Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Cluster-Randomized Trials with Cross-Cluster Interference

Michael P. Leung

To cite this article: Michael P. Leung (10 Oct 2025): Cluster-Randomized Trials with Cross-Cluster Interference, Journal of the American Statistical Association, DOI: 10.1080/01621459.2025.2566416

To link to this article: https://doi.org/10.1080/01621459.2025.2566416

Cluster-Randomized Trials with Cross-Cluster Interference

Michael P. Leung*

Department of Economics, University of California, Santa Cruz

*E-mail: leungm@ucsc.edu.

Abstract

The literature on cluster-randomized trials typically allows for interference within but not across clusters. This may be implausible when units are irregularly distributed across space without well-separated communities, as clusters in such cases may not align with significant geographic, social, or economic divisions. This paper develops methods for reducing bias due to cross-cluster interference. We first propose an estimation strategy that excludes units not surrounded by clusters assigned to the same treatment arm. We show that this substantially reduces bias relative to conventional difference-in-means estimators without significant cost to variance. Second, we formally establish a bias-variance trade-off in the choice of clusters: constructing fewer, larger clusters reduces bias due to interference but increases variance. We provide a rule for choosing the number of clusters to balance the asymptotic orders of the bias and variance of our estimator. Finally, we consider unsupervised learning for cluster construction and provide theoretical guarantees for *k*-medoids.

KEYWORDS: experimental design, causal inference, cluster-randomized trials, interference

1 Introduction

The literature on cluster-randomized trials (CRTs) predominantly assumes partial interference, which allows for interference within but not across clusters. However, researchers often conduct CRTs in what Hayes and Moulton (2017) refer to as "arbitrary geographical zones," where clusters are not given by nature since "the population is widely scattered and not divided up into clearly defined and well separated communities." A well-known concern in these settings is *cross-cluster interference*, where units within a cluster may respond to treatments assigned to units in adjacent clusters. Some references refer to this phenomenon as *contamination* (Hudgens and Halloran, 2008; Staples et al., 2015), while others define contamination to mean that control units procure treatment from treated neighbors, which is suggestive of interference (Hayes and Moulton, 2017). In either case, a simple comparison of treatment and control clusters may be biased.

Example 1 (Infectious disease trials).

CRTs are widely used in the infection control and hospital epidemiology literature (O'Hara et al., 2019). An example is the SolarMal trial (Homan et al., 2016), designed to evaluate the impact of mosquito trapping on malaria transmission. The CRT was conducted on an island of Kenya, partitioned into contiguous clusters using an unsupervised learning algorithm. As noted by Jarvis et al. (2017) in a review of infectious disease trials, partial interference "can be violated due to movement of people or diseases across borders, such as mosquitoes flying between control and intervention households."

Example 2 (Large-scale social experiments).

CRTs conducted on large populations have become increasingly common in academia (Muralidharan and Niehaus, 2017) and industry (Karrer et al., 2021). For example, Egger et al. (2022) study the general equilibrium effects of unconditional cash transfers using a CRT conducted in rural Kenya. The trial clustered villages into administrative units called "sublocations." The authors observe that "villages are relatively close to each other [and] sublocation boundaries are not 'hard' in any sense nor reflective of salient ethnic or social divides…there is extensive economic interaction in nearby markets regardless of sublocation."

This paper proposes methods for reducing bias due to cross-cluster interference in the analysis and design of cluster-randomized trials. Instead of partial interference, we consider a spatial interference model proposed by Leung (2022) which posits that interference decays with geographic distance. This captures the central concern expressed in the previous examples, namely potential interference between geographically proximate units.

At the analysis stage, a common bias-reduction strategy is the "fried-egg design" (Hayes and Moulton, 2017, Ch. 4.4.3). This is not an experimental design but rather entails excluding from estimation units in the trial located near cluster boundaries, the "whites" of the "fried eggs" that are the clusters. An open question is whether excluding observations in such a fashion is worth the loss of efficiency. We show that conventional difference-in-means estimators can have large biases due to interference near cluster boundaries. We then propose to improve the efficiency of the fried-egg design by excluding not all units near cluster boundaries but rather only the subset not surrounded by clusters assigned to the same

treatment arm. We prove that this can substantially reduce the asymptotic order of the bias relative to difference in means with no increase in the asymptotic order of the variance.

We then turn to optimal design of clusters. Partitioning the region into a small number of large clusters reduces bias since fewer units are located near cluster boundaries where they are most prone to cross-cluster interference. However, this comes at the cost of higher variance because the sample size in a CRT is the number of clusters. Unlike the conventional partial interference framework, our spatial interference model induces a bias-variance trade-off, which enables the study of cluster design for optimizing the trade-off.

We prove that the number of clusters *k* that balances the asymptotic bias and variance of our estimators depends on a parameter γ that measures the speed at which interference decays with distance. This formally characterizes how domain knowledge of interference informs optimal cluster construction. Such knowledge is implicitly used in practice when researchers define fried-egg boundaries or construct clusters with "buffer zones" to minimize interaction between units (Hayes and Moulton, 2017).

In practice, cluster construction is *ad hoc*. Homan et al. (2016) utilize a "traveling salesman algorithm," essentially an unsupervised learning method. More commonly, researchers use administrative divisions (e.g. Egger et al., 2022) or manually partition the study region (e.g. Moulton et al., 2001). Binka et al. (1998) study a malaria trial conducted in Northern Ghana, noting that, "Where possible, small paths or road were used to delineate the clusters. However, in most cases, the cluster boundaries did not correspond to natural barriers." We contribute to the literature by providing formal justification for constructing clusters using *k*-medoids, a well-known unsupervised learning algorithm.

We study CRTs under a design-based framework, as in Hudgens and Halloran (2008), Imai et al. (2009), and Schochet et al. (2022), among others. Unlike these papers, we do not assume partial interference and therefore require a different formulation of standard estimands.

Our work is most closely related to Leung (2022). He studies designs targeting the global average treatment effect (GATE) in which clusters are squares in $\mathbb{R}^2$ with identical areas. We consider a more general set of estimands and richer designs that may utilize unsupervised learning to construct clusters. Also, to accommodate denser spatial settings, we consider "infill-increasing" asymptotics in which the number of units is of larger asymptotic order than the volume of the study region. Finally, we propose new standard errors that are asymptotically conservative without restrictions on the superpopulation.

Our analysis of *k*-medoids builds on Cao et al. (2024). We extend a key result of theirs to the case where *k* diverges and characterize other *k*-medoid properties in this regime which may be of independent interest. We generalize their Ahlfours-regularity condition on the metric space to allow for infill-increasing asymptotics, though unlike them, we also require an additional boundary condition.

Faridani and Niehaus (2024) substantially generalize the theoretical results in Leung (2022) while retaining his focus on the GATE. Their results hold for general spaces defined using topological conditions that differ from Ahlfours regularity. For this reason, their proofs differ substantially from ours. They study designs that essentially cover the study region with non-intersecting balls of approximate radius g_n, where the radius is chosen to grow at an optimal

rate. Our design specifies an optimal choice for the number of clusters and uses unsupervised learning to construct clusters.

The paper is organized as follows. The next section defines the model and estimands. In § 3, we discuss the disadvantages of existing estimators and propose an alternative. We study the theoretical properties of the estimators in § 4 and propose an optimal design in § 5. We present simulation results in § 6 and an empirical application in § 7 using data from the Egger et al. (2022) trial. Finally, § 8 concludes and summarizes the practical outputs of our analysis.

We will use the following asymptotic order notation. For two sequences of random variables $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$, we write $X_n \precsim Y_n$ if $|X_n / Y_n| = O_p(1)$, $X_n \prec Y_n$ if $|X_n / Y_n| = o_p(1)$, $X_n \succsim Y_n$ if $|Y_n / X_n| = O_p(1)$, $X_n \succ Y_n$ if $|Y_n / X_n| = o_p(1)$, and $X_n \sim Y_n$ if both $X_n \precsim Y_n$ and $X_n \succsim Y_n$. For two sequences of constants, we use the same notation for analogous notions of asymptotic boundedness and domination.

