Title: Inclusion of Unexposed Clusters Improves the Precision of Fixed Effects Analysis of Stepped-Wedge Cluster Randomized Trials

Short Title: Unexposed Clusters in Fixed Effects Analysis of Stepped-Wedge CRTs

Author List: Kenneth Menglin Lee,[1]* Xiangmei Ma,[1] Grace Meijuan Yang,[2,3] Yin Bun Cheung[1,4,5]

[1] Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore 169857

[2] Division of Supportive and Palliative Care, National Cancer Centre Singapore, Singapore 169610

[3] Lien Centre for Palliative Care, Duke-NUS Medical School, Singapore 169857

[4] Signature Programme in Health Services & Systems Research, Duke-NUS Medical School, Singapore 169857

[5] Tampere Center for Child, Adolescent and Maternal Health Research, Tampere University, Tampere 33520, Finland

*Corresponding author. Center for Quantitative Medicine, Duke-NUS Medical School, Singapore, 20 College Road, Singapore 169857

E-mail: klee@u.duke.nus.edu.

Phone: +65 8950 7699

Abstract

Stepped-Wedge Cluster Randomized Trials (SW-CRTs) are typically analyzed using mixed effects models. The fixed effects model is a useful alternative that controls for all time-invariant cluster-level confounders and has proper control of type I error when the number of clusters is small. In principle, all clusters in SW-CRTs are designed to eventually receive the intervention, but in real-world research, some trials can end with unexposed clusters (clusters that never received the intervention), such as when a trial is terminated early based on interim analysis results. Typically, unexposed clusters are expected to contribute no information to the fixed effects intervention effect estimator and are excluded from fixed effects analyses. In this article we mathematically prove that inclusion of unexposed clusters improves the precision of the fixed effects least squares dummy variable (LSDV) intervention effect estimator, re-analyze data from a recent SW-CRT of a novel palliative care intervention containing an unexposed cluster, and evaluate the methods by simulation. We found that including unexposed clusters improves the precision of the fixed effects LSDV intervention effect estimator in both real and simulated datasets. Our simulations also reveal an increase in power and decrease in root mean square error. These improvements are present even if the assumptions of constant residual variance and period effects are violated. In the case that a SW-CRT concludes with unexposed clusters, these unexposed clusters can be included in the fixed effects LSDV analysis to improve precision, power, and root mean square error.

Keywords: stepped wedge trials; cluster randomized trials; fixed effects model; precision

1. Introduction

In a cluster randomized trial (CRT), randomization is carried out on the cluster-level rather than the individual-level. The stepped-wedge cluster randomized trial (SW-CRT) is a specific type of CRT where clusters begin the trial unexposed to the intervention and are randomized to start the intervention across different periods or "steps". The crossover is uni-directional and continues until all clusters are exposed to the intervention.[1,2] The logistical advantage of the phased implementation makes the SW-CRT an increasingly popular design.[1,3] Due to its unique features as a uni-directional crossover CRT, different statistical models have been proposed and applied to SW-CRT data.[3]

The most popular statistical model for cross-sectional SW-CRT designs, where each individual is only observed during one period, is the "Hussey and Hughes" mixed effects model.[3] This mixed effects model treats clusters as random and periods as fixed effects.[2] Extensions to the Hussey and Hughes mixed effects model have been proposed and applied over the past few years with new models adjusting for cluster-by-period random interaction effects,[4,5] between-period correlation that decays over time,[6] and more.[3] Alternatively, previous studies have also treated clusters as fixed effects in a fixed effects least squares dummy variable (LSDV) model.[7–11] While different mixed effects models with additional random effects terms have been proposed, developed, and studied in the context of SW-CRTs,[3] there have been comparatively few methodological developments in fixed effects methods for stepped-wedge designs.

The Hussey & Hughes mixed effects model is often preferred over the fixed effects LSDV model due to reduced standard error (SE) for the intervention effect estimates.[12,13] Mixed effects models, in general, treat clusters as random effects that are uncorrelated with both the residual error term and other model covariates.[14] If the cluster random effect terms are correlated with the other model covariates, as is the case when there are unmeasured cluster-level time-invariant confounders, the mixed effects intervention effect estimator becomes biased, inconsistent, and fails to remove the confounding.[12,14–17] Since mixed effects models rely on randomization to control for known and unknown confounders, the benefits of randomization may be lost when the number of clusters is small, making it difficult to balance cluster characteristics.[18,19]

In contrast, the fixed effects LSDV model treats clusters as fixed effects using dummy variables and estimates the intervention effect using ordinary least squares (OLS).[12,17] Accordingly, the fixed effects LSDV model estimates the intervention effect using within-cluster comparisons and controls for all cluster-level time-invariant confounders.[12] Therefore, a major distinction between modelling clusters as random or fixed depends on whether such confounders may exist.[15]

Furthermore, modelling clusters as random effects in a mixed effects model tends to lead to inflated type I error rates and overly narrow confidence intervals for the intervention effect estimates when the number of clusters is small.[13,20,21] Due to real world constraints, it is not uncommon for SW-CRTs to have such low numbers of clusters.[22] This inflated type I error rate is not observed in fixed effects models, making it an attractive alternative.[13]

Previous studies have elected to use a fixed effects LSDV model to analyze data collected from a SW-CRT, citing difficulties that arise from small number of clusters,[9,23] practical and logistical issues that prevented randomization,[24] and concerns over confounding between cluster and outcomes.[8] Under such conditions, the fixed effects model has been an effective complementary approach to the more widely adopted mixed effects models for analyzing SW-CRTs.

A potential drawback of the fixed effects LSDV model is its inability to estimate coefficients for variables that have no within-cluster variation.[12] In principle, all clusters in a SW-CRT spend periods unexposed and exposed to the intervention over the duration of the trial. Therefore, all clusters are designed to have within-cluster variation in intervention status. In reality, however, some clusters may end the trial without having received the intervention. For example, a SW-CRT of seasonal malaria chemoprevention was stopped following interim analysis and clusters that were randomized to receive the intervention at later periods ended the trial unexposed to the intervention.[25] In Section 3, we will introduce and discuss a motivating and illustrative case study that encountered an unexposed cluster due to hospital management restructuring.

Fixed effects analyses are often referred to as only making "within-unit comparisons"[26] where "only covariates that vary within-subjects at the observational level should be used in the model",[15] and "cases that do not change either (1) do not contribute much information to the analysis or (2) are altogether omitted by design".[27] Under such phrasing and guidance, one may have the impression that an unexposed cluster would not contribute meaningfully to the fixed effects intervention effect estimate. While these authors[15,27] did not explicitly state whether unexposed units should or should not be included in the fixed effects analysis, others have explicitly stated that "Comparisons are made within individuals [units] rather than between individuals [units]… Thus, only those who have experienced both the outcome and the exposure of interest are included".[28]

However, research on the conditional Poisson model for drug safety assessment, which relies on within-subject comparisons, has shown that the inclusion of unexposed subjects provides useful information about time-varying covariates and reduces confounding by these covariates.[29] Ma, Lam and Cheung further show that the inclusion of unexposed subjects in the conditional Poisson model improves the precision of the exposure effect estimator when the analysis adjusts for time-varying covariates.[30] Likewise, models for SW-CRTs need to adjust for periods as time-varying covariates. Drawing on this similarity, we hypothesize that including clusters that are never exposed to the intervention (hereon referred to as "unexposed clusters") in a fixed effects LSDV analysis of cross-sectional SW-CRT data will improve the precision of the intervention effect estimator.

In Section 2, we mathematically prove that the inclusion of unexposed clusters in the analysis of a SW-CRT design increases the precision of the fixed effects LSDV intervention effect estimator. In Section 3, we illustrate the benefits of including unexposed clusters in a fixed effects LSDV model by re-analyzing a SW-CRT of a novel palliative care model that had four exposed clusters and one unexposed cluster. In Section 4, we conduct extensive simulations to

assess the impact of including unexposed clusters on the fixed effects LSDV intervention effect estimator in terms of precision, bias, coverage probability, power, type I error, and root mean square error. In Section 5 we end with some concluding remarks.

