{ik}=1 \mid C_{ik}, B_i, \boldsymbol{\alpha}\right)=g^{-1}\left(\boldsymbol{\alpha}^{\top} C_{ik}+B_i\right)$. Hence, the log-likelihood for the mixed effects model is written as $\log\left(\prod_{i=1}^m \mathbb{P}_{TND}\left(\boldsymbol{V}_i=\boldsymbol{v}_i \mid \mathbf{Y}_i=\mathbf{0}, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)\right)=\sum_{i=1}^m l(\boldsymbol{V}_i, \mathbf{C}_i; \boldsymbol{\alpha}, \sigma_B^2)$, where

$$l(\boldsymbol{V}_i, \mathbf{C}_i; \boldsymbol{\alpha}, \sigma_B^2)=\log\left[\int_{-\infty}^{\infty} \prod_{k \in\{\mathcal{S}_{TND,i}, \mathbf{Y}_i=\mathbf{0}\}} g^{-1}\left(\boldsymbol{\alpha}^{\top} C_{ik}+B_i\right)^{v_j}\left\{1-g^{-1}\left(\boldsymbol{\alpha}^{\top} C_{ik}+B_i\right)\right\}^{(1-v_j)} \mathbb{P}\left(B_i=b\right) db\right]. \tag{D.2}$$

For the score function system that is $\boldsymbol{\psi}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)=\left(\partial l\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) / \partial \boldsymbol{\alpha}^{\top}, \partial l\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) / \partial \sigma_B^2\right)^{\top}$ of Equation 21, we denote each element in $\boldsymbol{\psi}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)$ as

$$\begin{gathered}\sum_i^m \psi_{cr}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)=0, \text{ where } \psi_{cr}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)=\partial l\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) / \partial \alpha_r,\ \alpha_r \text{ is the element } r \text{ of } \boldsymbol{\alpha}, \text{ for } r=1, \ldots, p; \\ \sum_i^m \psi_{\sigma_B^2}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)=0, \text{ where } \psi_{\sigma_B^2}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)=\partial l\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) / \partial \sigma_B^2,\end{gathered} \tag{D.3}$$

and $\hat{\boldsymbol{\alpha}}, \hat{\sigma}_B^2$ that maximize the log-likelihood are the solutions.

As mentioned in the main context, it is important to note that $\log(\widehat{DE}(Q))$ is a solution for $\log(DE(Q))$ to the estimating equation

$$\sum_i^m \psi_{v,Q}\left(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(DE(Q))\right)=0,$$

where

$$\psi_{v,Q}\left(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(DE(Q))\right)=\log\left(\frac{\mu_i^{\text{ipw,TND}}(1, Q)}{\mu_i^{\text{ipw,TND}}(0, Q)}\right)-\log[DE(Q)]. \tag{D.4}$$

Therefore, with parameters $\boldsymbol{\theta}_0=(\boldsymbol{\alpha}, \sigma_B^2, \log(DE(Q)))$, we have the estimator, $\hat{\boldsymbol{\theta}}=\left(\hat{\boldsymbol{\alpha}}, \hat{\sigma}_B^2, \log(\widehat{DE}(Q))\right)$ which is a solution to the estimating equation systems $\sum_i \boldsymbol{\psi}\left(\mathbf{O}_i, \boldsymbol{\theta}\right)=\mathbf{0}$, with $\mathbf{O}_i=(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i)$, where

$$\boldsymbol{\psi}\left(\mathbf{O}_i, \boldsymbol{\theta}_0\right)=\left(\boldsymbol{\psi}^{\top}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right), \psi_{v,Q}\left(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(DE(Q))\right)\right)^{\top}=\begin{pmatrix} \psi_{c1}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) \\ \vdots \\ \psi_{cp}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) \\ \psi_{\sigma_B^2}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) \\ \psi_{v,Q}\left(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(DE(Q))\right) \end{pmatrix}. \tag{D.5}$$

Hence, according to M-estimation theory (Boos et al. (2013)), $\hat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}_0$ in probability, and the difference $\sqrt{m}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$, where the variance matrix $\boldsymbol{\Sigma}$ adopts a sandwich form of

$$\boldsymbol{\Sigma}=\mathbb{B}\left(\boldsymbol{\theta}_0\right)^{-1} \mathbb{F}\left(\boldsymbol{\theta}_0\right)\left\{\mathbb{B}\left(\boldsymbol{\theta}_0\right)^{-1}\right\}^{\top},$$

where

$$\mathbb{B}(\boldsymbol{\theta}_0) = \mathbb{E}_{F_0}\left\{-\boldsymbol{\psi}'(\mathbf{O}_i, \boldsymbol{\theta}_0)\right\}, \text{with } \boldsymbol{\psi}'(\mathbf{O}_i, \boldsymbol{\theta}) = \partial\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}) / \partial\boldsymbol{\theta}^\top,$$

$$\mathbb{F}(\boldsymbol{\theta}_0) = \mathbb{E}_{F_0}\left\{\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}_0)\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}_0)^\top\right\}.$$

Note that $\boldsymbol{\theta}_0$ is the true parameter value defined Stefanski and Boos (2002) by

$$\mathbb{E}_{F_0}\left[\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}_0)\right] = \int \boldsymbol{\psi}(\mathbf{o}, \boldsymbol{\theta}_0)\, dF_0(\mathbf{o}) = \mathbf{0},$$

where here F_0 denotes the cumulative distribution function of $\mathbf{O}_i$.