2 Setup

Let $(\mathcal{X}, \rho)$ be a metric space where $\mathcal{X}$ is the set of spatial locations and ρ the metric. We observe a set of *n* units $\mathcal{N}_n \subseteq \mathcal{X}$, so $\rho(i, j)$ is the spatial distance between units $i, j \in \mathcal{N}_n$. Let D_i denote unit *i*'s binary treatment assignment and $\boldsymbol{D} = (D_i)_{i\in\mathcal{N}_n}$ the vector of observed assignments, which is the only random quantity in our analysis. Let $\{Y_i(\cdot)\}_{i\in\mathcal{N}_n}$ be a set of functions with domain $\{0,1\}^n$ and range $\mathbb{R}$. For $\boldsymbol{d} = (d_i)_{i\in\mathcal{N}_n} \in \{0,1\}^n$, $Y_i(\boldsymbol{d})$ denotes the potential outcome of unit *i* under the counterfactual that all units are assigned treatments according to $\boldsymbol{d}$. Unit *i*'s observed outcome is $Y_i = Y_i(\boldsymbol{D})$. We maintain the following standard assumption, that potential outcomes are uniformly asymptotically bounded. All asymptotic statements are with respect to a sequence indexed by $n \in \mathbb{N}$ unless otherwise indicated.

Assumption 1 (Bounded Outcomes).

$\max_{i\in\mathcal{N}_n} \max_{\boldsymbol{d}\in\{0,1\}^n} |Y_i(\boldsymbol{d})| \precsim 1.$

2.1 Spatial Interference

The primary metric space of interest is $\mathbb{R}^2$, but our results apply to a more general "Ahlfors-regular" space, augmented with a boundary condition. Define the *r -neighborhood* of unit *i*

$$\mathcal{N}(i, r) = \{j \in \mathcal{N}_n : \rho(i, j) \le r\}.$$

Assumption 2 (Metric Space).

There exist constants $C, d > 0$ *and a positive sequence* $\{\xi_n\}_{n\in\mathbb{N}}$ *such that* $1 \precsim \xi_n \prec n$ *and the following hold. (a) For any* $n \in \mathbb{N}$, $i \in \mathcal{N}_n$, *and* $r > 0$,
$\min\{C^{-1}\xi_n r^d, n\} \le |\mathcal{N}(i, r)| \le \max\{C\xi_n r^d, 1\}$. *(b)* $d \ge 1$, *and*
$\max_{i\in\mathcal{N}_n} |\mathcal{N}(i, r+1) \setminus \mathcal{N}(i, r)| \le C \max\{\xi_n r^{d-1}, 1\}$ *for any* $r > 0$.

The constant *d* represents the spatial dimension and ξ_n a measure of density (units per unit volume). The dependence on ξ_n is new to this paper relative to prior work on spatial interference and accommodates applications with denser regions. To understand the assumption, consider the standard *increasing-domain* case in which $\xi_n = 1$ for all *n*. Part (a) says that the number of units in an *r*-neighborhood is the same order r^d as the neighborhood's volume. This defines a (C, d)-finite Ahlfors-regular space, the same metric space studied by Cao et al. (2024). The boundary condition in part (b) is new relative to their setup, but both (b) and the upper bound in (a) are satisfied if $\mathcal{X} = \mathbb{R}^d$ under the usual increasing-domain assumption that units are minimally separated in space (Jenish and Prucha, 2009, Lemma A.1).

Assumption 2 also accommodates the *infill-increasing* case in which ξ_n diverges with *n*. Here the number of units in any neighborhood is of larger order than the neighborhood volume by a factor of ξ_n, corresponding to a denser region. Because we require $\xi_n \prec n$, this is a hybrid of infill and increasing-domain asymptotics in which the volume of the study region $\mathcal{N}_n$ grows but at a slower rate n/ξ_n than the population size *n* (as in Lahiri and Zhu, 2006, for example).

Under Assumption 2(a), the number of units in any *r*-ball is proportional to $\xi_n r^d$, which allows for some variation in density across the study region. The assumption is violated if two subregions of similar volume exist but the number of units in one is a large multiple of the other. This could be the case in practice if the study region encompasses urban and rural areas. In §SA.2.1, we discuss how our proposed methodology can be modified to accommodate large variations in density.

Assumption 3 (Interference).

There exist $c > 0$ *and* $\gamma > d$ *for d defined in Assumption 2 such that for all* $r \geq 0$,

$$\sup_{n\in\mathbb{N}} \max_{i\in\mathcal{N}_n} \max\{|Y_i(\boldsymbol{d}) - Y_i(\boldsymbol{d}')| : \boldsymbol{d}, \boldsymbol{d}' \in \{0,1\}^n, d_j = d_j' \ \forall j \in \mathcal{N}(i,r)\} \leq c\min\{r^{-\gamma}, 1\}.$$

This corresponds to Assumption 3 of Leung (2022). Unlike partial interference, it is formulated independently of clusters, enabling us to develop a theory of optimal cluster construction. To interpret the condition, consider a unit *i* centered at a "cluster" $\mathcal{N}(i,r)$ of radius *r*. The quantity $|Y_i(\boldsymbol{d}) - Y_i(\boldsymbol{d}')|$ measures interference induced by manipulating the treatment assignments of units outside the cluster, and the assumption requires this to decay like $r^{-\gamma}$ or faster. Hence, the larger the minimum distance *r* between *i* and the units with manipulated treatments, the smaller the spillover effect. We thus interpret $1/\gamma$ as (an upper bound on) the *degree of interference*. In § 5, we discuss the optimal design of clusters using knowledge of γ.

Assumption 3 further requires γ to decay fast enough relative to the spatial dimension *d*. This coincides with the requirement imposed by Leung (2022) for $d = 2$. In the increasing-domain case, the intuition is that *r*-neighborhood sizes grow like r^d under Assumption 2, so weak dependence requires interference to decay faster than the rate at which neighborhoods

densify with *r*. Central limit theorems for spatial processes impose analogous requirements on mixing coefficients which control the degree of spatial dependence, as γ does in our framework (e.g. Jenish and Prucha, 2009, Assumption 3(b)).

2.2 Design

We interchangeably use *k* or k_n to denote the number of clusters in the design. For a given population $\mathcal{N}_n$, denote by $\mathcal{C}_n = \{C_j\}_{j=1}^k$ the set of clusters, which is a partition of $\mathcal{N}_n$. We study standard two-stage randomized-saturation designs.

Assumption 4 (Assignment Mechanism).

For $q \in (0,1)$, $p_0, p_1 \in [0,1]$, *and* $t \in \{0,1\}$, $\{D_i\}_{i \in C_j} \overset{iid}{\sim} \text{Bernoulli}(p_t)$ *conditional on* $W_j = t$ *where* $\{W_j\}_{j=1}^k \overset{iid}{\sim} \text{Bernoulli}(q)$.

Under this design, *k* clusters are independently randomized into treatment with probability *q* with W_j denoting cluster *j*'s treatment assignment. If $W_j = t$, all units in cluster *j* are randomized into treatment with probability p_t. The literature often refers to p_1, p_0 as *saturation levels*.

Let ξ_n and *d* be given from Assumption 2. We consider (sequences of) clusters satisfying the following properties.

Assumption 5 (Spatial Clusters).

There exist positive sequences $\{L_n\}_{n \in \mathbb{N}}$ *and* $\{U_n\}_{n \in \mathbb{N}}$ *for which the following hold. (a) For any sequence of clusters* $\{C_n\}_{n \in \mathbb{N}}$ *with* $C_n \in \mathcal{C}_n$ *for all n, there exists a sequence of "centroid" units* $\{i_n\}_{n \in \mathbb{N}}$ *with* $i_n \in C_n$ *for all n such that* $\mathcal{N}(i_n, L_n) \subseteq C_n \subseteq \mathcal{N}(i_n, U_n)$. *(b)* $L_n \sim U_n \sim (n/(k_n \xi_n))^{1/d}$.

This requires clusters to be globular in that they contain and are contained by balls with radii of the same order. Under Assumption 2(a), part (b) implies that the number of units in any cluster is order n/k_n, which implies that clusters are comparable in size. We show below that clusters generated by *k*-medoids satisfy these requirements. Note that even under partial interference, restrictions on cluster size heterogeneity are required for inference using conventional clustered standard errors (Hansen and Lee, 2019, Assumption 2), although our requirements are stronger. These may be possible to relax, but we leave this to future research.

Assumption 5 allows clusters to have quite heterogeneous shapes, as can be seen in Figure 1. It depicts *k*-medoid clusters, which satisfy Assumption 5 by Theorem 1 below. However, the assumption is violated if clusters are highly size-imbalanced, such as elongated clusters that narrowly encompass a geographical feature such as a river or road. This may be addressed by subdividing large or abnormally shaped clusters or grouping adjacent small clusters. The assumption can also be violated if clusters are not constructed based on spatial proximity. For

instance, if units are connected through an online social network connecting units far apart in space, then clusters based on network connectivity would likely violate the assumption.