2. Analysis Model and Precision

We begin with a 5-cluster, 5-period SW-CRT design (Figure 1) based on the motivating case study to be described in Section 3. Let $n_{i,j}$ be the number of individuals in the i^{th} cluster ($i = 1, 2, 3, 4, 5$) during the j^{th} period ($j = 1, 2, 3, 4, 5$). Note that individuals in Cluster 5 are never exposed to the intervention (an unexposed cluster).

The outcome of the SW-CRT is modelled using the fixed effects LSDV model:

$$Y_{ijk} = \delta Z_{ij} + \sum_{p=2}^{5} \phi_j \, \mathrm{I}_{[j=p]} + \sum_{c=1}^{5} \alpha_i \mathrm{I}_{[i=c]} + e_{ijk}$$

where Y_{ijk} is the health outcome of the k^{th} individual ($k = 1, \dots, n_{i,j}$) in the i^{th} cluster ($i = 1, \dots, 5$) and j^{th} period ($j = 1, \dots, 5$), δ is the intervention effect, Z_{ij} is the intervention indicator for the i^{th} cluster during the j^{th} period ($Z_{ij} = 1$, if exposed to intervention, $Z_{ij} = 0$ otherwise), ϕ_j is the fixed effect for the categorical j^{th} period ($\phi_1 = 0$ for identifiability), α_i is the fixed effect for the i^{th} cluster, $\mathrm{I}_{[j=p]}$ and $\mathrm{I}_{[i=c]}$ are dummy variables for periods and clusters, respectively, and e_{ijk} is the residual error assumed to be independently and identically distributed with variance σ^2.

Writing the model in the matrix form:

$$Y = X\beta + e$$

where X is an $N \times 10$ design matrix, with N being the total number of study participants, and $\beta = (\delta, \phi_2, .., \phi_5, \alpha_1, .., \alpha_5)'$. Since all variables in the model are dummy coded, the entire matrix X is composed of 0's and 1's with column 1 containing data on the intervention status, columns 2 to 5 on whether a participant is in period 2 to 5, and columns 6 to 10 on whether a participant is in cluster 1 to 5. For example, a participant from cluster 2 who received the intervention in period 2 would contribute a row of (1, 1, 0, 0, 0, 0, 1, 0, 0, 0).

Using OLS, the intervention effect estimator $\hat{\delta}$ is the first element in the vector of coefficients:

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'Y,$$

where $X'X$ is a matrix product assumed to be positive definite. Given the OLS variance-covariance matrix:

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(X'X)^{-1}.$$

The variance of the intervention effect estimator $\mathrm{Var}(\hat{\delta})$ is $\mathrm{Var}(\widehat{\boldsymbol{\beta}})_{1,1}$ and $\mathrm{Precision}(\hat{\delta}) = 1/\mathrm{Var}(\hat{\delta})$. Since OLS assumes constant σ^2, $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) \propto (X'X)^{-1}$.

Specifically, to distinguish different ways of using the data collected from the unexposed cluster, let $\check{X}$ be the $\check{N} \times 10$ design matrix for the fixed effects LSDV analysis containing observations from all 5 periods in all 5 clusters (including unexposed Cluster 5), and $\dot{X}$ be the $\dot{N} \times 9$ design matrix containing observations from all 5 periods in the 4 exposed clusters only (Clusters 1 to 4), where $\check{N}$ and $\dot{N}$ are the total sample sizes in the two models, respectively. Furthermore, let $\tilde{X}$ be the $\tilde{N} \times 10$ design matrix for the analysis containing observations from all 5 periods in the 4 exposed clusters and observations from only period 1 of the unexposed Cluster 5. The construction of this design matrix $\tilde{X}$ with a sample of size $\tilde{N}$ serves two purposes. First, it helps demonstrate that unless the unexposed cluster provides information for estimating period effects, which requires observations from at least two periods, its inclusion will not improve the precision of the intervention effect estimate beyond analyzing only the exposed clusters with $\dot{N}$ observations. Second, as will be seen later, it facilitates the proof by providing a same-sized product matrix for comparison with the product matrix based on the sample of $\check{N}$ observations from all clusters and all periods. Assuming $n_{5,1} > 0$ and $\sum_{j=2}^{5} n_{5,j} > 0$, $\check{N} > \tilde{N} > \dot{N}$.

Then, $\check{X}'\check{X}$, $\dot{X}'\dot{X}$ and $\tilde{X}'\tilde{X}$ are the 10×10, 9×9 and 10×10 product matrices explicitly defined in terms of $n_{i,j}$ in the Online Supplementary Material S1a, S1b, and S1c, respectively. The similarities and differences in the elements of these three product matrices will be detailed in Sections 2.1 and 2.2. Furthermore, let $\check{\delta}$, $\dot{\delta}$, and $\tilde{\delta}$ be the least-square estimators of intervention effect based on the three models.

The proof that inclusion of the unexposed cluster reduces the variance and increases the precision of the fixed effects LSDV intervention effect estimator $\check{\delta}$ proceeds in two steps:

1. We demonstrate that the variance of the intervention effect estimator $\mathrm{Var}(\dot{\delta})$ obtained from $\dot{X}'\dot{X}$ (exposed clusters only) is equal to $\mathrm{Var}(\tilde{\delta})$ obtained from $\tilde{X}'\tilde{X}$ (exposed clusters and one period in the unexposed cluster):

$$\mathrm{Var}(\dot{\delta}) = \sigma^2(\dot{X}'\dot{X})^{-1}_{1,1} = \sigma^2(\tilde{X}'\tilde{X})^{-1}_{1,1} = \mathrm{Var}(\tilde{\delta}).$$

2. We demonstrate that the variance of the intervention effect estimator $\mathrm{Var}(\check{\delta})$ obtained from $\check{X}'\check{X}$ (all clusters, all periods) is smaller than $\mathrm{Var}(\tilde{\delta})$ obtained from $\tilde{X}'\tilde{X}$:

$$\mathrm{Var}(\check{\delta}) = \sigma^2(\check{X}'\check{X})^{-1}_{1,1} < \sigma^2(\tilde{X}'\tilde{X})^{-1}_{1,1} = \mathrm{Var}(\tilde{\delta}).$$

Altogether, we prove $\mathrm{Var}(\check{\delta}) < \mathrm{Var}(\dot{\delta})$.

2.1 Proof of $\mathbf{Var}(\dot{\boldsymbol{\delta}}) = \mathbf{Var}(\tilde{\boldsymbol{\delta}})$

Since $\tilde{X}'\tilde{X}$ contains observations from the four exposed clusters and only period 1 of the unexposed Cluster 5, the upper-left block of the 9×9 elements of $\tilde{X}'\tilde{X}$ (Online Supplementary Material S1c) is equal to $\dot{X}'\dot{X}$ (Online Supplementary Material S1b), where $\dot{X}'\dot{X}$ can interpreted as a submatrix of $\tilde{X}'\tilde{X}$:

$$\tilde{X}'\tilde{X} = \begin{pmatrix} \dot{X}'\dot{X} & 0 \\ 0 & n_{5,1} \end{pmatrix}.$$

The blockwise inverted matrix $(\tilde{X}'\tilde{X})^{-1}$ can then be defined as:

$$(\tilde{X}'\tilde{X})^{-1} = \begin{pmatrix} \dot{X}'\dot{X} & 0 \\ 0 & n_{5,1} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} (\dot{X}'\dot{X})^{-1} & 0 \\ 0 & 1/n_{5,1} \end{pmatrix}.$$

Therefore, $(\tilde{X}'\tilde{X})^{-1}_{1,1} = (\dot{X}'\dot{X})^{-1}_{1,1}$ and on the OLS model assumption of constant σ^2:

$$\mathrm{Var}(\tilde{\delta}) = \sigma^2(\tilde{X}'\tilde{X})^{-1}_{1,1} = \sigma^2(\dot{X}'\dot{X})^{-1}_{1,1} = \mathrm{Var}(\dot{\delta}).$$

2.2 Proof of $\mathbf{Var}(\breve{\delta}) < \mathbf{Var}(\tilde{\delta})$

For brevity, we present a slightly abbreviated proof of $\mathrm{Var}(\breve{\delta}) < \mathrm{Var}(\tilde{\delta})$. The complete proof can be found in the Online Supplementary Material S2.