In the subsequent sections, we present the specific formulation for the asymptotic variance of $\hat{\psi}_{v,Q}(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(DE(Q))) = \log(\widehat{DE}(Q))$, explicitly expressing $\mathbb{B}(\boldsymbol{\theta}_0)$ and $\mathbb{F}(\boldsymbol{\theta}_0)$ in terms of $\boldsymbol{\psi}'(\mathbf{O}_i, \boldsymbol{\theta})$ and $\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta})$. First, $\boldsymbol{\psi}'(\mathbf{O}_i, \boldsymbol{\theta})$ is the $(p+2) \times (p+2)$ matrix:

$$\begin{bmatrix} \partial\psi_{c1}/\partial\alpha_1 & \cdots & \partial\psi_{c1}/\partial\alpha_p & \partial\psi_{c1}/\partial\sigma_B^2 & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ \partial\psi_{cp}/\partial\alpha_1 & \cdots & \partial\psi_{cp}/\partial\alpha_p & \partial\psi_{cp}/\partial\sigma_B^2 & 0 \\ \partial\psi_{\sigma_B^2}/\partial\alpha_1 & \cdots & \partial\psi_{\sigma_B^2}/\partial\alpha_p & \partial\psi_{\sigma_B^2}/\partial\sigma_B^2 & 0 \\ \partial\psi_{v,Q}/\partial\alpha_1 & \cdots & \partial\psi_{v,Q}/\partial\alpha_p & \partial\psi_{v,Q}/\partial\sigma_B^2 & -1 \end{bmatrix}. \tag{D.6}$$

Then, using block matrix notation, $\mathbb{B}(\boldsymbol{\theta}_0)$ can be written as

$$\mathbb{B}(\boldsymbol{\theta}_0) = \mathbb{E}_{F_0}\left\{-\boldsymbol{\psi}'(\mathbf{O}_i, \boldsymbol{\theta}_0)\right\} = -\mathbb{E}_{F_0}\left[\begin{array}{cccc|c} \partial\psi_{c1}/\partial\alpha_1 & \cdots & \partial\psi_{c1}/\partial\alpha_p & \partial\psi_{c1}/\partial\sigma_B^2 & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ \partial\psi_{cp}/\partial\alpha_1 & \cdots & \partial\psi_{cp}/\partial\alpha_p & \partial\psi_{cp}/\partial\sigma_B^2 & 0 \\ \partial\psi_{\sigma_B^2}/\partial\alpha_1 & \cdots & \partial\psi_{\sigma_B^2}/\partial\alpha_p & \partial\psi_{\sigma_B^2}/\partial\sigma_B^2 & 0 \\ \hline \partial\psi_{v,Q}/\partial\alpha_1 & \cdots & \partial\psi_{v,Q}/\partial\alpha_p & \partial\psi_{v,Q}/\partial\sigma_B^2 & -1 \end{array}\right] = \begin{bmatrix} \mathbb{B}_{11} & \mathbf{0} \\ \mathbb{B}_{21} & 1 \end{bmatrix}, \tag{D.7}$$

where $\mathbb{B}_{11}$ is a $(p+1) \times (p+1)$ information matrix of the cluster PS, and $\mathbb{B}_{21}$ is a $1 \times (p+1)$ vector. $\mathbb{F}(\boldsymbol{\theta}_0)$ can be written as

$$\mathbb{F}(\boldsymbol{\theta}_0) = \mathbb{E}_{F_0}\left\{\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}_0)\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}_0)^\top\right\} = \mathbb{E}_{F_0}\left[\begin{array}{cccc|c} \psi_{c1}^2 & \cdots & \psi_{c1}\psi_{cp} & \psi_{c1}\psi_{\sigma_B^2} & \psi_{c1}\psi_{v,Q} \\ \vdots & & \vdots & \vdots & \vdots \\ \psi_{cp}\psi_{c1} & \cdots & \psi_{cp}^2 & \psi_{cp}\psi_{\sigma_B^2} & \psi_{cp}\psi_{v,Q} \\ \psi_{\sigma_B^2}\psi_{c1} & \cdots & \psi_{\sigma_B^2}\psi_{cp} & \psi_{\sigma_B^2}^2 & \psi_{\sigma_B^2}\psi_{v,Q} \\ \hline \psi_{v,Q}\psi_{c1} & \cdots & \psi_{v,Q}\psi_{cp} & \psi_{v,Q}\psi_{\sigma_B^2} & \psi_{v,Q}^2 \end{array}\right] = \begin{bmatrix} \mathbb{F}_{11} & \mathbb{F}_{21}^\top \\ \mathbb{F}_{21} & F_{22} \end{bmatrix}, \tag{D.8}$$