Assumption 5 is satisfied if the researcher partitions the space into k_n identically-sized cubes, but such clusters do not adapt to the spatial distribution of units. We therefore suggest using unsupervised learning algorithms. The next result provides theoretical guarantees for the well-known *k*-medoids algorithm stated in §SA.1.

Theorem 1 .

Suppose clusters are the output of k-medoids, given in Algorithm SA.1.1. Under Assumption 2(a), the clusters satisfy Assumption 5.

In practice, *k*-means can also be used since it typically delivers clusters similar to those of *k*-medoids. The globular clusters produced by both algorithms are sometimes viewed unfavorably relative to algorithms such as spectral clustering for certain unsupervised learning tasks. However, for our purposes, globular clusters are preferable because they minimize the number of units near cluster boundaries, which are the primary source of bias due to cross-cluster interference. More broadly, Assumption 5 provides general design principles for constructing clusters, namely to aim for balance and globularity, which may be achieved either using these algorithms or manually.

2.3 Causal Estimands

Because we allow for cross-cluster interference, we need to redefine conventional estimands in a manner free of this source of bias. To this end, define for $t \in \{0,1\}$ the p_t *-counterfactual design*, which sets $k = 1$ and $q = t$ in Assumption 4. This design groups the population into a single cluster and assigns treatment with probability p_t . Throughout the paper, let $\mathbf{E}[\cdot]$ denote the expectation taken with respect to the observed design in Assumption 4 and $\mathbf{E}^*_{p_t}[\cdot]$ the expectation taken with respect to the p_t -counterfactual design.

Let $\boldsymbol{D}_{-i}$ denote the assignment vector excluding the *i*th component and $Y_i(d, \boldsymbol{D}_{-i})$ denote *i*'s potential outcome under the counterfactual that *i*'s observed assignment is $d \in \{0,1\}$, holding fixed the realized assignments of other units. We study estimands of the form

$$\theta^* \equiv \theta^*(d_1, d_0; p_1, p_0) = \frac{1}{n} \sum_{i \in \mathcal{N}_n} (\mathbf{E}^*_{p_1}[Y_i(d_1, \boldsymbol{D}_{-i})] - \mathbf{E}^*_{p_0}[Y_i(d_0, \boldsymbol{D}_{-i})])$$

for $d_1, d_0 \in \{\varnothing, 0, 1\}$, where we define $Y_i(\varnothing, \boldsymbol{D}_{-i}) \equiv Y_i(\boldsymbol{D})$. The following special cases are analogous to estimands defined by Hudgens and Halloran (2008) and Hayes and Moulton (2017). The *direct effect* compares treated and untreated units under the p_1 -counterfactual design:

$$\theta^*_D = \theta^*(1, 0, p_1, p_1) = \frac{1}{n} \sum_{i \in \mathcal{N}_n} \mathbf{E}^*_{p_1}[Y_i(1, \boldsymbol{D}_{-i}) - Y_i(0, \boldsymbol{D}_{-i})].$$

This is directly identified if the p_1-counterfactual design were implemented in practice. The estimands that follow, however, require multiple clusters assigned to different arms because they compare different counterfactual designs. Under our framework, this is the primary motivation for cluster randomization.

The *indirect effect* compares outcomes of untreated units under counterfactual designs with different saturation levels:

$$\theta_I^* = \theta^*(0,0; p_1, p_0) = \frac{1}{n}\sum_{i\in\mathcal{N}_n}(\mathbf{E}^*_{p_1}[Y_i(0, \boldsymbol{D}_{-i})] - \mathbf{E}^*_{p_0}[Y_i(0, \boldsymbol{D}_{-i})]).$$

Interest often centers on the "pure control" baseline of $p_0 = 0$, so that $\mathbf{E}^*_{p_0}[Y_i(0, \boldsymbol{D}_{-i})] = Y_i(\mathbf{0})$. The *total effect* is the sum of the direct and indirect effects, equal to $\theta_T^* = \theta^*(1,0; p_1, p_0)$. Finally, the *overall effect* compares outcomes under different saturation levels: $\theta_O^* = \theta^*(\varnothing, \varnothing, p_1, p_0)$. Leung (2024) provides conditions under which these have causal interpretations.

3 Estimators

3.1 Existing Approaches

Let $c(i) \in \{1, \ldots, k\}$ denote the index of the cluster containing unit *i*. A common strategy for estimating θ_I^* is to compute the difference in means between control units in clusters assigned saturation level p_1 and those in clusters assigned level p_0:

$$\frac{\sum_{i\in\mathcal{N}_n}(1-D_i)W_{c(i)}Y_i}{\sum_{i\in\mathcal{N}_n}(1-D_i)W_{c(i)}} - \frac{\sum_{i\in\mathcal{N}_n}(1-W_{c(i)})Y_i}{\sum_{i\in\mathcal{N}_n}(1-W_{c(i)})}. \quad (1)$$

This is the sample analog of

$$\frac{1}{n}\sum_{i\in\mathcal{N}_n}(\mathbf{E}[Y_i \mid (1-D_i)W_{c(i)} = 1] - \mathbf{E}[Y_i \mid W_{c(i)} = 0]),$$

which may be quite far from the target θ_I^*. Units in control clusters near cluster boundaries may be spatially proximate to units in treated clusters and therefore at greater risk of contamination. For such units *i*, $\mathbf{E}[Y_i \mid W_{c(i)} = 0]$ may be substantially different from $\mathbf{E}^*_{p_0}[Y_i(0, \boldsymbol{D}_{-i})]$.

The fried-egg design attempts to reduce bias by restricting the comparison in (1) to the subset of units deemed sufficiently far from cluster boundaries. Unfortunately, this has two problems. First, the resulting estimator is in fact asymptotically biased because boundary units are excluded with probability one, so it only estimates an average effect for the

subpopulation of units in cluster interiors. As noted by McCann et al. (2018), this differs from the target estimand since units in cluster interiors may be systematically different from those near the boundaries due to spatial heterogeneity. Second, it is inefficient. The purpose of only including units in the interiors of control clusters is that such units are "well surrounded" by control clusters, or equivalently, relatively far from treated clusters. However, boundary units may be well surrounded in the same fashion if all neighboring clusters are assigned to control, so it would be just as useful to include these units.

3.2 Our Approach

Fix a neighborhood radius r_n to be defined in (2) below. Call a unit *i well-surrounded* if

$$S_i \equiv \max_{t\in\{0,1\}} \prod_{j\in\mathcal{N}(i,r_n)} W_{c(j)}^{t}(1-W_{c(j)})^{1-t} = 1,$$

that is, if a unit *i*'s r_n-neighborhood only intersects clusters assigned to the same treatment arm. Figure 1 depicts *k*-medoid clusters, marking with an "X" units that are not well surrounded. Our strategy is to only exclude from estimation units that are not well surrounded. If $r_n = 0$, then all units are well surrounded, and our estimators reduce to difference in means. Choosing a larger radius r_n is analogous to choosing a larger boundary region for exclusion in a fried-egg design.

Under our assumptions, the number of well-surrounded units is of asymptotic order equal to the population size *n* because any unit has nontrivial probability of being well-surrounded (Lemma SA.5.3). As a consequence, our strategy solves the fried-egg design's boundary bias problem and typically excludes strictly fewer units.

Let d_1, d_0 be given from the estimand θ^*. For any $t\in\{0,1\}$ and $i\in\mathcal{N}_n$, define $T_{ti} = \mathbf{1}\{D_i = d_t, W_{c(i)} = t\}S_i$ if $d_t \in \{0,1\}$ and $T_{ti} = \mathbf{1}\{W_{c(i)} = t\}S_i$ if $d_t = \varnothing$. Let $p_{ti} = \mathbf{E}[T_{ti}]$ be the propensity score, which has a closed-form expression given in Remark 1 below. We propose the following Hájek estimator for θ^*:

$$\hat{\theta} = \hat{\mu}_1 - \hat{\mu}_0 \quad \text{for} \quad \hat{\mu}_t = \frac{\sum_{i\in\mathcal{N}_n} T_{ti}Y_i / p_{ti}}{\sum_{i\in\mathcal{N}_n} T_{ti} / p_{ti}}.$$

When $r_n = 0$, $S_i = 1$ for all *i*, so propensity scores are homogeneous across *i*, and $\hat{\theta}$ reduces to difference in means. For $r_n > 0$, the scores are generally spatially heterogeneous. For instance, boundary units are less likely to be well surrounded than interior units because that would require more clusters to be assigned the same saturation level.