First, we represent $\breve{X}'\breve{X}$ (Online Supplementary Material S1a) and $\tilde{X}'\tilde{X}$ in terms of submatrices:

$$\breve{X}'\breve{X} = \begin{pmatrix} \breve{A} & \breve{B}' \\ \breve{B} & \breve{D} \end{pmatrix},$$

$$\tilde{X}'\tilde{X} = \begin{pmatrix} \tilde{A} & \tilde{B}' \\ \tilde{B} & \tilde{D} \end{pmatrix}.$$

Submatrix $\breve{A}$ is a scalar that equals the dot product of the first row of $\breve{X}'$ by the first column of $\breve{X}$. Since the first column of $\breve{X}$ indicates the intervention status of each participant, $\breve{A}$ equals the total number of trial participants who receive the intervention.

Submatrix $\breve{B}$ is a column vector of length 9 that equals the dot product of all rows but the first of $\breve{X}'$ (containing dummy variables for the four period and five cluster effects) by the first column of $\breve{X}$ (indicating intervention status). Therefore, $\breve{B}$ equals the marginal numbers of participants who receive the intervention in each period and each cluster. Since Cluster 5 is unexposed, the marginal number of participants who receive the intervention under the two study designs are the same, $\breve{A} = \tilde{A}$ and $\breve{B} = \tilde{B}$.

The 9×9 submatrix $\breve{D}$ equals the matrix product of the rows of $\breve{X}'$ and columns of $\breve{X}$ that represent dummy variables for the periods and clusters.

With these submatrices, the blockwise inverted matrix $(\breve{X}'\breve{X})^{-1}$ is defined as:

$$(\check{X}'\check{X})^{-1} = \begin{pmatrix} (\check{A} - \check{B}'\check{D}^{-1}\check{B})^{-1} & -(\check{A} - \check{B}'\check{D}^{-1}\check{B})^{-1}\check{B}'\check{D}^{-1} \\ -\check{D}^{-1}\check{B}(\check{A} - \check{B}'\check{D}^{-1}\check{B})^{-1} & \check{D}^{-1} + \check{D}^{-1}\check{B}(\check{A} - \check{B}'\check{D}^{-1}\check{B})^{-1}\check{B}'\check{D}^{-1} \end{pmatrix},$$

where:

$$\mathrm{Var}(\check{\delta}) = \sigma^2(\check{X}'\check{X})^{-1}_{1,1} = \sigma^2(\check{A} - \check{B}'\check{D}^{-1}\check{B})^{-1}.$$

The blockwise inverted matrix $(\tilde{X}'\tilde{X})^{-1}$ and $\mathrm{Var}(\tilde{\delta})$ are similarly defined.

Given that $\check{X}'\check{X}$ and $\tilde{X}'\tilde{X}$ are same-sized positive definite matrices ($\check{X}'\check{X} \succ 0$ and $\tilde{X}'\tilde{X} \succ 0$), the principal submatrices $\check{D}$ and $\tilde{D}$ are also positive definite matrices ($\check{D} \succ 0$ and $\tilde{D} \succ 0$).[31] $\check{D}$ and $\tilde{D}$ are explicitly defined in terms of $n_{i,j}$ in the Online Supplementary Material S1d.

The difference between $\check{D}$ and $\tilde{D}$ is positive semi-definite, where $x'(\check{D} - \tilde{D})x \geq 0$ for all x in $\mathbb{R}^9$,[31] as proven in the Online Supplementary Material S2. Therefore, we can order the submatrices as induced by Loewner partial ordering[31]:

$$\check{D} \succcurlyeq \tilde{D},$$

and:

$$\check{D}^{-1} \preccurlyeq \tilde{D}^{-1}.$$

Given that vector $\check{B} = \tilde{B}$ (here on referred to as $\mathbb{B}$) and scalar $\check{A} = \tilde{A}$ (here on referred to as $\mathbb{A}$):

$$\left(\mathbb{A} - \mathbb{B}'\check{D}^{-1}\mathbb{B}\right)^{-1} \leq \left(\mathbb{A} - \mathbb{B}'\tilde{D}^{-1}\mathbb{B}\right)^{-1}.$$

Therefore, on the OLS assumption of constant σ^2:

$$\mathrm{Var}(\check{\delta}) \leq \mathrm{Var}(\tilde{\delta}).$$

Furthermore, in Online Supplementary Material S2, we provide a proof by contradiction revealing that $\mathrm{Var}(\check{\delta}) \neq \mathrm{Var}(\tilde{\delta})$, therefore $\mathrm{Var}(\check{\delta}) < \mathrm{Var}(\tilde{\delta})$.

Combining the proofs in Sections 2.1 and 2.2, we demonstrate that $\mathrm{Var}(\check{\delta}) < \mathrm{Var}(\dot{\delta})$, where the inclusion of the unexposed cluster improves precision of the fixed effects LSDV intervention effect estimator. While this proof is demonstrated with a 5-cluster, 5-period SW-CRT design, the proof's utilization of submatrices implies that this result applies to SW-CRT designs with any number of exposed and unexposed clusters. A simple extension of the proof while maintaining the standard OLS assumption of equal variance demonstrates that the effect of including an always-exposed cluster (a cluster that never receives the control condition) on the precision of the intervention effect estimator is equivalent to the effect of including an unexposed cluster (Online Supplementary Material S3).

3. Case Study

We re-analyzed data from a recent SW-CRT comparing a standard palliative care delivery model (control) against a novel co-rounding model (intervention) for hospital inpatients

with cancer.[32] In the standard care model, oncologists conducted the daily ward rounds and referred patients to the palliative care department if considered appropriate. In the novel co-rounding model, oncologists and palliative care specialists jointly conducted the ward rounds and initiated palliative care as per their consensus. The primary endpoint for this trial was hospital length of stay (LOS). It was hypothesized that LOS would be reduced under the novel co-rounding model.

The study was initially planned as a 4-cluster SW-CRT with 5 four-month periods. The clusters were different oncology teams in the Singapore General Hospital. However, due to the restructuring of hospital management, a fifth oncology team was developed and deployed by the hospital. This happened after randomization of the original 4 clusters and just before study initiation. Since this fifth cluster was not formally part of the trial or randomization process, it implemented the standard care model for the entire duration of the trial. As a result, this study resembles the 5-cluster, 5-period SW-CRT design described above (Section 2).

The original publication on the efficacy of the novel co-rounding model used a fixed effects LSDV model to account for the cluster effects and control for confounding.[32] The fixed effects LSDV model was used due to the difficulty in controlling for confounding (e.g. the clusters differed not only in patient characteristics but also in unmeasured clinician characteristics) and concerns on the robustness of applying the mixed effects model to a SW-CRT with such a small number of clusters.[13,20] The analysis did not include the unexposed cluster because it was not part of the original trial plan and it was not yet methodologically clear what the implications of including an unexposed cluster in a fixed effects LSDV model were.

We re-analyzed the data using the fixed effects LSDV model, with and without the unexposed cluster. For the present purpose, we only kept the first admission if a participant was admitted more than once over the study duration. In total, there were 3462 admissions. We analyzed ln(LOS) as the outcome variable. As shown in Figure 2, which pooled data from all 5 clusters, ln(LOS) was approximately normally distributed under both co-rounding and standard care models, but LOS was not.