where $\mathbb{F}_{11}$ is a $(p+1) \times (p+1)$ matrix, $\mathbb{F}_{21}$ is a $1 \times (p+1)$ vector, and F_{22} is a scalar. In addition,

$$\mathbb{B}(\boldsymbol{\theta}_0)^{-1} = \begin{bmatrix} \mathbb{B}_{11}^{-1} & \mathbf{0} \\ -\mathbb{B}_{21}\mathbb{B}_{11}^{-1} & 1 \end{bmatrix}, \tag{D.9}$$

and the variance matrix $\mathbf{\Sigma}$ can be written as

$$
\begin{aligned}
\mathbf{\Sigma} &= \mathbb{B}\left(\boldsymbol{\theta}_0\right)^{-1} \mathbb{F}\left(\boldsymbol{\theta}_0\right) \left\{\mathbb{B}\left(\boldsymbol{\theta}_0\right)^{-1}\right\}^{\top} \\
&= \begin{bmatrix} \mathbb{B}_{11}^{-1} & \mathbf{0} \\ -\mathbb{B}_{21}\mathbb{B}_{11}^{-1} & 1 \end{bmatrix} \begin{bmatrix} \mathbb{F}_{11} & \mathbb{F}_{21}^{\top} \\ \mathbb{F}_{21} & F_{22} \end{bmatrix} \begin{bmatrix} \mathbb{B}_{11}^{-1} & \mathbf{0} \\ -\mathbb{B}_{21}\mathbb{B}_{11}^{-1} & 1 \end{bmatrix}^{\top} \\
&= \begin{bmatrix} \mathbb{B}_{11}^{-1}\mathbb{F}_{11} & \mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} \\ -\mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{11} + \mathbb{F}_{21} & -\mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} + F_{22} \end{bmatrix} \begin{bmatrix} (\mathbb{B}_{11}^{-1})^{\top} & (-\mathbb{B}_{21}\mathbb{B}_{11}^{-1})^{\top} \\ \mathbf{0} & 1 \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{B}_{11}^{-1}\mathbb{F}_{11}(\mathbb{B}_{11}^{-1})^{\top} & \mathbb{B}_{11}^{-1}\mathbb{F}_{11}(-\mathbb{B}_{21}\mathbb{B}_{11}^{-1})^{\top} + \mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} \\ (-\mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{11} + \mathbb{F}_{21})(\mathbb{B}_{11}^{-1})^{\top} & (-\mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{11} + \mathbb{F}_{21})(-\mathbb{B}_{21}\mathbb{B}_{11}^{-1})^{\top} + (-\mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} + F_{22}) \end{bmatrix}.
\end{aligned}
$$

As $\mathbb{B}_{11}$ and $\mathbb{F}_{11}$ are associated with the score equations of the log-likelihood function in the mixed effects model, it can be inferred that $\mathbb{B}_{11}$ is equal to $\mathbb{F}_{11}$ (Stefanski and Boos (2002)). Then, we can simplify the matrix $\mathbf{\Sigma}$:

$$
\begin{aligned}
\mathbf{\Sigma} &= \begin{bmatrix} \mathbb{B}_{11}^{-1}\mathbb{F}_{11}(\mathbb{B}_{11}^{-1})^{\top} & \mathbb{B}_{11}^{-1}\mathbb{F}_{11}(-\mathbb{B}_{21}\mathbb{B}_{11}^{-1})^{\top} + \mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} \\ (-\mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{11} + \mathbb{F}_{21})(\mathbb{B}_{11}^{-1})^{\top} & (-\mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{11} + \mathbb{F}_{21})(-\mathbb{B}_{21}\mathbb{B}_{11}^{-1})^{\top} + (-\mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} + F_{22}) \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{B}_{11}^{-1} & (-\mathbb{B}_{21}\mathbb{B}_{11}^{-1})^{\top} + \mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} \\ (-\mathbb{B}_{21} + \mathbb{F}_{21})\,\mathbb{B}_{11}^{-1} & (\mathbb{B}_{21} - \mathbb{F}_{21})\,\mathbb{B}_{11}^{-1}\mathbb{B}_{21}^{\top} - \mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} + F_{22} \end{bmatrix}.
\end{aligned}
$$

Therefore, we conclude that $\sqrt{m}\left(\log(\widehat{DE}(Q)) - \log(DE(Q))\right)$ converges in distribution to $N(0, \Sigma_{log(DE)})$, as $m \to \infty$, where $\Sigma_{log(DE)}$ is expressed as

$$\Sigma_{log(DE)} = (\mathbb{B}_{21} - \mathbb{F}_{21})\,\mathbb{B}_{11}^{-1}\mathbb{B}_{21}^{\top} - \mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} + F_{22}.$$

Further, applying the delta method with the exponential function such that $\exp[\log(DE(Q))] = DE(Q)$, we can conclude that $\sqrt{m}\left(\widehat{DE}(Q) - DE(Q)\right)$ converges in distribution to $N(0, \Sigma_{DE})$, as $m \to \infty$, where $\Sigma_{DE} = (DE(Q))^2\Sigma_{log(DE)}$.