For any cluster C_j, let m_j be its centroid from Assumption 5 and $R_j = \max_{i\in C_j}\rho(i,m_j)$, the "radius" of the cluster. We propose setting

$$r_n = 0.5\cdot\text{Median}(\{R_j\}_{j=1}^{k}). \quad (2)$$

In the case where clusters are equally sized squares, this coincides with the radius suggested by Leung (2022).

Remark 1 (Overlap).

Let ϕ_i be the number of clusters intersecting $\mathcal{N}(i, r_n)$. The propensity score can be written explicitly as

$$p_{ti} = p_t^{d_t}(1-p_t)^{1-d_t} q^{\phi_i t}(1-q)^{\phi_i(1-t)}$$

where $p_t^{\varnothing} \equiv 1$. We show in Lemma SA.5.3 that ϕ_i is asymptotically bounded uniformly in *i* under (2). Since $q \in (0,1)$ by Assumption 4, p_{ti} is similarly bounded away from zero when the saturation levels p_1, p_0 are nontrivial (see Assumption 6 below). Hence, using a CRT paired with our estimation strategy inherently avoids positivity violations or limited overlap.

Remark 2 .

The choice of 0.5 in (2) is immaterial for the asymptotic theory, but in finite samples it controls a bias-variance trade-off. We choose 0.5 to informally balance the two. Constants close to zero will yield high bias because $r_n = 0$ corresponds to difference in means. Constants close to one result in high variance. To see this, suppose each cluster is a ball with homogeneous radius *R*. Setting the constant to 1 implies $r_n = R$, so the only unit in any cluster that is well surrounded with probability one is the centroid. A unit located slightly north of the centroid finds its r_n-neighborhood intersecting at least two clusters, so the chance that it is well surrounded, and hence not excluded, is substantially lower. In contrast, suppose $r_n = 0.5R$, and define a cluster C_j's "interior" as the ball centered at m_j with radius $R/2$. Then all units in the interior are well surrounded with probability one because their r_n-neighborhoods are contained within C_j.

Lastly, we propose a variance estimator for $\hat{\theta}$. Let

$$\Lambda_i = \{\ell \in \mathcal{N}_n : \max_{j \in \{1,\ldots,k_n\}} |C_j \cap \mathcal{N}(i, r_n)| |C_j \cap \mathcal{N}(\ell, r_n)| > 0\}, \quad (3)$$

the set of units ℓ for which some cluster intersects the r_n-neighborhood of ℓ and *i*. These can be thought of as the units "most potentially correlated" with *i*. Define $A_{ij}(1) = \mathbf{1}\{j \in \Lambda_i\}$, $A_{ij}(2) = \mathbf{1}\{j \in C_{c(i)}\}$, and $\hat{Z}_i = (T_{1i}(Y_i - \hat{\mu}_1))/p_{1i} - (T_{0i}(Y_i - \hat{\mu}_0))/p_{0i}$. The variance estimator is

$$\hat{\sigma}^2 = \max_{t \in \{1,2\}} \hat{\sigma}^2(u) \quad \text{where} \quad \hat{\sigma}^2(u) = \frac{k_n}{n^2} \sum_{i \in \mathcal{N}_n} \sum_{j \in \mathcal{N}_n} \hat{Z}_i \hat{Z}_j A_{ij}(u). \quad (4)$$

Notice that $\hat{\sigma}^2(2)$ is the conventional cluster-robust variance estimator (e.g. Baird et al., 2018), which only accounts for within-cluster dependence. The estimator $\hat{\sigma}^2(1)$ is analogous

to that of Leung (2022) and additionally accounts for cross-cluster dependence since $C_{c(i)} \subseteq \Lambda_i$. In § 4, we discuss the advantages of taking the larger of the two.

4 Asymptotic Theory

We next derive bounds on the rates of convergence of our estimator and difference in means. We then characterize the asymptotic distribution of our estimator and prove that the variance estimator is asymptotically conservative.

Assumption 6 (Overlap).

For all $t \in \{0,1\}$, *if* $d_t = 1$ *(*$d_t = 0$*), then* $p_t > 0$ *(*$p_t < 1$*).*

As discussed in Remark 1, this ensures overlap or positivity. For the next two theorems, abbreviate the estimand $\theta^*(d_1, d_0, p_1, p_0)$ as θ^*. Recall the asymptotic order notation from the end of § 1.

Theorem 2 (Our Estimator).

Suppose $k_n \precsim n/\xi_n$. *Under Assumptions 1–6,* $|\hat{\theta} - \theta^*| \precsim r_n^{-\gamma} + k_n^{-1/2} \sim (k_n \xi_n / n)^{\gamma/d} + k_n^{-1/2}$.

The result establishes a bias-variance trade-off in k_n. The $k_n^{-1/2}$ term is the contribution of (the square root of) the variance since the effective sample size in a CRT is the number of clusters k_n. The asymptotic bias is order $r_n^{-\gamma}$. Intuitively, if r_n is large, then the expected outcome of a unit *i* assigned treatment d_t and well surrounded by clusters assigned to saturation level p_t should well approximate $\mathbf{E}^*_{p_t}[Y_i(d_t, \boldsymbol{D}_{-i})]$, corresponding to lower bias. The bias decreases at a faster rate for larger γ since this corresponds to a lower degree of spatial interference.

The requirement $k_n \precsim n/\xi_n$ is mild, and typically we would have $k_n \prec n/\xi_n$. In the increasing-domain case where $\xi_n = 1$, necessarily $k_n \precsim n/\xi_n$ since $k_n \leq n$, and usually the number of clusters is of smaller order than the population size. In the infill-increasing case, if k_n grows faster than n/ξ_n, which is the volume of the study region under Assumption 2, then we would be increasingly subdividing the region into smaller clusters of shrinking volume, analogous to having more clusters than units.

Denote the difference-in-means estimator of θ^* by $\hat{\theta}^+$, which corresponds to setting $r_n = 0$ in the definition of $\hat{\theta}$. Let $\hat{\theta}^+_D$, $\hat{\theta}^+_T$, and $\hat{\theta}^+_O$ be the difference-in-means estimates of θ^*_D, θ^*_T, and θ^*_O defined in § 2.3.

Theorem 3 (Difference in Means).

(a) Under Assumptions 1–6, $|\hat{\theta}^+ - \theta^*| \precsim (k_n \xi_n / n)^{1/d} + k_n^{-1/2}$. *(b) Suppose the design satisfies Assumption 4 for* $p_1 \in (0,1)$, $p_0 = 0$. *There exist a metric space and sequence of units,*

clusters, and potential outcomes satisfying Assumptions 1–3 and 5 such that $|\hat{\theta}_Q^+ - \theta_Q^*| \gtrsim (k_n\xi_n / n)^{1/d} + k_n^{-1/2}$ *for any* $Q \in \{D,T,O\}$.

Part (a) provides an upper bound on the rate. Part (b) shows that the rate is tight for several illustrative cases. Recall that setting $p_0 = 0$ corresponds to the "pure control" baseline used in the estimands of Hayes and Moulton (2017).

The $k_n^{-1/2}$ term in the bound is (the square root of) the variance contribution, as in Theorem 2, while $(k_n\xi_n / n)^{1/d}$ is the bias contribution. The latter is notably worse than that of $\hat{\theta}$ because a reduction in the degree of interference $1/\gamma$ has no effect on the rate. The reason is that units situated at cluster boundaries may be directly proximate to clusters assigned to different treatment arms, and as shown in the proof, the share of such units can be of order $(k_n\xi_n / n)^{1/d}$.

These results demonstrate that excluding units in the manner of $\hat{\theta}$ can significantly reduce the asymptotic order of the bias relative to difference in means. While excluding units may come at the cost of efficiency, there is no increase to the asymptotic order of the variance because the variances of both estimators scale not with the number of units but with the number of clusters. Hence, the efficiency loss is second-order relative to the potential reduction in bias.

Define $\mu_t = n^{-1}\sum_{i\in\mathcal{N}_n} \mathbf{E}[Y_i \mid T_{ti} = 1]$, $\bar{\theta} = \mu_1 - \mu_0$, and

$$\sigma_n^2 = \mathrm{Var}\left(\sqrt{k_n}\,\frac{1}{n}\sum_{i\in\mathcal{N}_n}\left(\frac{T_{1i}(Y_i - \mu_1)}{p_{1i}} - \frac{T_{0i}(Y_i - \mu_0)}{p_{0i}}\right)\right).$$

Note that $\sigma_n^2 \precsim 1$ by the proof of Theorem 2.

Theorem 4 (CLT).