The pooled analysis was only sensible if the underlying pattern of period effects were expected to be the same between the exposed and unexposed clusters, an assumption of the fixed effects LSDV model.[33] There are currently no standard practices for evaluating this assumption. As such, we applied multiple methods and confirmed that they gave consistent findings. First, we tested for equivalent period effects between the unexposed cluster and exposed clusters using an ANOVA to compare a typical fixed effects LSDV model against a fixed effects LSDV model with interaction terms between the unexposed cluster and period indicators. The ANOVA showed no evidence of difference, with $P = 0.302$ (on 4 degrees of freedom). We also estimated the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) of the models excluding and including the interaction terms (AIC: 9524.341 and 9527.461, respectively. BIC: 9591.987 and 9619.705, respectively). Both favored the model that excluded the interaction terms and assumed same period effects between exposed and unexposed clusters.

Table 1 shows the analysis results. We estimated the intervention effect among all cancer patients or among only stage III and stage IV cancer patients, who were expected to be the primary beneficiaries of palliative care. Without including the unexposed cluster, the standard errors of the intervention effect estimates were 0.073 and 0.077 among all patients and stage III/IV patients, respectively. In contrast, the inclusion of the unexposed cluster in the fixed effects LSDV analysis yielded ~12% smaller standard errors and narrower 95% Wald Confidence Intervals (95% CI), both in the analyses of all patients and only stage III/IV patients (Table 1). The inclusion of the unexposed cluster also altered the point estimate, from -0.101 to -0.126 among all patients. Despite this change, the 95% confidence intervals of the analyses with and without the unexposed cluster continued to overlap substantially. The difference of 0.025 between the point estimates was small compared to the standard errors both with and without the unexposed cluster. A similar pattern was seen among stage III/IV patients. Additional analyses of all patients and only stage III/IV patients from the exposed clusters and only period 1 of the unexposed cluster (resembling the $\tilde{X}$ design matrix described in Section 2) yielded intervention effect estimates and standard errors identical to the corresponding analyses of the exposed clusters only.

4. Simulation

4.1 Simulation Settings

Realistic simulation parameters were generated with reference to the results from the case study of ln(LOS) under different palliative care models (Section 3). We examined scenarios where there were 4, 8, or 12 exposed clusters and 0, 1, 2, or 3 unexposed clusters. Accordingly, the total number of clusters in each simulation scenario was C = 4+0, 4+1, 4+2, 4+3, 8+1, …, 12+3. All scenarios had a fixed total of 5 periods (4 steps) with either 1, 2, or 3 clusters crossing from control to intervention at each step, depending on the number of exposed clusters (4, 8, or 12) (Figure 3)

4.1.1 Simulation based on the Hussey and Hughes Mixed Effects Model

We simulated SW-CRT data based on the Hussey and Hughes mixed effects model[2]:

$$Y_{ijk} = \delta Z_{ij} + \phi_j + \alpha_i + e_{ijk}$$

for the k^{th} individual ($k = 1, \dots, n_{i,j}$) in the i^{th} cluster ($i = 1, \dots, C$) and j^{th} period ($j = 1, \dots, 5$), where Y_{ijk} was a continuous outcome, δ was the intervention effect, Z_{ij} was the intervention indicator for the i^{th} cluster during the j^{th} period ($Z_{ij} = 1$, if exposed to intervention, $Z_{ij} = 0$ otherwise), ϕ_j was the fixed effect for the j^{th} period ($\phi_1 = 0$ for identifiability), $\alpha_i \sim Normal(\mu, \tau_\alpha^2)$ was the random effect for the i^{th} cluster, and $e_{ijk} \sim Normal(0, \sigma_e^2)$ was the residual independent of α_i.

We generated the true intervention effects δ equal to 0, -0.1, and -0.2, and a linear period effect equal to an increase of 0.1 per period. We also considered scenarios where the true period effects $\phi_j = 0$ (for all j) to examine the effects of including unexposed clusters when period effects are absent from the underlying data generating process. We generated residual

error e_{ijk} by setting $e_{ijk} \sim Normal(0, 1)$. To simulate cluster effects, we set the between-cluster variance τ_α^2 to $1/99$, $1/19$, and $1/9$, to generate corresponding intracluster correlation (ICC) values of 0.01, 0.05, and 0.1, where ICC $= \frac{\tau_\alpha^2}{\tau_\alpha^2 + \sigma_e^2}$.[2]

For each of these scenarios, we used a range of different cluster sample sizes n_i for each period in the i^{th} cluster, where $n_i \sim Gamma(k, \theta)$ with $k = 30$, 100, and 300, and $\theta = 1$. This produces an average size of 30, 100 or 300. Realized sample sizes $n_{i,j}$ for the i^{th} cluster during the j^{th} period were subsequently generated with $n_{i,j} \sim Poisson(n_i)$, so the sample size could vary between periods within a cluster. The trial's total sample size was $N = \sum_i \sum_j n_{i,j}$.

In total, 324 scenarios were investigated (3 # of exposed clusters × 4 # of unexposed clusters × 3 intervention effect sizes × 3 values of τ_α^2 × 3 cluster sample sizes).

4.1.2 Misspecification with Non-constant Residual Variance across Clusters

The fixed effects LSDV model assumes constant residual variance. To assess the model robustness and impact of including unexposed clusters in the presence of model misspecification, we simulated SW-CRT data by extending the mixed effects model in Section 4.1.1 to allow the standard deviation of the residuals to vary across clusters. The residual standard deviation in the i^{th} cluster, $\sigma_{e,i}$, was set to follow a gamma distribution, $\sigma_{e,i} \sim Gamma\left(k = \text{CV}^2, \theta = 1/\text{CV}^2\right)$, with $E\left[\sigma_{e,i}\right] = 1$ and coefficient of variation CV $= 0.1$, 0.5, or 1.

For the purposes of these simulations, we fixed the true intervention effects δ to -0.1, between-cluster variance τ_α^2 to $1/19$, and the average cluster size $E[n_i]$ to 100. In total, 36 scenarios were investigated (3 # of exposed clusters × 4 # of unexposed clusters × 3 CVs for cluster-specific residual standard deviation $\sigma_{e,i}$).

4.1.3 Misspecification with Non-constant Period Effects across Clusters

We evaluated whether varying the period effects between clusters affects the impact of including unexposed clusters on the precision and other properties of the fixed effects LSDV estimator. We varied period effects between clusters by simulating SW-CRT data based on the "Hooper-Girling" mixed effects model[4,5]:

$$Y_{ijk} = \delta Z_{ij} + \phi_j + \alpha_i + \gamma_{ij} + e_{ijk}$$

Along with the cluster random effect α_i and period fixed effect ϕ_j, we simulated a cluster-by-period random interaction effect, $\gamma_{ij} \sim Normal\left(0, \tau_\gamma^2\right)$. We set τ_γ^2 to 0, $1/76$, and $2/57$, to generate corresponding realistic cluster autocorrelation (CAC) values of 1, 0.8, 0.6, where CAC $= \frac{\tau_\alpha^2}{\tau_\alpha^2 + \tau_\gamma^2}$.[4,34,35]

As in Section 4.1.2, we fixed the true intervention effects δ to -0.1, between-cluster variance τ_α^2 to $1/19$, and the average cluster size $E[n_i]$ to 100. In total, 36 scenarios were investigated (3 # of exposed clusters × 4 # of unexposed clusters × 3 values of τ_γ^2).

4.1.4 Analysis of Simulated data

In each simulation scenario, we generated $s = 10{,}000$ simulated data sets and estimated the intervention effect $\hat{\delta}_s$ using the fixed effects LSDV model. We present the properties of the intervention effect estimator in terms of bias, precision, power, coverage probability (CP), and root mean squared error (RMSE). We present the absolute bias ($\text{Abs Bias} = [\sum_{s=1}^{10{,}000} \hat{\delta}_s/10{,}000] - \delta$) when $\delta = 0$, and the relative bias ($\text{Rel Bias} = [\text{Absolute bias}/\delta] \times 100$) when $\delta \neq 0$. Precision is the reciprocal of the average estimated variance ($\text{Precision} = 1/[\sum_{s=1}^{10{,}000} \text{Var}(\hat{\delta}_s)/10{,}000]$). Power is the empirical power (if $\delta < 0$) or empirical type I error rate (if $\delta = 0$) for rejecting the null hypothesis of $\delta \geq 0$ at the one-sided significance level of 0.05. CP is the probability that the 95% confidence interval contains the true effect. RMSE is the square root of the average squared difference between the estimated effect $\hat{\delta}_s$ and the true effect δ over the 10,000 simulated data sets for each scenario $\left(\text{RMSE} = \sqrt{\sum_{s=1}^{10{,}000}[\hat{\delta}_s - \delta]^2 /10{,}000}\right)$. The Monte Carlo standard errors (standard deviation of the 10,000 estimated intervention effects $\hat{\delta}_s$ for each scenario) are included in the Online Supplementary Material S4.