By replacing $\mathbb{B}(\hat{\boldsymbol{\theta}})$ and $\mathbb{F}(\hat{\boldsymbol{\theta}})$ with their empirical estimators, we obtain:

$$\mathbb{B}_m(\hat{\boldsymbol{\theta}}) = m^{-1}\sum_i \left\{-\boldsymbol{\psi}'\left(\mathbf{O}_i, \hat{\boldsymbol{\theta}}\right)\right\}, \text{ and } \mathbb{F}_m(\hat{\boldsymbol{\theta}}) = m^{-1}\sum_i \left\{\boldsymbol{\psi}\left(\mathbf{O}_i, \hat{\boldsymbol{\theta}}\right)\boldsymbol{\psi}\left(\mathbf{O}_i, \hat{\boldsymbol{\theta}}\right)^{\top}\right\}.$$

Thus, the empirical sandwich variance estimator is given by: $\mathbf{\Sigma}_m = \mathbb{B}_m(\hat{\boldsymbol{\theta}})^{-1}\mathbb{F}_m(\hat{\boldsymbol{\theta}})\left\{\mathbb{B}_m(\hat{\boldsymbol{\theta}})^{-1}\right\}^{\top}$. The element in the bottom-right position, i.e., $(p+2, p+2)$ of the matrix $\mathbf{\Sigma}_m$, when multiplied by m^{-1}, provides the estimated asymptotic variance for $\log(\widehat{DE}(Q))$ and thus $\widehat{DE}(Q)$.

In conclusion, this implies the estimation of the asymptotic variance of $\widehat{DE}(Q)$ by

$$\frac{\widehat{DE}(Q)^2}{m}\left\{\left(\hat{\mathbb{B}}_{21} - \hat{\mathbb{F}}_{21}\right)\hat{\mathbb{B}}_{11}^{-1}\hat{\mathbb{B}}_{21}^{\top} - \hat{\mathbb{B}}_{21}\hat{\mathbb{B}}_{11}^{-1}\hat{\mathbb{F}}_{21}^{\top} + \hat{F}_{22}\right\}, \tag{D.10}$$

where $\hat{\mathbb{B}}_{21}$ represents the first $p+1$ terms in the bottom row of $\mathbb{B}_m(\hat{\boldsymbol{\theta}})$, and $\hat{\mathbb{F}}_{21}$, $\hat{\mathbb{B}}_{11}$, and F_{22} are the corresponding sub matrices or elements of $\mathbb{F}_m(\hat{\boldsymbol{\theta}})$.

Remark 1: For further computation of variance of DE, we note that the explicit form of $\mathbb{B}_{21}$ and $\mathbb{F}_{21}$ can be derived as follows. First, regarding $\mathbb{B}_{21}$, we have

$$\mathbb{E}_{F_0}\left\{\frac{\partial}{\partial\alpha_r}\psi_{v,Q}\left[\mathbf{Y}_i,\boldsymbol{V}_i,\mathbf{C}_i,\log(DE(Q));\log(DE),\boldsymbol{\alpha},\sigma_B^2\right]\right\}=\mathbb{E}_{F_0}\left\{\frac{\partial}{\partial\alpha_r}\log\left[\mu_i^{\text{ipw,TND}}(1,Q)\right]-\frac{\partial}{\partial\alpha_r}\log\left[\mu_i^{\text{ipw,TND}}(0,Q)\right]\right\},$$

where for $v=\{0,1\}$,

$$\begin{aligned}
\frac{\partial}{\partial\alpha_r}\log\left[\mu_i^{\text{ipw,TND}}(v,Q)\right] &= \left[\mu_i^{\text{ipw,TND}}(v,Q)\right]^{-1}\frac{\partial}{\partial\alpha_r}\left[\mu_i^{\text{ipw,TND}}(v,Q)\right]\\
&= \left[\mu_i^{\text{ipw,TND}}(v,Q)\right]^{-1}\left(\sum_{k\in\mathcal{S}_{TND,i}}\frac{Y_{ik}\mathbb{I}\left(V_{ik}=v\right)Q\left(\boldsymbol{V}_{i(-k)}\mid\mathbf{C}_i\right)}{|\mathcal{S}_{TND,i}|}\rho_i^0\right)\left(\frac{\partial}{\partial\alpha_r}\frac{1}{\tau_i\left(\boldsymbol{V}_i\mid\mathbf{Y}_i=\mathbf{0},\mathbf{C}_i;\boldsymbol{\alpha},\sigma_B^2\right)}\right)\\
&= -\left[\mu_i^{\text{ipw,TND}}(v,Q)\right]^{-1}\left(\sum_{k\in\mathcal{S}_{TND,i}}\frac{Y_{ik}\mathbb{I}\left(V_{ik}=v\right)Q\left(\boldsymbol{V}_{i(-k)}\mid\mathbf{C}_i\right)}{|\mathcal{S}_{TND,i}|}\rho_i^0\right)\left(\frac{\frac{\partial}{\partial\alpha_r}\log\tau_i\left(\boldsymbol{V}_i\mid\mathbf{Y}_i=\mathbf{0},\mathbf{C}_i;\boldsymbol{\alpha},\sigma_B^2\right)}{\tau_i\left(\boldsymbol{V}_i\mid\mathbf{Y}_i=\mathbf{0},\mathbf{C}_i;\boldsymbol{\alpha},\sigma_B^2\right)}\right)\\
&= -\left[\mu_i^{\text{ipw,TND}}(v,Q)\right]^{-1}\left(\sum_{k\in\mathcal{S}_{TND,i}}\frac{Y_{ik}\mathbb{I}\left(V_{ik}=v\right)Q\left(\boldsymbol{V}_{i(-k)}\mid\mathbf{C}_i\right)}{\tau_i\left(\boldsymbol{V}_i\mid\mathbf{Y}_i=\mathbf{0},\mathbf{C}_i\right)|\mathcal{S}_{TND,i}|}\rho_i^0\right)\psi_{cr}\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)\\
&= -\psi_{cr}\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)\\
&= -\partial l\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)/\partial\alpha_r,
\end{aligned}$$