Suppose $1 \prec k_n \prec n/\xi_n$ *and* $\sigma_n^2 \gtrsim 1$. *Under Assumptions 1–6,*

$$\sigma_n^{-1}\sqrt{k_n}(\hat{\theta} - \bar{\theta}) \xrightarrow{d} \mathcal{N}(0,1). \quad (5)$$

Furthermore, if $k_n \prec (n/\xi_n)^{\frac{2\gamma}{2\gamma+d}}$, *then*

$$\sigma_n^{-1}\sqrt{k_n}(\hat{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0,1). \quad (6)$$

The first result (5) centers the estimator at $\bar{\theta}$, the probability limit of $\hat{\theta}$. This is not the estimand of interest since there is an asymptotic bias $|\bar{\theta} - \theta^*|$ by Theorem 2. To use the normal limit to justify the validity of conventional CIs for θ^*, the second result (6) requires "undersmoothed designs" in which the number of clusters is of smaller order than the optimal

rate discussed in the next section. This ensures that the asymptotic bias is small. It is analogous to nonparametric regression where rate-optimal tuning parameter choices result in asymptotic bias, so conventional CIs require undersmoothing. We discuss practical choices of k_n in § 5. In §SA.2.3, we discuss "bias-aware" inference, an alternative to undersmoothing.

Theorem 5 (Variance Estimator).

Suppose $1 \prec k_n \prec n/\xi_n$ *and* $\sigma_n^2 \succsim 1$. *Under Assumptions 1–6,* $\hat{\sigma}^2(t) = \sigma_n^2 + \mathcal{B}_n + o_p(1)$ *for* $t \in \{1,2\}$ *and some sequence of non-negative constants* $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$. *Hence, both* $\hat{\sigma}^2(1)$ *and* $\hat{\sigma}^2(2)$ *are asymptotically conservative.*

The cluster-robust variance estimator $\hat{\sigma}^2(2)$ has the advantage of being non-negative in finite sample, unlike $\hat{\sigma}^2(1)$ which is a truncation estimator (Andrews, 1991, p. 823). On the other hand, $\hat{\sigma}^2(2)$ only accounts for within-cluster dependence, whereas $\hat{\sigma}^2(1)$ also accounts for cross-cluster dependence in the definition $A_{ij}(1)$. As shown in the proof, this dependence vanishes, so $\hat{\sigma}^2(1)$ is a valid estimator. However, the dependence vanishes at the slow rate $(k_n\xi_n/n)^{1/d}$, the same order as the bias of difference in means. Thus in smaller samples, ignoring second-order terms may result in anti-conservativeness. By taking the larger of the two estimators, we obtain the benefits of both. See §SA.3 for a comparison of $\hat{\sigma}^2$ with other variance estimators in the literature.

5 Optimal Design

Recall that *d* is the dimension of the spatial region, ξ_n is the density of the region (number of units per unit volume/area), and γ is a lower bound on the speed at which interference decays with distance. By Theorem 2, choosing

$$k \sim (n/\xi_n)^{\frac{2\gamma}{2\gamma+d}} \quad (7)$$

optimizes the rate of convergence of $\hat{\theta}$. This formalizes how domain knowledge of interference γ informs the design of clusters. The right-hand side is increasing in γ since less interference means the same level of bias reduction can be achieved with more clusters. It is decreasing in density ξ_n because having more units in a given area effectively corresponds to greater interference (an infectious disease may spread more easily).

Choosing the number of clusters according to (7) balances the asymptotic orders of the bias and variance of $\hat{\theta}$. However, as discussed in § 4, validity of the CI

$$\hat{\theta} \pm 1.96 \cdot \hat{\sigma} k^{-1/2} \quad (8)$$

requires the bias to be of smaller order than the variance. This requires an *undersmoothed design* where *k* is of smaller order than (7).

We next recommend an undersmoothed choice of *k* given domain knowledge of γ. Let $\mathcal{V}$ denote the area or volume of the study region containing $\mathcal{N}_n$. As discussed below, $\mathcal{V}$ is of order n/ξ_n, so the rate-optimal formula (7) can be rewritten as $k \sim \mathcal{V}^{\frac{2\gamma}{2\gamma+d}}$. While one could choose *k* equal to the right-hand side, this choice would be extremely sensitive to the unit of length used to measure geographic distance. Switching from square kilometers to square millimeters would dramatically increase $\mathcal{V}$.

Our observation is that both the unit of length and γ determine the speed at which interference decays with distance. For any given choice of γ, the rate of decay is significantly faster if distance is measured in millimeters compared to kilometers, for example. Therefore, domain knowledge of interference informs both γ and the unit of length, and one can determine *k* from these ingredients as follows.

1. Based on domain knowledge, specify a unit of length and rate of decay γ under which Assumption 3 is believed to hold. Choose a strict lower bound $\tilde{\gamma} < \gamma$. For a given unit of length, the conservative choice allowed by Assumption 3 is $\tilde{\gamma} = d$. We suggest this choice absent any prior information on γ.

2. Let $\mathcal{V}$ denote the volume of the study region containing $\mathcal{N}_n$ under the chosen unit of length. Set the number of clusters to

$$k = [\min\{\mathcal{V}, n\}^{\frac{2\tilde{\gamma}}{2\tilde{\gamma}+d}}], \quad (9)$$

where *[c]* means round *c* to the nearest integer.

The next subsection provides empirical examples of calibrating *k* under the conservative choice $\tilde{\gamma} = d$. The last subsection discusses how to determine γ to obtain less conservative estimates.

Remark 3 .

The theoretical motivation for (9) is as follows. First, using $\tilde{\gamma} < \gamma$ corresponds to undersmoothing. Second, Assumption 2(a) says that the number of units in an *r*-neighborhood is of asymptotic order equal to the density ξ_n times the volume of the neighborhood r^d. It follows that the number of units *n* is of asymptotic order equal to ξ_n times the volume of the region: $n \sim \xi_n \mathcal{V}$. Hence, in the infill-increasing case where ξ_n is diverging, we replace n/ξ_n in the rate-optimal formula (7) with $\mathcal{V}$, which can be directly computed from the data. In the increasing-domain case where $\xi_n \sim 1$, we can replace n/ξ_n with either *n* or $\mathcal{V}$ since they are of the same order, so (9) conservatively chooses the smaller option.

5.1 Empirical Examples

Sur et al. (2009) conduct a CRT in an urban slum in India spanning about 1.2 by 0.7 km with 38k participants. To compute our suggested number of clusters (9), we need to select $\tilde{\gamma}$ and the unit of length. Suppose we choose the conservative bound $\tilde{\gamma} = d = 2$, so that interference is assumed to decay like r^{-2} or faster with each unit of length *r*. If we take 35m to be the unit of length, then for units *i*, *j*, *k* such that $\rho(i, j) = 35\,\text{m}$ and $\rho(i, k) = 70\,\text{m}$, the extent to which *k*'s treatment affects *i* is less than $1/4\text{th}$ ($2^{-\tilde{\gamma}} = 0.25$) as small as the extent to which *j*'s affects *i*. This is with only a 35m difference in distance. For this unit of length, (9) yields $k = 78$. In other words, these are the assumptions on interference that justify the authors' choice of $k = 80$, originally determined by a conventional power analysis based on a partial interference model.

The previous rate of decay may be over-optimistic, so suppose the relevant unit of length is in 100m increments. If $\rho(i, j) = 100\,\text{m}$ and $\rho(i, k) = 200\,\text{m}$, the extent to which *k*'s treatment affects *i* is less than $1/4\text{th}$ as small as the extent to which *j*'s affects *i*, now with a 100m difference in distance. With a slower rate of decay, bias is higher, which requires constructing fewer, larger clusters. As a result, (9) yields only $k = 19$.

Next consider Homan et al. (2016) whose trial region is substantially larger at roughly 12 by 4 km with nearly the same number of units (34k). Due to the lower density, there is less bias from interference, so a choice of $k = 81$ can be justified under weaker assumptions on interference. Specifically, if $\tilde{\gamma} = 2$ but the unit of length is now in 250m increments, then (9) results in $k = 84$. For context, the *Anopheles* mosquito that is the subject of their trial typically does not fly more than 2 km from their breeding grounds (CDC, 2025).

These examples illustrate what domain knowledge of the degree of interference entails. They also show how our proposed choice of *k* accounts for regional density, unlike the standard power analysis (e.g. Hemming et al., 2011, eq. (9)). In both examples, weaker assumptions on interference require choosing *k* smaller than the standard analysis to better balance bias and variance. To justify choosing larger *k*, the researcher must either collect data over a larger area or be willing to entertain stronger assumptions on interference.