4.2 Simulation Results

4.2.1 Simulation based on the Hussey and Hughes Mixed Effects Model

Figure 4 shows the simulation results in the scenarios with true intervention effect $\delta = -0.1$ and between-cluster variance $\tau_\alpha^2 = 1/19$. Among the scenarios considered, the inclusion of unexposed clusters generally increased the precision, power, and decreased the RMSE of the fixed effects LSDV intervention effect estimator. Within the range considered, the more unexposed clusters included, the better the improvement. Overall, the fixed effects LSDV method yielded practically unbiased estimates with CP of 95% confidence interval (0.945-0.955) close to the nominal level. The inclusion of unexposed clusters had no impact on these two properties.

The results across all values of δ ($\delta = 0, -0.1, -0.2$) and τ_α^2 ($\tau_\alpha^2 = 1/99\,, 1/19\,, 1/9$) were qualitatively similar to those reported above, details in the Online Supplementary Material S5. Scenarios when $\delta = 0$ all maintained near 5% type I error rates, regardless of the inclusion of unexposed clusters (Online Supplementary Material S5). Additionally, the benefits of including up to 3 unexposed clusters are still observed when there are large numbers of randomized clusters across 5 periods (4 steps), although the relative improvements in precision were milder (Online Supplementary Material S6).

Similar to Figure 4, Figure 5 shows the simulation results in the scenarios with true intervention effect $\delta = -0.1$, between-cluster variance $\tau_\alpha^2 = 1/19$, but where the true period effects $\phi_j = 0$ (for all j). These simulation results reveal that the improvements from including unexposed cases in a fixed effects LSDV model (adjusting for period effects) are unaffected by the absence of period effects in the underlying data generating process (Figure 5).

4.2.2 Misspecification with Non-constant Residual Variance across Clusters

Figure 6 shows the properties of the intervention effect estimator when the true intervention effect $\delta = -0.1$, the between-cluster variance $\tau_\alpha^2 = 1/19$, the average cluster sample size $E[n_i] = 100$, and CV of the cluster-specific residual standard deviation = 0.1, 0.5 or 1.0.

Despite violating OLS assumptions, the inclusion of unexposed clusters generally increased the precision, power, and decreased the RMSE of the fixed effects LSDV intervention effect estimator (Figure 6). Within the range considered, the more unexposed clusters included, the better the improvement. These improvements are observed across different amounts of heterogeneity in the residual standard deviation $\sigma_{e,i}$.

Relative bias and CP were largely unaffected by the inclusion of unexposed clusters. Furthermore, increasing the CV of the residual standard deviation had little effect on the relative bias and CP of the fixed effects LSDV intervention effect estimates (Figure 6). However, increasing the CV of the residual standard deviation resulted in reduced precision and power. This occurred regardless of the inclusion of unexposed clusters, and the inclusion of unexposed clusters did not make it worse. On the contrary, the inclusion of unexposed clusters improved precision, power, and RMSE despite variable cluster-specific residual standard deviation $\sigma_{e,i}$.

4.2.3 Misspecification with Non-constant Period Effects across Clusters

Figure 7 shows the properties of the intervention effect estimator when the true intervention effect $\delta = -0.1$, the between-cluster variance $\tau_\alpha^2 = 1/19$, the average cluster sample size $E[n_i] = 100$, and $\tau_\gamma^2 = 0$, $1/76$, or $2/57$.

Despite the random cluster-by-period interaction effects γ_{ij}, the inclusion of unexposed clusters increased the precision, power, and decreased the RMSE of the intervention effect estimator (Figure 7). Across different amounts of τ_γ^2 and number of unexposed clusters considered, the more unexposed clusters included, the better the improvement.

Figure 7 also reveals that the magnitude of τ_γ^2 had little effect on the relative bias, precision, and power of the intervention effect estimators. However, CP decreased and RMSE increased substantially in relation to increase in τ_γ^2. Nevertheless, this occurred regardless of whether unexposed clusters were included in the analysis and the inclusion of unexposed clusters did not worsen the CP.

5. Discussion

Unexposed clusters (clusters never exposed to intervention) are seen in SW-CRTs in real-world research.[25,32] For example, SW-CRTs that are terminated early can result in clusters that are never exposed to intervention. SW-CRTs with unexposed clusters somewhat resemble the "optimal design" discussed by Thompson et al.[36] and Girling and Hemming.[4] However, the analysis of the "optimal design" is based on the mixed effects model which, in addition to horizontal within-cluster comparisons, also makes vertical within-period comparisons. This differs from the use of fixed effects LSDV (Least Squares Dummy Variable) model.

The fixed effects LSDV model for estimating the intervention effect in a SW-CRT design uses within-cluster comparisons and controls for all unmeasured cluster-level time-invariant confounders.[12] The more commonly used mixed effects models, in contrast, are unable to control for these confounders[12,14–17] and struggle with inflated type I error rates when the number of clusters is small.[13,20] However, a potential drawback of the fixed effects LSDV model is its inability to estimate coefficients for variables that have no within-cluster variation.[12] Some articles recommend that clusters should only be modelled as fixed if they are the only clusters that exist or of interest, otherwise the variability of the cluster effects will be underestimated and results may not be generalizable to unsampled clusters.[37] This concern is largely irrelevant in the context of SW-CRTs where the primary interest is to estimate the coefficient of the intervention effect rather than cluster effects.[26] Furthermore, clusters are often selected due to practical reasons instead of randomly sampled.[24] In such cases, it is appropriate to consider them the only clusters of interest and use the fixed effects LSDV model.

Previous publications on fixed effects and other methods that make within-unit comparisons imply that unexposed units (in this case, clusters) should not be included in the analysis.[12,15,28] The SW-CRT of palliative care models that motivated this work did not include the unexposed cluster in its original analysis due to the lack of methodological guidance at the time.[32]

In this paper, we demonstrated that the inclusion of unexposed clusters in a fixed effects LSDV model is a viable strategy for improving the precision of the intervention effect estimate in a SW-CRT design. We mathematically proved this in Section 2. Furthermore, we re-analyzed data from the SW-CRT of a novel palliative care model in Section 3 and conducted simulations of a variety of realistic scenarios in Section 4. We found that including unexposed clusters improves the precision, power, and decreases the RMSE of the fixed effects LSDV intervention effect estimator. These improvements are unaffected by the absence of true period effects ($\phi_j = 0$, for all j) in the data generating process and persist provided the fixed effects LSDV model adjusts for period effects as per standard practice. Furthermore, these improvements persist even if the OLS assumptions of constant residual variance and period effects are violated. It is straightforward to generalize the results here, on the inclusion of unexposed clusters, to the inclusion of always-exposed clusters (Online Supplementary Material S3). This may happen, for example, when intervention is provided to a small number of clusters from the beginning to pilot the logistics of distributing the intervention in a large-scale study.[25]

Unexposed clusters may be included in the analysis if they arise from the same study population or share the same underlying pattern of period effects as the exposed clusters. In the example of a trial that was terminated early following the results of an interim analysis, the unexposed clusters were clusters randomized to receive the intervention in the last period. Here, the unexposed clusters certainly arise from the same study population and there is strong motivation to include them in the analysis. Alternatively, in the palliative care trial case study, reorganization of hospital management introduced an additional fifth cluster that ran concurrently alongside the other study clusters to serve the existing target population and reduce the patient load of some of the original clusters.[32] There was no expansion of study population

and no change to the case-mix. In this case, we consider it appropriate to include the fifth cluster as an unexposed cluster in our fixed effects analysis. In contrast, if an unexposed cluster arises from expansion of service coverage, one should exercise caution and carefully consider the comparability of the new clusters.