where $\psi_{cr}\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)=\partial l\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)/\partial\alpha_r$, the score function of the cluster PS, related to the PS log-likelihood for the mixed effects model (D.2). Similarly, a corresponding process can be derived for the parameter σ_B^2 to complete the derivation of the $\partial\psi_{v,Q}/\partial\sigma_B^2$ element in $\mathbb{B}_{21}$:

$$\frac{\partial}{\partial\sigma_B^2}\log\left[\mu_i^{\text{ipw,TND}}(v,Q)\right]=\left[\mu_i^{\text{ipw,TND}}(v,Q)\right]^{-1}\frac{\partial}{\partial\sigma_B^2}\left[\mu_i^{\text{ipw,TND}}(v,Q)\right]=-\psi_{\sigma_B^2}\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)=-\partial l\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)/\partial\sigma_B^2.$$

Then, for $\mathbb{F}_{21}$, with the score function of the cluster PS, $\psi_{cr}\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)$, we have

$$\mathbb{E}_{F_0}\left\{\psi_{v,Q}\left[\mathbf{Y}_i,\boldsymbol{V}_i,\mathbf{C}_i,\log(DE(Q))\right]\psi_{cr}\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)\right\}=\mathbb{E}_{F_0}\left\{\left(\log\left(\frac{\mu_i^{\text{ipw,TND}}(1,Q)}{\mu_i^{\text{ipw,TND}}(0,Q)}\right)-\log[DE(Q)]\right)\psi_{cr}\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)\right\}.$$

Remark 2: The estimation of large-sample variances for indirect, total, and overall effect estimators follows a similar approach, with different estimation equations that:

$$\sum_i^m\psi_{v,Q}\left(\mathbf{Y}_i,\boldsymbol{V}_i,\mathbf{C}_i,SE_v(Q',Q)\right)=0,$$

where

$$\psi_{v,Q}\left(\mathbf{Y}_i,\boldsymbol{V}_i,\mathbf{C}_i,SE_v(Q',Q)\right)=\log\left(\frac{\mu_i^{\text{ipw,TND}}(v,Q')}{\mu_i^{\text{ipw,TND}}(v,Q)}\right)-\log[SE_v(Q',Q)]. \tag{D.11}$$

For the total and overall effect, we have that

$$\psi_{v,Q}\left(\mathbf{Y}_i,\boldsymbol{V}_i,\mathbf{C}_i,TE(Q',Q)\right)=\log\left(\frac{\mu_i^{\text{ipw,TND}}(1,Q')}{\mu_i^{\text{ipw,TND}}(0,Q)}\right)-\log[TE(Q',Q)];$$

$$\psi_{v,Q}\left(\mathbf{Y}_i,\boldsymbol{V}_i,\mathbf{C}_i,OE(Q',Q)\right)=\log\left(\frac{\mu_i^{\text{ipw,TND}}(Q')}{\mu_i^{\text{ipw,TND}}(Q)}\right)-\log[OE(Q',Q)],$$

where, again,

$$\mu_i^{\text{ipw,TND}}(v, Q) = \sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik}\mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\hat{\tau}_i(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\,|\mathcal{S}_{TND,i}|}\rho_i^0 \quad (v \in \mathcal{V}), \text{ and } \mu_i^{\text{ipw,TND}}(Q) = \sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik} Q(\boldsymbol{V}_i \mid \mathbf{C}_i)}{\hat{\tau}_i(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\,|\mathcal{S}_{TND,i}|}\rho_i^0$$

and

$$\begin{aligned}\tau_i(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i) &= \mathbb{P}_{TND}\left(\boldsymbol{V}_i = \boldsymbol{v}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) \\ &= \int_{-\infty}^{\infty} \prod_{k \in \{\mathcal{S}_{TND,i}, \mathbf{Y}_i = \mathbf{0}\}} \mathbb{P}(V_{ik} = v_{ik} \mid C_{ik}, B_i, \boldsymbol{\alpha})\,\mathbb{P}(B_i = b)\,db.\end{aligned}$$