5.2 Bounding Interference

Optimal design generally requires prior information on certain population parameters. The standard power analysis for CRTs assumes partial interference and requires knowledge of the intracluster correlation coefficient, a measure of within-cluster dependence in potential outcomes (e.g. Baird et al., 2018; Hemming et al., 2017). Our results suggest that, if cross-cluster interference is of first-order importance, the focus of attention should instead be γ. As illustrated in the previous subsection, the conservative choice $\gamma = d$ can result in relatively small values of *k* because it allows for a greater degree of interference, so to the extent that one can justify stronger assumptions on interference, that is, values of γ that are larger than *d* for a given unit of length, this would improve asymptotic power.

In the context of infectious disease trials, Halloran et al. (2017) argue that CRT design should be informed by simulating models of disease transmission. Several papers utilize parametric models and simulation methods to estimate or bound contamination bias. Alexander et al.

(2020) and Jarvis et al. (2019) use spatial models to provide evidence of contamination in prior CRTs. Multerer et al. (2021) employ models of disease transmission for a similar purpose. Our theory provides a precise way in which modeling can inform design, namely by providing plausible values of γ. This relates to Staples et al. (2015) who show how to estimate a different measure of cross-cluster interference to assess the degree to which the conventional power analysis overstates trial power.

To estimate γ, models may be combined with external data sources such as data from pilot studies. In the context of malaria vector control, mosquito mark-release-recapture experiments (e.g. Guerra et al., 2014) provide data on geographic dispersion of malaria carriers, which is informative of spatial interference. We provide additional suggestions in §SA.2.2.

6 Simulation Study

We conduct a simulation study to illustrate the finite-sample properties of our estimator and difference in means under our proposed design. We randomly draw unit locations from the square $[-(n\alpha_n)^{1/2},(n\alpha_n)^{1/2}]^2$ with $\alpha_n = 0.8, 0.7, 0.6$, respectively. This corresponds to the infill-increasing case where the regional volume shrinks with the population size. We create clusters using *k*-medoids with *k* given by (9) using the conservative choice $\tilde{\gamma} = d = 2$ and set the assignment probabilities in Assumption 4 to $(q, p_1, p_0) = (0.7, 0.5, 0)$.

Let $\{\tilde{\varepsilon}_i\}_{i\in\mathcal{N}_n} \overset{iid}{\sim} \mathcal{N}(-0.5,1)$, $\{\beta_i\}_{i\in\mathcal{N}_n} \overset{iid}{\sim} \mathcal{N}(2,1)$, and $\{\gamma_i\}_{i\in\mathcal{N}_n} \overset{iid}{\sim} \mathcal{N}(1,1)$ be independent and drawn independently of locations. We generate spatially autocorrelated errors $\varepsilon_i = \tilde{\varepsilon}_i + \sum_{j\in\mathcal{N}_n} \mathbf{1}\{\rho(i,j)\le 1\}\tilde{\varepsilon}_j \,/\, \sum_{k\in\mathcal{N}_n} \mathbf{1}\{\rho(i,k)\le 1\}$. For $w_{ij} = \min\{\rho(i,j)^{-5},1\}$, we generate outcomes according to

$$Y_i = \sum_{j\in\mathcal{N}_n} w_{ij} D_j \beta_j + \sum_{j\in\mathcal{N}_n} w_{ij} D_i D_j \gamma_j + \varepsilon_i.$$

Under this model, the unit-level direct and indirect effects are respectively given by

$$Y_i(1,\boldsymbol{D}_{-i}) - Y_i(0,\boldsymbol{D}_{-i}) = \gamma_i + \beta_i + \sum_{j\neq i} w_{ij} D_j \gamma_j \quad \text{and} \quad Y_i(0,\boldsymbol{D}_{-i}) - Y_i(\mathbf{0}) = \sum_{j\neq i} w_{ij} D_j \beta_j.$$

Due to the choice of -5 in the spatial weights w_{ij}, Assumption 3 holds for $\gamma = 3$ (Leung, 2022, eq. (3)), which is a fairly slow rate of decay given that Assumption 3 requires $\gamma > 2$.

We present results for the indirect and total effects using 5000 simulation draws where within each draw, we redraw potential outcomes and recompute the design-based estimand. In Table 1, the "Spatial Interference" columns correspond to the outcome model described above, whereas the "Partial Interference" columns redefine $w_{ij} = 0$ if *i, j* lie in different clusters to eliminate cross-cluster interference. The "CI" rows report the coverage of 95-percent CIs using the indicated standard errors. For our estimator, the "SE" row corresponds to standard errors obtained from our variance estimator (4), while for difference in means, it corresponds

to conventional cluster-robust SEs. The "SE*" rows are the true superpopulation standard errors obtained by taking the standard deviation of the estimator across the simulation draws. As such, SE should be consistent for SE* rather than conservative, while the CIs have asymptotically conservative coverage for the design-based estimand. Finally, row "% Excl" is the percentage of units that are not well surrounded.

The results are consistent with the theory. As *n* grows, the biases of our estimators shrink, while coverage tends to or exceeds the nominal level. The bias of difference in means is more than twice that of our estimators under spatial interference, resulting in severe undercoverage even with the true superpopulation SEs. While the SEs are smaller than those of our estimators, this is not by a significant amount, and the efficiency advantage comes at a large cost to bias under spatial interference.

Table 2 presents results for our estimator under spatial interference, but we multiply (2) by $c \in \{0.8, 1.2\}$. This is to assess robustness and illustrate a bias-variance trade-off in the choice of r_n. The $c = 0.8$ columns show that this results in about half the proportion of excluded units relative to Table 1. As a result, the bias is higher, resulting in undercoverage. The $c = 1.2$ columns show that the proportion of excluded units a little less than doubles. The bias is lower, and as a result, the probability of coverage is higher. On the other hand, the variance increases, as can be seen in the standard error columns.

7 Empirical Application

We apply our estimator to data from the unconditional cash transfer experiment mentioned in Example 2. In the experiment, households eligible for the transfer (the treatment) live in homes with thatched roofs, which is a proxy for poverty. Houses are grouped in villages, which are grouped in "sublocations." See Figure A.2 of Egger et al. (2022) for a map of the study area. The CRT randomizes sublocations (the clusters) into treatment with probability $q = 0.5$. Within treatment (control) sublocations, villages are randomized into treatment with probability $p_1 = 2/3$ ($p_0 = 1/3$). Within treated villages, all eligible households receive cash transfers totaling 1000 USD, which is about 75 percent of average annual household spending. Sublocations contain 7.8 villages on average (SD 3.9).

Egger et al. (2022) study the effect of the cash transfers on the following household-level outcomes, which can be grouped into three categories: (1) annualized expenditures on consumption (of food and other purchases described in their footnote 31), non-durables, food alone, temptation goods, and durables; (2) asset stocks, housing value, and land value; and (3) annualized household income, transfers, taxes paid, profits, and wage earnings. The authors report model-assisted estimates of causal effects at the household level. Our analysis will be entirely design-based but at the village level. We aggregate household outcomes to the village level by averaging.

Egger et al. (2022) estimate the effects of the transfers on the population of eligible households using two main specifications described in their §3.2. Their "RF" (reduced form) specification is an OLS regression of an outcome on household-level and sublocation-level treatment indicators and covariates. They report the coefficient on the household-level indicator, which corresponds to a model-assisted estimate of the direct effect $\theta^*(1, 0, 2/3, 2/3)$ for eligible households.

However, the authors note the potential for interference across sublocations (their quote in Example 2). For this reason, their preferred specification is the following "IV" (instrumental variables) regression. The main regressors are the amount of cash transferred per capita to the household's village and the amount transferred to neighboring villages within a band of $r-2$ to *r* km from the ego's village for a range of *r* values. The corresponding instruments are respectively an indicator for the ego's village being treated and the share of eligible households in the band assigned to treatment. They use a BIC criterion to select the maximum range of *r*, which is 2 km. Using these estimates, the authors compute a model-assisted estimate of the total effect $\theta^*(1,0,2/3,0)$. This compares saturation levels of two-thirds and zero, which is nonparametrically unidentified since control villages have a saturation level of one-third. We instead report our estimates of the total effect $\theta^*(1,0,2/3,1/3)$.

Table 3 reports the results for a subset of the outcomes, and Table SA.4.1 in the supplementary appendix reports the remainder. Both tables choose r_n in $\hat{\theta}$ according to (2), which results in $r_n = 1.6$ and 39.66 percent of units not well surrounded. In §SA.4 we describe how we construct cluster radii R_j used in this formula. The $\hat{\theta}^+$ columns correspond to difference-in-means estimates with clustered standard errors.