The fixed effects LSDV model operates under the assumption that exposed and unexposed clusters have the same underlying pattern of period effects.[33] Currently, there are no standard practices for testing this assumption in the context of fixed effects analyses for SW-CRTs. In the palliative care case study, we elected to test these assumptions using an ANOVA, AIC, and BIC to determine whether including interaction terms between an unexposed cluster and period effects improves the model fit. Future research can explore the operating characteristics of these or other methods for the detection of differences in period effects, and whether a level of difference that is difficult to detect with these methods can cause any material bias in the point estimation.

In the palliative care case study, the inclusion of the unexposed cluster was shown to reduce the standard error and change the magnitude of the intervention effect estimate. Given that we did not detect any differences in the pattern of period effects between exposed and unexposed clusters, the inclusion of unexposed clusters is not expected to affect the point estimate of the intervention effect more than by chance. Although in this case study, the relative change of the point estimates when including the unexposed cluster was substantial (about 25%), it could be the result of the small absolute value of the point estimate as the denominator in calculating the relative change. Nevertheless, the absolute difference of the point estimates when including versus excluding unexposed clusters was small compared to the level of uncertainty in the point estimates as reflected by the standard errors. Notably, our simulations showed that including unexposed clusters did not affect the unbiasedness of the point estimate of the fixed effects LSDV analysis, even with non-constant period effects between clusters in the data generating process.

However, our simulation results also revealed that non-constant period effects between clusters resulted in low coverage probability of the fixed effects LSDV intervention effect estimate. The fixed effects LSDV is a useful model for SW-CRTs when cluster-level confounding is suspected but is susceptible to additional heterogeneity in the period effect. Thompson et al.[21] pointed out that the Hussey and Hughes mixed effects model is also susceptible to misspecification of the period effect and instead recommends a non-parametric vertical within-period estimator when period effect varies between clusters.[21] The final decision as to which model to use ultimately depends on the conditions of the SW-CRT and where the sources of heterogeneity and confounding lie.

The CONSORT extension to SW-CRTs suggests including the intraclass correlation coefficient (ICC) estimates in results reporting to help inform future studies.[38] Unlike mixed effects models, the fixed effects LSDV model is unable to automatically estimate the ICC between clusters.[17] However, there are alternative ways to estimate the ICC alongside a fixed effects analysis. For example, researchers can estimate the ICC through a one-way ANOVA of cluster effects during the first time-period of the SW-CRT when all clusters are still

unexposed.[39,40] The use of fixed effects analysis for the estimation of the intervention effect does not preclude the estimation of ICC which can be accomplished with additional analyses.

In this paper, we consider the fixed effects LSDV model for the analysis of continuous outcomes. Recent work by Ma, Lam and Cheung shows that inclusion of unexposed subjects in the conditional Poisson model, which is equivalent to a fixed effects Poisson model,[41] improves the precision of the exposure effect estimator when the analysis adjusts for time-varying covariates.[30] It is intuitive to assume by extension that the inclusion of unexposed clusters will also improve the precision of the fixed effects intervention effect estimate in SW-CRTs with discrete and binary outcomes. Future research is needed to confirm this.

Until now, we have restricted our attention to cluster randomized trials. However, the concepts and findings here are applicable to other research settings. Examples include observational studies of child health, which are susceptible to confounding by family socio-economic status. As such, fixed effects models have been proposed to study the impact of risk factors that are variable between children within families and remove confounding by unmeasured family-level covariates.[42] If all children in a family are exposed (or unexposed) to the measured risk factors, they contribute no information to fixed effects models that do not involve adjustment for child-level covariates. However, such analyses often require adjustment for child-level covariates, such as age. Based on the observations in the present study, we expect that the inclusion of families that contain only children exposed (or unexposed) to the measured risk factors may result in a more precise fixed effects exposure effect estimate when adjusting for time-varying covariates. The magnitude of precision gained in such situations requires further investigation.

In conclusion, cross-sectional SW-CRTs may conclude with clusters that are never or always exposed to the intervention of interest in scenarios where a fixed effects LSDV analysis may be preferred. In such cases, it is preferable to include these clusters in the fixed effects LSDV analysis due to the improvements in precision, power, and RMSE of the intervention effect estimator.

Disclosure statement

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The methodological work of YBC was supported by the National Medical Research Council, Singapore (MOH-000526).

Disclaimer

Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of Ministry of Health / National Medical Research Council, Singapore.

ORCID

Kenneth Menglin Lee https://orcid.org/0000-0002-0454-4537

Yin Bun Cheung http://orcid.org/0000-0003-0517-7625

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study. R codes for the simulations will be deposited to figshare by Wiley.

References

1. Hayes RJ, Moulton LH. *Cluster Randomised Trials*. Second edition. CRC Press; 2017.

2. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*. 2007;28(2):182-191. doi:10.1016/j.cct.2006.05.007

3. Li F, Hughes JP, Hemming K, Taljaard M, Melnick ER, Heagerty PJ. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res*. 2021;30(2):612-639. doi:10.1177/0962280220932962

4. Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Stat Med*. 2016;35(13):2149-2166. doi:10.1002/sim.6850

5. Hooper R, Teerenstra S, Hoop E de, Eldridge S. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Stat Med*. 2016;35(26):4718-4728. doi:10.1002/sim.7028

6. Kasza J, Hemming K, Hooper R, Matthews J, Forbes A. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res*. 2019;28(3):703-716. doi:10.1177/0962280217734981

7. Cowan JF, Micek M, Cowan JFG, et al. Early ART initiation among HIV-positive pregnant women in central Mozambique: a stepped wedge randomized controlled trial of an optimized Option B+ approach. *Implement Sci*. 2015;10(1):61. doi:10.1186/s13012-015-0249-6

8. Craine N, Whitaker R, Perrett S, Zou L, Hickman M, Lyons M. A stepped wedge cluster randomized control trial of dried blood spot testing to improve the uptake of hepatitis C antibody testing within UK prisons. *Eur J Public Health*. 2015;25(2):351-357. doi:10.1093/eurpub/cku096

9. Kelly PJ, Baker AL, Deane FP, et al. Study protocol: a stepped wedge cluster randomised controlled trial of a healthy lifestyle intervention for people attending residential substance abuse treatment. *BMC Public Health*. 2015;15(1):465. doi:10.1186/s12889-015-1729-y

10. Mouchoux C, Rippert P, Duclos A, et al. Impact of a multifaceted program to prevent postoperative delirium in the elderly: the CONFUCIUS stepped wedge protocol. *BMC Geriatr*. 2011;11(1):25. doi:10.1186/1471-2318-11-25

11. van den Broek IV, Hoebe CJ, van Bergen JE, et al. Evaluation design of a systematic, selective, internet-based, Chlamydiascreening implementation in the Netherlands, 2008-2010: implications of first results for the analysis. *BMC Infect Dis*. 2010;10(1):89. doi:10.1186/1471-2334-10-89

12. Allison PD. *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. SAS Press; 2005.

13. Barker D, D'Este C, Campbell MJ, McElduff P. Minimum number of clusters and comparison of analysis methods for cross sectional stepped wedge cluster randomised trials with binary outcomes: A simulation study. *Trials*. 2017;18(1):1-11. doi:10.1186/s13063-017-1862-2

14. Hausman JA. Specification Tests in Econometrics. *Econometrica*. 1978;46(6):1251-1271. doi:10.2307/1913827

15. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: What are the differences? *Stat Med*. 2009;28(2):221-239. doi:https://doi.org/10.1002/sim.3478

16. Gunasekara FI, Richardson K, Carter K, Blakely T. Fixed effects analysis of repeated measures data. *Int J Epidemiol*. 2014;43(1):264-269. doi:10.1093/ije/dyt221

17. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. 2. ed. MIT Press; 2010.