Specifically, this involves substituting Equation (D.4) with the target estimation equation (e.g., for SE, $\psi_{v,Q}(\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, SE_v(Q', Q))$) in the estimation equations from the system (D.5) and subsequently computing the corresponding $\mathbb{B}(\boldsymbol{\theta}_0)$ and $\mathbb{F}(\boldsymbol{\theta}_0)$ based on $\boldsymbol{\psi}'(\mathbf{O}_i, \boldsymbol{\theta})$ and $\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta})$. For instance, for the computation of IE, we derive the explicit form of $\mathbb{B}_{21}$ and $\mathbb{F}_{21}$. First, regarding $\mathbb{B}_{21}$, we have

$$\mathbb{E}_{F_0}\left\{\frac{\partial}{\partial \alpha_r}\psi_{v,Q}\left[\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(SE_v(Q', Q)); \log(SE), \boldsymbol{\alpha}, \sigma_B^2\right]\right\} = \mathbb{E}_{F_0}\left\{\frac{\partial}{\partial \alpha_r}\log\left[\mu_i^{\text{ipw,TND}}(v, Q')\right] - \frac{\partial}{\partial \alpha_r}\log\left[\mu_i^{\text{ipw,TND}}(v, Q)\right]\right\},$$

where for $v = \{0, 1\}$,

$$\begin{aligned}\frac{\partial}{\partial \alpha_r}\log\left[\mu_i^{\text{ipw,TND}}(v, Q)\right] &= \left[\mu_i^{\text{ipw,TND}}(v, Q)\right]^{-1}\frac{\partial}{\partial \alpha_r}\left[\mu_i^{\text{ipw,TND}}(v, Q)\right] \\ &= \left[\mu_i^{\text{ipw,TND}}(v, Q)\right]^{-1}\left(\sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik}\mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{|\mathcal{S}_{TND,i}|}\rho_i^0\right)\left(\frac{\partial}{\partial \alpha_r}\frac{1}{\tau_i\left(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i; \boldsymbol{\alpha}, \sigma_B^2\right)}\right) \\ &= -\left[\mu_i^{\text{ipw,TND}}(v, Q)\right]^{-1}\left(\sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik}\mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{|\mathcal{S}_{TND,i}|}\rho_i^0\right)\left(\frac{\frac{\partial}{\partial \alpha_r}\log\tau_i\left(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i; \boldsymbol{\alpha}, \sigma_B^2\right)}{\tau_i\left(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i; \boldsymbol{\alpha}, \sigma_B^2\right)}\right) \\ &= -\left[\mu_i^{\text{ipw,TND}}(v, Q)\right]^{-1}\left(\sum_{k \in \mathcal{S}_{TND,i}} \frac{Y_{ik}\mathbb{I}(V_{ik} = v)\, Q\left(\boldsymbol{V}_{i(-k)} \mid \mathbf{C}_i\right)}{\tau_i(\boldsymbol{V}_i \mid \mathbf{Y}_i = \mathbf{0}, \mathbf{C}_i)\,|\mathcal{S}_{TND,i}|}\rho_i^0\right)\psi_{cr}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) \\ &= -\psi_{cr}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)\end{aligned}$$

where $\psi_{cr}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right) = \partial l\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)/\partial \alpha_r$, the score function of the cluster PS, related to the PS log-likelihood Equation D.2. Similarly, a corresponding process can be derived for the parameter σ_B^2 to complete the derivation of the $\partial\psi_{v,Q}/\partial\sigma_B^2$ element in $\mathbb{B}_{21}$.

Then, for $\mathbb{F}_{21}$, with the score function of the cluster PS $\psi_{cr}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)$, we have

$$\mathbb{E}_{F_0}\left\{\psi_{v,Q}\left[\mathbf{Y}_i, \boldsymbol{V}_i, \mathbf{C}_i, \log(SE_v(Q', Q))\right]\psi_{cr}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)\right\} = \mathbb{E}_{F_0}\left\{\left(\log\left(\frac{\mu_i^{\text{ipw,TND}}(v, Q')}{\mu_i^{\text{ipw,TND}}(v, Q)}\right) - \log[SE_v(Q', Q)]\right)\psi_{cr}\left(\boldsymbol{V}_i, \mathbf{C}_i, \boldsymbol{\alpha}, \sigma_B^2\right)\right\}.$$

1. $\sqrt{m}\left(\log(\widehat{SE}(Q', Q)) - \log(SE(Q', Q))\right)$ converges in distribution to $N(0, \Sigma_{log(SE)})$, as $m \to \infty$, where $\Sigma_{log(SE)}$ is expressed as

$$(\mathbb{B}_{21} - \mathbb{F}_{21})\,\mathbb{B}_{11}^{-1}\mathbb{B}_{21}^{\top} - \mathbb{B}_{21}\mathbb{B}_{11}^{-1}\mathbb{F}_{21}^{\top} + F_{22},$$

where

$$\mathbb{B}_{11}(\boldsymbol{\alpha},\sigma_B^2)=\mathbb{E}_{F_0}\left\{-\partial\boldsymbol{\psi}\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)/\partial(\boldsymbol{\alpha},\sigma_B^2)^\top\right\};\mathbb{B}_{21}(\boldsymbol{\alpha},\sigma_B^2)=\mathbb{E}_{F_0}\left\{-\partial\psi_{v,Q}\left(\mathbf{Y}_i,\boldsymbol{V}_i,\mathbf{C}_i,\log(SE(Q',Q))\right)/\partial(\boldsymbol{\alpha},\sigma_B^2)^\top\right\};$$