We find that the difference-in-means estimates of the direct and total effects are comparable to the RF and IV estimates, respectively, despite the distinctions outlined above. Our estimators find larger direct and total effects. Decomposing the total effect into direct and indirect effects, we find that the latter are substantially smaller in magnitude with large standard errors. Compared to difference in means, our estimates tend to be larger in magnitude but with larger standard errors due to the restriction to well-surrounded units.

Table SA.4.2 in the supplementary appendix reports results for $r_n = 2$, resulting 57.89 percent of units not being well surrounded. This choice of r_n coincides with the largest range of *r* selected by the BIC procedure of Egger et al. (2022). Our estimates and standard errors become larger still in magnitude relative to difference in means, but the results are qualitatively similar.

To estimate spillover effects, Egger et al. (2022) rerun their IV specification using only non-eligible households, which did not receive any transfers (their §3.3). In Table SA.4.3 of the supplementary appendix, we compare their results with design-based estimates of the overall effect $\theta^*(\varnothing,\varnothing,2/3,1/3)$ on the population of non-eligibles. The effect sizes of our estimators and those of difference in means are fairly similar in magnitude to their IV results, but the standard errors are large. Combined with the results in Table 3, we ultimately find strong direct effects of the cash transfers but weaker evidence for spillover effects compared to Egger et al. (2022).

8 Conclusion

When interference occurs across clusters, conventional analyses of CRTs suffer from bias induced by units near cluster boundaries. To reduce bias at the analysis stage, we provide in § 3.2 an estimator $\hat{\theta}$ that improves upon the fried-egg design by excluding from estimation units that are not surrounded by clusters assigned to the same treatment arm. This may be employed as a robustness check for difference in means. To reduce bias at the design stage, we propose a rate-optimal formula for the number of clusters *k* in (9). Unlike the standard power analysis that assumes partial interference, our choice balances power against the need to reduce bias and accounts for the density of units in the spatial region. Given *k*, we suggest automating cluster construction using *k*-medoids and prove that the resulting clusters are balanced and globular, thereby approximately minimizing the number of units near boundaries. We also provide valid design-based standard errors.

Under the conventional superpopulation, partial interference framework, power calculations for choosing *k* require prior knowledge of the intracluster correlation coefficient (ICC). Under our design-based framework, the optimal choice of *k* requires knowledge of γ, the speed at which interference decays with distance, rather than the ICC. Absent domain knowledge, one can conservatively set γ to the dimension of the spatial region. We discuss in § 5.2 how to obtain less conservative estimates via modeling or prior data.

9 Acknowledgements

I thank Dennis Egger for generously providing supplemental distance data used in the empirical application. I also thank Forrest Crawford and seminar participants at Texas A&M and the UCSC brown bag for helpful comments.

10 Disclosure Statement

The author reports there are no competing interests to declare.

References

Alexander, N., A. Lenhart, and K. Anaya-Izquierdo, "Spatial Spillover Analysis of a Cluster-Randomized Trial Against Dengue Vectors in Trujillo, Venezuela," *PLoS Neglected Tropical Diseases*, 2020, *14* (9), e0008576.

Andrews, D., "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 1991, *59* (3), 817–858.

Baird, S., J A. Bohren, C. McIntosh, and B. Özler, "Optimal Design of Experiments in the Presence of Interference," *Review of Economics and Statistics*, 2018, *100* (5), 844–860.

Binka, F., F. Indome, and T. Smith, "Impact of Spatial Distribution of Permethrin-Impregnated Bed Nets on Child Mortality in Rural Northern Ghana," *The American Journal of Tropical Medicine and Hygiene*, 1998, *59* (1), 80–85.

Cao, J., C. Hansen, D. Kozbur, and L. Villacorta, "Inference for Dependent Data with Learned Clusters," *Review of Economics and Statistics (forthcoming)*, 2024.

CDC, "Life Cycle of *Anopheles* Species Mosquitoes," 2025. https://www.cdc.gov/mosquitoes/about/life-cycle-of-anopheles-mosquitoes.html, accessed 6/7/2025.

Egger, D., J. Haushofer, E. Miguel, P. Niehaus, and M. Walker, "General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya," *Econometrica*, 2022, *90* (6), 2603–2643.

Faridani, S. and P. Niehaus, "Rate-Optimal Linear Estimation of Average Global Effects," *arXiv preprint arXiv:2209.14181*, 2024.

Guerra, C., R. Reiner, T. Perkins, S. Lindsay, J. Midega et al., "A Global Assembly of Adult Female Mosquito Mark-Release-Recapture Data to Inform the Control of Mosquito-Borne Pathogens," *Parasites & Vectors*, 2014, *7*, 1–15.

Halloran, M., K. Auranen, S. Baird, N. Basta, S. Bellan et al., "Simulations for Designing and Interpreting Intervention Trials in Infectious Diseases," *BMC Medicine*, 2017, *15* (1), 1–8.

Hansen, B. and S. Lee, "Asymptotic Theory for Clustered Samples," *Journal of Econometrics*, 2019, *210* (2), 268–290.

Hayes, R. and L. Moulton, *Cluster Randomised Trials*, CRC press, 2017.

Hemming, K., A. Girling, A. Sitch, J. Marsh, and R. Lilford, "Sample Size Calculations for Cluster Randomised Controlled Trials with a Fixed Number of Clusters," *BMC Medical Research Methodology*, 2011, *11* (1), 1–11.

—— , S. Eldridge, G. Forbes, C. Weijer, and M. Taljaard, "How to Design Efficient Cluster Randomised Trials," *BMJ*, 2017, *358*.

Homan, T., A. Hiscox, C. Mweresa, D. Masiga, W. Mukabana et al., "The Effect of Mass Mosquito Trapping on Malaria Transmission and Disease Burden (SolarMal): A Stepped-Wedge Cluster-Randomised Trial," *The Lancet*, 2016, *388* (10050), 1193–1201.

Hudgens, M. and M. Halloran, "Toward Causal Inference with Interference," *Journal of the American Statistical Association*, 2008, *103* (482), 832–842.

Imai, K., G. King, and C. Nall, "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation," *Statistical Science*, 2009, *24* (1), 29–53.

Jarvis, C., G. Di Tanna, D. Lewis, N. Alexander, and J. Edmunds, "Spatial Analysis of Cluster Randomised Trials: A Systematic Review of Analysis Methods," *Emerging Themes in Epidemiology*, 2017, *14*, 1–9.

—— , L. Multerer, D. Lewis, F. Binka, W. Edmunds, N. Alexander, and T. Smith, "Spatial Effects of Permethrin-Impregnated Bed Nets on Child Mortality: 26 Years on, a Spatial Reanalysis of a Cluster Randomized Trial," *The American Journal of Tropical Medicine and Hygiene*, 2019, *101* (6), 1434.

Jenish, N. and I. Prucha, "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields," *Journal of Econometrics*, 2009, *150* (1), 86–98.

Karrer, B., L. Shi, M. Bhole, M. Goldman, T. Palmer, C. Gelman, M. Konutgan, and F. Sun, "Network Experimentation at Scale," in "Proceedings of the 27th ACM Sigkdd Conference on Knowledge Discovery & Data Mining" 2021, pp. 3106–3116.

Lahiri, S. and J. Zhu, "Resampling Methods for Spatial Regression Models Under a Class of Stochastic Designs," *Annals of Statistics*, 2006, *34* (4), 1774–1813.

Leung, M., "Rate-Optimal Cluster-Randomized Designs for Spatial Interference," *The Annals of Statistics*, 2022, *50* (5), 3064–3087.

—— , "Identifying Treatment and Spillovers Effects Using Exposure Contrasts," *arXiv preprint arXiv:2403.08183*, 2024.

McCann, R., H. van den Berg, W. Takken, A. Chetwynd, E. Giorgi et al., "Reducing Contamination Risk in Cluster-Randomized Infectious Disease-Intervention Trials," *International Journal of Epidemiology*, 2018, *47* (6), 2015–2024.

Moulton, L., K. O'Brien, R. Kohberger, I. Chang, R. Reid, R. Weatherholtz, J. Hackell, G. Siber, and M. Santosham, "Design of a Group-Randomized Streptococcus Pneumoniae Vaccine Trial," *Controlled Clinical Trials*, 2001, *22* (4), 438–452.

Multerer, L., T. Glass, F. Vanobberghen, and T. Smith, "Analysis of Contamination in Cluster Randomized Trials of Malaria Interventions," *Trials*, 2021, *22* (1), 1–17.

Muralidharan, K. and P. Niehaus, "Experimentation at Scale," *Journal of Economic Perspectives*, 2017, *31* (4), 103–124.