18. Taljaard M, Teerenstra S, Ivers NM, Fergusson DA. Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. *Clin Trials*. 2016;13(4):459-463. doi:10.1177/1740774516634316

19. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2-21. doi:10.1016/j.socscimed.2017.12.005

20. Leyrat C, Morgan KE, Leurent B, Kahan BC. Cluster randomized trials with a small number of clusters: which analyses should be used? *Int J Epidemiol*. 2018;47(1):321-331. doi:10.1093/ije/dyx169

21. Thompson JA, Davey C, Fielding KL, Hargreaves JR, Hayes RJ. Robust analysis of stepped wedge trials using cluster-level summaries within periods. *Stat Med*. 2018;37(16):2487-2500.

22. Barker D, McElduff P, D'Este C, Campbell MJ. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. *BMC Med Res Methodol*. 2016;16(1):69. doi:10.1186/s12874-016-0176-5

23. The CIPHER Investigators, Williamson A, Redman S, et al. Supporting Policy In health with Research: an Intervention Trial (SPIRIT)--protocol for a stepped wedge trial. *BMJ Open*. 2014;4(7):e005293-e005293. doi:10.1136/bmjopen-2014-005293

24. Groshaus H, Boscan A, Khandwala F, Holroyd-Leduc J. Use of Clinical Decision Support to Improve the Quality of Care Provided to Older Hospitalized Patients. *Appl Clin Inform*. 2012;03(01):94-102. doi:10.4338/ACI-2011-08-RA-0047

25. Cissé B, Ba EH, Sokhna C, et al. Effectiveness of Seasonal Malaria Chemoprevention in Children under Ten Years of Age in Senegal: A Stepped-Wedge Cluster-Randomised Trial. *PLoS Med*. 2016;13(11). doi:10.1371/journal.pmed.1002175

26. Allison PD. *Fixed Effects Regression Models*.; 2009.

27. Hill TD, Davis AP, Roos JM, French MT. Limitations of Fixed-Effects Models for Panel Data. *Sociol Perspect*. 2020;63(3):357-369. doi:10.1177/0731121419863785

28. Forbes H, Douglas I, Finn A, et al. Risk of herpes zoster after exposure to varicella to explore the exogenous boosting hypothesis: self controlled case series study using UK electronic healthcare data. *BMJ*. Published online January 22, 2020:l6987. doi:10.1136/bmj.l6987

29. Musonda P, Hocine MN, Whitaker HJ, Farrington CP. Self-controlled case series analyses: Small-sample performance. *Comput Stat Data Anal*. 2008;52(4):1942-1957. doi:10.1016/j.csda.2007.06.016

30. Ma X, Lam KF, Cheung YB. Inclusion of unexposed subjects improves the precision and power of self-controlled case series method. *J Biopharm Stat*. Published online November 15, 2021:1-10. doi:10.1080/10543406.2021.1998099

31. Horn RA, Johnson CR. *Matrix Analysis*. Second edition, corrected reprint. Cambridge University Press; 2017.

32. Yang GM, Zhou S, Xu Z, et al. Comparing the effect of a consult model versus an integrated palliative care and medical oncology co-rounding model on health care utilization in an acute hospital – an open-label stepped-wedge cluster-randomized trial. *Palliat Med*. 2021;35(8):1578-1589. doi:10.1177/02692163211022957

33. Vaisey S, Miles A. What You Can—and Can't—Do With Three-Wave Panel Data. *Sociol Methods Res*. 2017;46(1):44-67. doi:10.1177/0049124114547769

34. Hemming K, Kasza J, Hooper R, Forbes A, Taljaard M. A tutorial on sample size calculation for multiple-period cluster randomized parallel, cross-over and stepped-wedge trials using the Shiny CRT Calculator. *Int J Epidemiol*. 2020;49(3):979-995. doi:10.1093/ije/dyz237

35. Hooper R, Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ*. 2015;350:h2925. doi:10.1136/bmj.h2925

36. Thompson JA, Fielding K, Hargreaves J, Copas A. The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. *Clin Trials*. 2017;14(6):639-647. doi:10.1177/1740774517723921

37. Donner A, Klar N. Pitfalls of and Controversies in Cluster Randomization Trials. *Am J Public Health*. 2004;94(3):416-422.

38. Hemming K, Taljaard M, McKenzie JE, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*. 2018;363:k1614. doi:10.1136/bmj.k1614

39. Campbell MK, Grimshaw JM, Elbourne DR. Intracluster correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. *BMC Med Res Methodol*. 2004;4(1):9. doi:10.1186/1471-2288-4-9

40. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428. doi:10.1037/0033-2909.86.2.420

41. Xu S, Zeng C, Newcomer S, Nelson J, Glanz J. Use of Fixed Effects Models to Analyze Self-Controlled Case Series Data in Vaccine Safety Studies. *J Biom Biostat*. Published online 2012:1-13.

42. Cheung YB, Lam KF. Three estimates of the association between linear growth failure and cognitive ability. *Trop Med Int Health*. 2009;14(9):1020-1024. doi:10.1111/j.1365-3156.2009.02321.x

Analysis	Patients	$\hat{\delta}$	SE	95% CI
− unexposed cluster	All	-0.101	0.073	(-0.243, 0.042)
+ unexposed cluster	All	-0.126	0.064	(-0.251, -0.001)
− unexposed cluster	Stage III/IV	-0.112	0.077	(-0.262, 0.038)
+ unexposed cluster	Stage III/IV	-0.141	0.067	(-0.273, -0.009)

Table 1. Intervention effect estimates $\hat{\delta}$ in analyses excluding (−) and including (+) the unexposed cluster in a fixed effects LSDV analysis of data from a SW-CRT testing a novel co-rounding model of palliative care. The intervention effect $\hat{\delta}$, standard errors, and 95% Wald Confidence Intervals were estimated for all cancer patients or Stage III/IV cancer patients only.

	P1	P2	P3	P4	P5	
C1	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,4}$	$n_{1,5}$	$\sum_{j=1}^{5} n_{1,j}$
C2	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	$n_{2,4}$	$n_{2,5}$	$\sum_{j=1}^{5} n_{2,j}$
C3	$n_{3,1}$	$n_{3,2}$	$n_{3,3}$	$n_{3,4}$	$n_{3,5}$	$\sum_{j=1}^{5} n_{3,j}$
C4	$n_{4,1}$	$n_{4,2}$	$n_{4,3}$	$n_{4,4}$	$n_{4,5}$	$\sum_{j=1}^{5} n_{4,j}$
C5	$n_{5,1}$	$n_{5,2}$	$n_{5,3}$	$n_{5,4}$	$n_{5,5}$	$\sum_{j=1}^{5} n_{5,j}$
	$\sum_{i=1}^{5} n_{i,1}$	$\sum_{i=1}^{5} n_{i,2}$	$\sum_{i=1}^{5} n_{i,3}$	$\sum_{i=1}^{5} n_{i,4}$	$\sum_{i=1}^{5} n_{i,5}$	N

Figure 1. A SW-CRT design with 4 exposed clusters and 1 unexposed cluster across 5 periods (4 steps). Cluster-period cross-sections that receive the intervention are shaded in gray; $n_{i,j}$ is the sample size for the i^{th} cluster and j^{th} period.