$$\mathbb{F}_{21}(\boldsymbol{\alpha},\sigma_B^2,\log(SE(Q',Q)))=\mathbb{E}_{F_0}\left\{\psi_{v,Q}\left(\mathbf{Y}_i,\boldsymbol{V}_i,\mathbf{C}_i,\log(SE(Q',Q))\right)\boldsymbol{\psi}\left(\boldsymbol{V}_i,\mathbf{C}_i,\boldsymbol{\alpha},\sigma_B^2\right)^\top\right\};$$

$$F_{22}(\log(SE(Q',Q)))=\mathbb{E}_{F_0}\left\{\psi_{v,Q}\left(\mathbf{Y}_i,\boldsymbol{V}_i,\mathbf{C}_i,\log(SE(Q',Q))\right)^2\right\},$$

regarding estimation equations in Equations D.11 and 21.

E Proof of Lemmas 1 and 4

Proof 3 (Proof of Lemma 1) *Let function $f(a,b):=\log(\frac{a}{b})$. By taking Taylor expansion on $(\mathbb{E}(a),\mathbb{E}(b))$ up to the second order of $f(a,b)$, we have*

$$f(a,b)=\log\left(\frac{\mathbb{E}(a)}{\mathbb{E}(b)}\right)+\frac{1}{\mathbb{E}(a)}(a-\mathbb{E}(a))-\frac{1}{\mathbb{E}(b)}(b-\mathbb{E}(b))-\frac{1}{2[\mathbb{E}(a)]^2}(a-\mathbb{E}(a))^2+\frac{1}{2[\mathbb{E}(b)]^2}(b-\mathbb{E}(b))^2+o((a-\mathbb{E}(a))^2+(b-\mathbb{E}(b))^2)$$

Therefore, $\mathbb{E}(f(a,b))$ satisfies

$$\begin{aligned}\mathbb{E}(f(a,b))&=\log\left(\frac{\mathbb{E}(a)}{\mathbb{E}(b)}\right)-\frac{1}{2[\mathbb{E}(a)]^2}var(a)+\frac{1}{2[\mathbb{E}(b)]^2}var(b)+\mathbb{E}\left[o((a-\mathbb{E}(a))^2+(b-\mathbb{E}(b))^2)\right]\\&=\log\left(\frac{\mathbb{E}(a)}{\mathbb{E}(b)}\right)+\frac{\mathbb{E}(b^2)}{2[\mathbb{E}(b)]^2}-\frac{\mathbb{E}(a^2)}{2[\mathbb{E}(a)]^2}+\mathbb{E}\left[o((a-\mathbb{E}(a))^2+(b-\mathbb{E}(b))^2)\right]\end{aligned}\tag{E.1}$$

Let $a=\mu_i(1,Q)$ and $b=\mu_i(0,Q)$, we then have

$$\mathbb{E}\left[\log\left(\frac{\mu_i(1,Q)}{\mu_i(0,Q)}\right)\right]=_{(1)}\log\left(\frac{\mathbb{E}(\mu_i(1,Q))}{\mathbb{E}(\mu_i(0,Q))}\right)+\frac{\mathbb{E}(\mu_i^2(1,Q))}{2[\mathbb{E}(\mu_i^2(1,Q))]^2}-\frac{\mathbb{E}(\mu_i^2(0,Q))}{2[\mathbb{E}(\mu_i(0,Q))]^2}+o(1).$$

(1) follows from Theorem 1, together with the fact that $\mu_i(1,Q)$ and $\mu_i(0,Q)$ are bounded. Therefore, we have

$$\left|\mathbb{E}\left[\log\left(\frac{\mu_i(1,Q)}{\mu_i(0,Q)}\right)\right]-\log\left(\frac{\mathbb{E}(\mu_i(1,Q))}{\mathbb{E}(\mu_i(0,Q))}\right)\right|=\left|\frac{\mathbb{E}(\mu_i^2(1,Q))}{2[\mathbb{E}(\mu_i^2(1,Q))]^2}-\frac{\mathbb{E}(\mu_i^2(0,Q))}{2[\mathbb{E}(\mu_i(0,Q))]^2}+o(1)\right|$$