O'Hara, L., N. Blanco, S. Leekha, K. Stafford, G. Slobogean, E. Ludeman, and A. Harris, "Design, Implementation, and Analysis Considerations for Cluster-Randomized Trials in Infection Control and Hospital Epidemiology: A Systematic Review," *Infection Control and Hospital Epidemiology*, 2019, *40* (6), 686–692.

Schochet, P., N. Pashley, L. Miratrix, and T. Kautz, "Design-Based Ratio Estimators and Central Limit Theorems for Clustered, Blocked RCTs," *Journal of the American Statistical Association*, 2022, *117* (540), 2135–2146.

Staples, P., E. Ogburn, and J. Onnela, "Incorporating Contact Network Structure in Cluster Randomized Trials," *Scientific Reports*, 2015, *5* (1), 17581.

Sur, D., R. Ochiai, S. Bhattacharya, N. K Ganguly, M. Ali, B. Manna, S. Dutta, A. Donner, S. Kanungo, J. Park et al., "A Cluster-Randomized Effectiveness Trial of Vi Typhoid Vaccine in India," *New England Journal of Medicine*, 2009, *361* (4), 335–344.

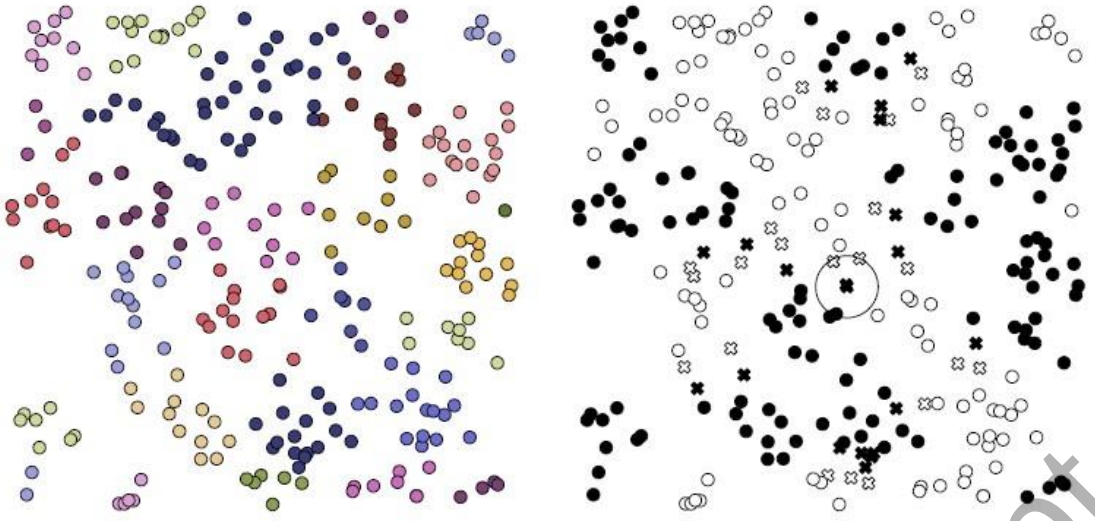

Figure 1: k -medoid clusters. The left panel colors units by cluster membership and the right panel by whether their respective clusters are assigned to treatment (black) or control (white). Units marked by "X" are excluded from estimation. The circle depicts a unit's r_n -neighborhood used to determine exclusion.

Table 1: Main results

	Spatial Interference						Partial Interference					
	Indirect Effect			Overall Effect			Indirect Effect			Overall Effect		
n	500	1000	2000	500	1000	2000	500	1000	2000	500	1000	2000
Our Estimator												
Bias	0.064	0.055	0.051	0.072	0.059	0.057	0.000	0.003	0.005	0.003	0.004	0.004
CI SE	0.945	0.951	0.954	0.940	0.949	0.948	0.951	0.958	0.966	0.952	0.960	0.964
CI SE*	0.956	0.952	0.948	0.956	0.959	0.952	0.962	0.961	0.959	0.961	0.964	0.960
SE	0.258	0.202	0.160	0.336	0.263	0.208	0.254	0.197	0.156	0.329	0.256	0.201
SE*	0.261	0.200	0.153	0.342	0.262	0.206	0.257	0.196	0.150	0.333	0.253	0.198
Difference in Means												
Bias	0.154	0.179	0.216	0.168	0.192	0.233	0.001	0.001	0.005	0.004	0.001	0.004
CI SE	0.906	0.852	0.689	0.904	0.872	0.770	0.953	0.961	0.970	0.951	0.960	0.962
CI SE*	0.920	0.853	0.654	0.935	0.894	0.784	0.960	0.963	0.963	0.963	0.963	0.960
SE	0.244	0.186	0.144	0.319	0.246	0.191	0.240	0.182	0.140	0.316	0.242	0.188
SE*	0.250	0.185	0.137	0.327	0.246	0.191	0.245	0.181	0.134	0.321	0.240	0.186
% Excl	5.035	7.816	11.25	5.035	7.816	11.25	5.035	7.816	11.25	5.035	7.816	11.25
r_n	1.395	1.518	1.623	1.395	1.518	1.623	1.395	1.518	1.623	1.395	1.518	1.623
$\hat{\theta}$	1.518	1.770	2.089	3.406	3.723	4.119	1.432	1.650	1.927	3.288	3.563	3.907
k	63	100	159	63	100	159	63	100	159	63	100	159

Table 2: Robustness results

	$c = 0.8$						$c = 1.2$					
	Indirect Effect			Overall Effect			Indirect Effect			Overall Effect		
n	500	1000	2000	500	1000	2000	500	1000	2000	500	1000	2000
Bias	0.092	0.085	0.083	0.102	0.092	0.090	0.044	0.035	0.033	0.050	0.038	0.036
CI	0.932	0.935	0.936	0.933	0.940	0.932	0.946	0.951	0.957	0.940	0.952	0.953
CI SE*	0.944	0.935	0.929	0.952	0.950	0.943	0.960	0.960	0.954	0.959	0.961	0.957
SE	0.252	0.195	0.152	0.330	0.256	0.201	0.267	0.213	0.171	0.346	0.274	0.219
SE*	0.256	0.194	0.146	0.335	0.255	0.200	0.272	0.212	0.164	0.353	0.274	0.218
% Excl	2.300	3.800	5.700	2.300	3.800	5.700	9.300	13.600	18.500	9.300	13.600	18.500
r_n	1.116	1.215	1.299	1.116	1.215	1.299	1.674	1.822	1.948	1.674	1.822	1.948
$\hat{\theta}_Q$	1.490	1.740	2.057	3.376	3.690	4.085	1.538	1.790	2.108	3.427	3.744	4.139
k	63	100	159	63	100	159	63	100	159	63	100	159

Table 3: Effects on eligibles

	Direct Effect		Indirect Effect		Total Effect		Egger et al.	
	$\hat{\theta}$	$\hat{\theta}^{+}$	$\hat{\theta}$	$\hat{\theta}^{+}$	$\hat{\theta}$	$\hat{\theta}^{+}$	RF	IV
consumption	390.70	253.93	9.51	64.43	400.21	318.37	293.59	338.57
	(109.19)	(76.32)	(155.56)	(103.07)	(158.01)	(102.18)	(60.11)	(109.38)
non-durable	248.71	151.99	43.18	61.08	291.90	213.07	187.65	227.2
	(104.20)	(66.39)	(139.26)	(96.22)	(140.41)	(94.55)	(58.59)	(99.63)
assets	249.58	180.43	-35.18	-11.43	214.40	168.99	178.78	183.38
	(59.26)	(39.32)	(104.75)	(66.00)	(98.33)	(59.98)	(24.66)	(44.26)
housing	422.82	376.36	18.77	31.02	441.60	407.38	376.92	477.29
	(49.74)	(31.12)	(93.55)	(51.73)	(86.81)	(48.74)	(26.37)	(38.8)
income	132.89	95.02	64.90	21.01	197.80	116.03	79.43	135.7
	(120.95)	(58.97)	(159.60)	(91.71)	(154.70)	(87.36)	(43.8)	(92.1)
earnings	59.35	45.70	22.31	6.46	81.67	52.16	42.43	73.66
	(78.90)	(37.67)	(123.04)	(66.82)	(119.93)	(60.62)	(32.23)	(60.82)

653 villages (units), 84 sublocations (clusters). Standard errors are in parentheses. Column RF (IV) is the reduced form (IV) estimate of the overall effect from Table I, column 1 (2) of Egger et al. (2022), $\hat{\theta}$ is our estimate, and $\hat{\theta}^{+}$ is difference in means. Our estimates use $r_n = 1.6$, which results in 39.66 percent of units not being well surrounded.