By utilizing mean value theorem on function $\exp(x)$, we have

$$\begin{aligned}\left|DE(Q)-DE^{HH}(Q)\right|&\le\exp\left\{\max\left(\mathbb{E}\left[\log\left(\frac{\mu_i(1,Q)}{\mu_i(0,Q)}\right)\right],\log\left(\frac{\mathbb{E}(\mu_i(1,Q))}{\mathbb{E}(\mu_i(0,Q))}\right)\right)\right\}\left|\frac{\mathbb{E}(\mu_i^2(1,Q))}{2[\mathbb{E}(\mu_i(1,Q))]^2}-\frac{\mathbb{E}(\mu_i^2(0,Q))}{2[\mathbb{E}(\mu_i(0,Q))]^2}+o(1)\right|\\&\le\max\left\{DE(Q),DE^{HH}(Q)\right\}\left|\frac{\mathbb{E}(\mu_i^2(1,Q))}{2[\mathbb{E}(\mu_i(1,Q))]^2}-\frac{\mathbb{E}(\mu_i^2(0,Q))}{2[\mathbb{E}(\mu_i(0,Q))]^2}+o(1)\right|\\&\le_{(1)}DE^{HH}(Q)\left|\frac{\mathbb{E}(\mu_i^2(1,Q))}{2[\mathbb{E}(\mu_i(1,Q))]^2}+\frac{\mathbb{E}(\mu_i^2(0,Q))}{2[\mathbb{E}(\mu_i(0,Q))]^2}+o(1)\right|\end{aligned}$$

(1) is by Jensen's inequality and $DE(Q),DE^{HH}(Q)\ge 0$.

Proof 4 (Proof of Lemma 4) *Based on Lemma 1, we have*

$$a\le DE^{HH}(Q)\le b$$

where $a = DE(Q) - \max\left\{DE(Q), \frac{\mathbb{E}(\mu_i(1,Q))}{\mathbb{E}(\mu_i(0,Q))}\right\} \left(\frac{\mathbb{E}(\mu_i^2(1,Q))}{2[\mathbb{E}(\mu_i(1,Q))]^2} + \frac{\mathbb{E}(\mu_i^2(0,Q))}{2[\mathbb{E}(\mu_i(0,Q))]^2} + o(1)\right)$ *and* $b = DE(Q) + \max\left\{DE(Q), \frac{\mathbb{E}(\mu_i(1,Q))}{\mathbb{E}(\mu_i(0,Q))}\right\} \left(\frac{\mathbb{E}(\mu_i^2(1,Q))}{2[\mathbb{E}(\mu_i(1,Q))]^2} + \frac{\mathbb{E}(\mu_i^2(0,Q))}{2[\mathbb{E}(\mu_i(0,Q))]^2} + o(1)\right)$. *A consistent estimator of b and its upper bound are*

$$\hat{b} = \widehat{DE}(Q) + \max\left\{\widehat{DE}(Q), \frac{\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(0,Q)}\right\} \left(\frac{m\sum_{i=1}^m(\hat{\mu}_i^{ipw,TND}(1,Q))^2}{2[\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(1,Q)]^2} + \frac{m\sum_{i=1}^m(\hat{\mu}_i^{ipw,TND}(0,Q))^2}{2[\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(0,Q)]^2}\right)$$
$$\leq_{(1)} \widehat{DE}(Q) + \frac{\Gamma_2^2}{\Gamma_1^2}\max\left\{\widehat{DE}(Q), \frac{\Gamma_2}{\Gamma_1}\frac{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q)}\right\} \left(\frac{m\sum_{i=1}^m(\frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q))^2}{2[\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q)]^2} + \frac{m\sum_{i=1}^m(\frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q))^2}{2[\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q)]^2}\right).$$

(1) *is based on taking the upper bound* Γ_2 *of* ρ_i *in the numerator and the lower bound of* Γ_1 *of* ρ_i *in the denomenator and each estimator is larger or equal to* 0. *Similarly, a consistent estimator of a and its lower bound are*

$$\hat{a} = \widehat{DE}(Q) - \max\left\{\widehat{DE}(Q), \frac{\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(0,Q)}\right\} \left(\frac{m\sum_{i=1}^m(\hat{\mu}_i^{ipw,TND}(1,Q))^2}{2[\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(1,Q)]^2} + \frac{m\sum_{i=1}^m(\hat{\mu}_i^{ipw,TND}(0,Q))^2}{2[\sum_{i=1}^m \hat{\mu}_i^{ipw,TND}(0,Q)]^2}\right)$$
$$\geq \widehat{DE}(Q) - \frac{\Gamma_2^2}{\Gamma_1^2}\max\left\{\widehat{DE}(Q), \frac{\Gamma_2}{\Gamma_1}\frac{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q)}\right\} \left(\frac{m\sum_{i=1}^m(\frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q))^2}{2[\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q)]^2} + \frac{m\sum_{i=1}^m(\frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q))^2}{2[\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q)]^2}\right).$$

Therefore, we have, with probability $1 - o_p(1)$, $\widehat{DE}(Q) - c(\Gamma_1,\Gamma_2) \leq DE^{HH}(Q) \leq \widehat{DE}(Q) + c(\Gamma_1,\Gamma_2)$ *where*

$$c(\Gamma_1,\Gamma_2) = \frac{\Gamma_2^2}{\Gamma_1^2}\max\left\{\widehat{DE}(Q), \frac{\Gamma_2}{\Gamma_1}\frac{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q)}{\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q)}\right\} \left(\frac{m\sum_{i=1}^m(\frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q))^2}{2[\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(1,Q)]^2} + \frac{m\sum_{i=1}^m(\frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q))^2}{2[\sum_{i=1}^m \frac{1}{\rho_i}\hat{\mu}_i^{ipw,TND}(0,Q)]^2}\right).$$