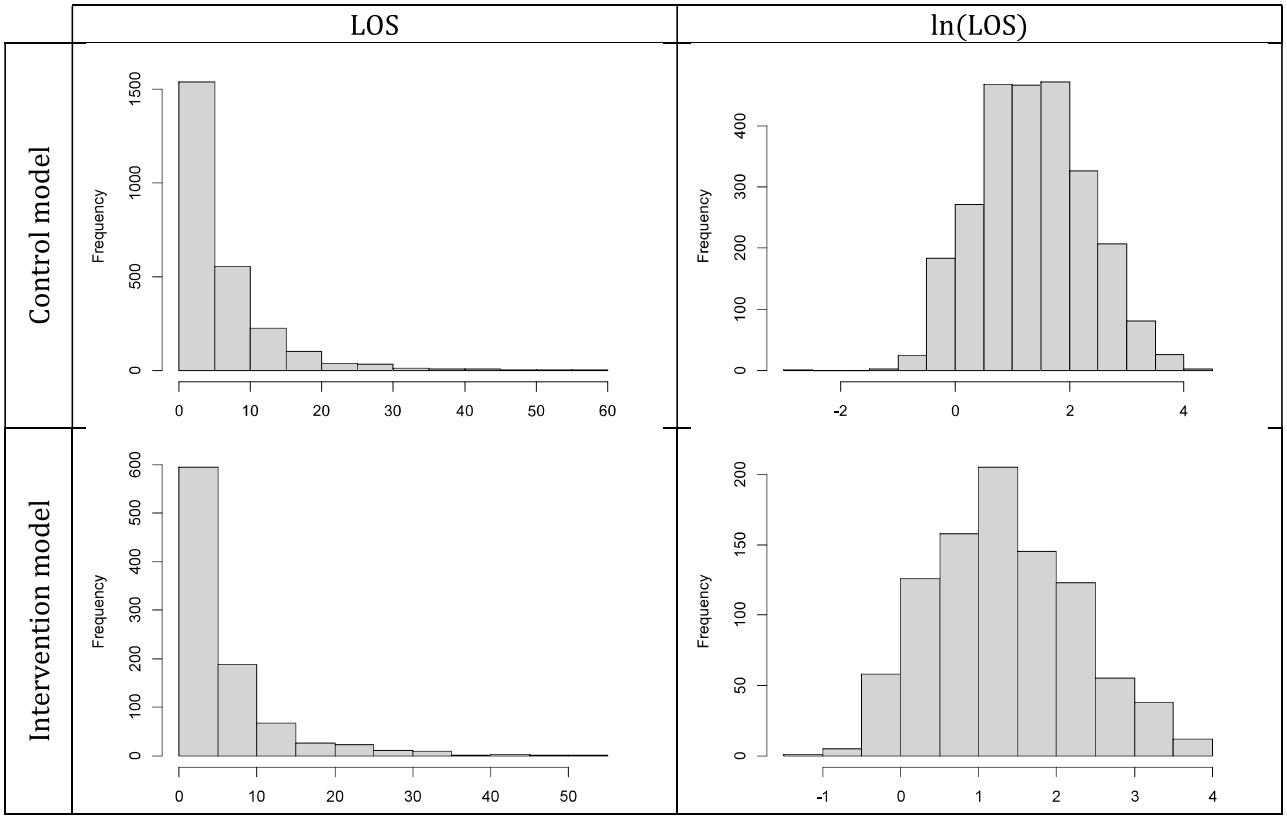

Figure 2. Distribution of hospital length of stay (LOS) and ln(LOS) for cancer patients who received either the standard palliative care delivery model (control) or a novel co-rounding model (intervention).

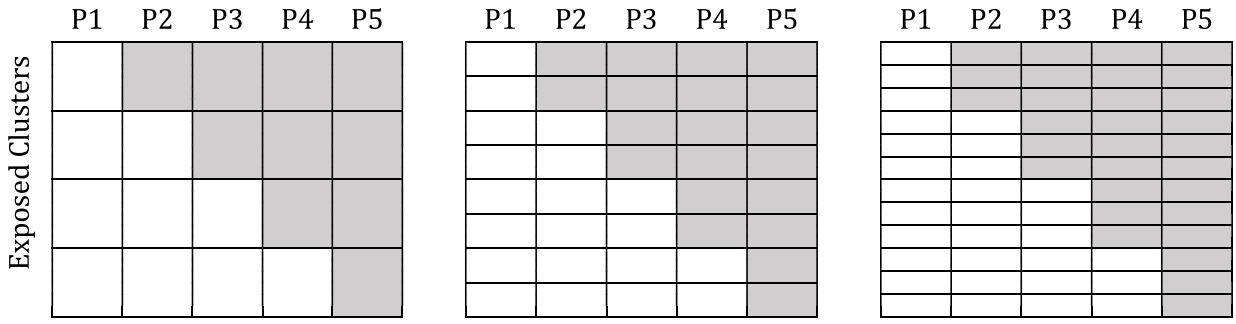

Figure 3. Simulation of three different designs with 4, 8, or 12 exposed clusters. There are 5 periods (4 steps) in each design. Cluster-period cross-sections that are receiving the intervention are shaded in gray.

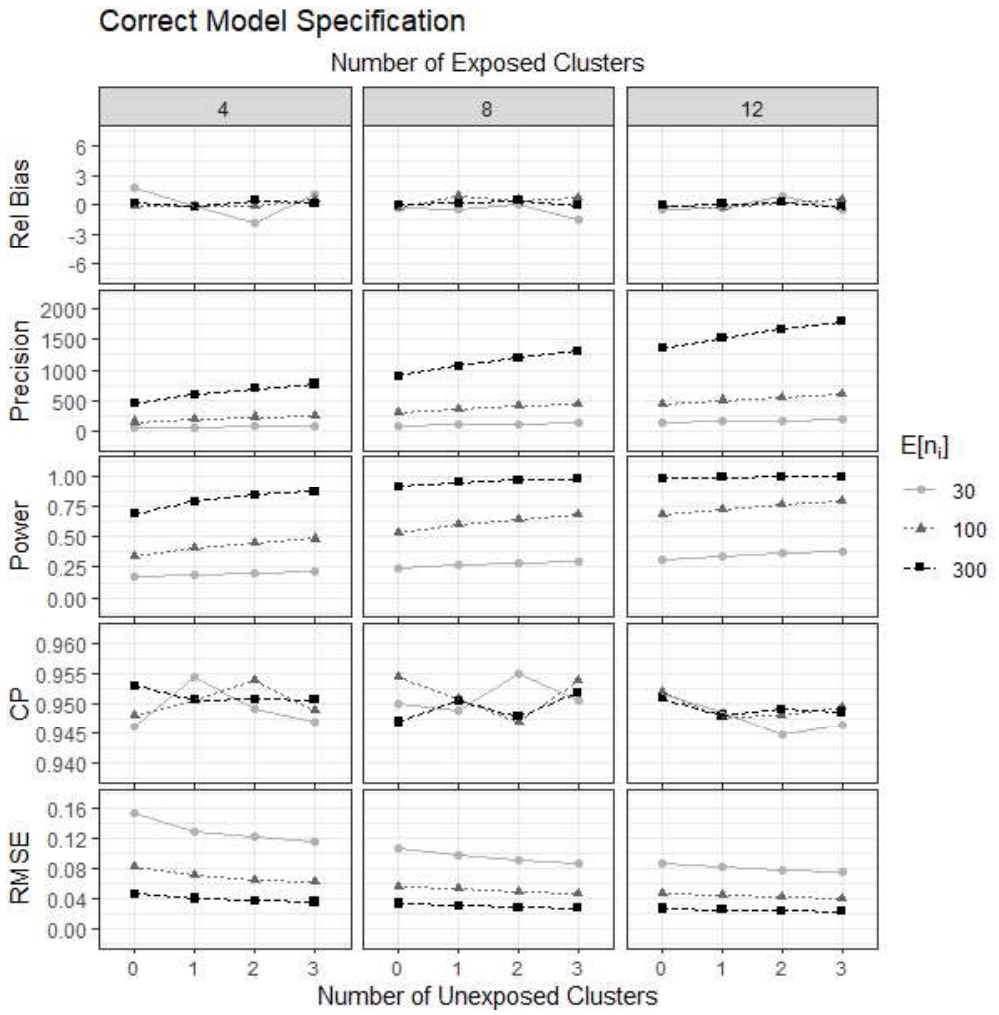

Figure 4. The impact of including unexposed clusters on properties of the fixed effects LSDV intervention effect estimator, presented across number of exposed clusters, unexposed clusters, and average cluster size $E[n_i]$, with fixed true intervention effect $\delta = -0.1$ and between-cluster variability $\tau_\alpha^2 = 1/19$.

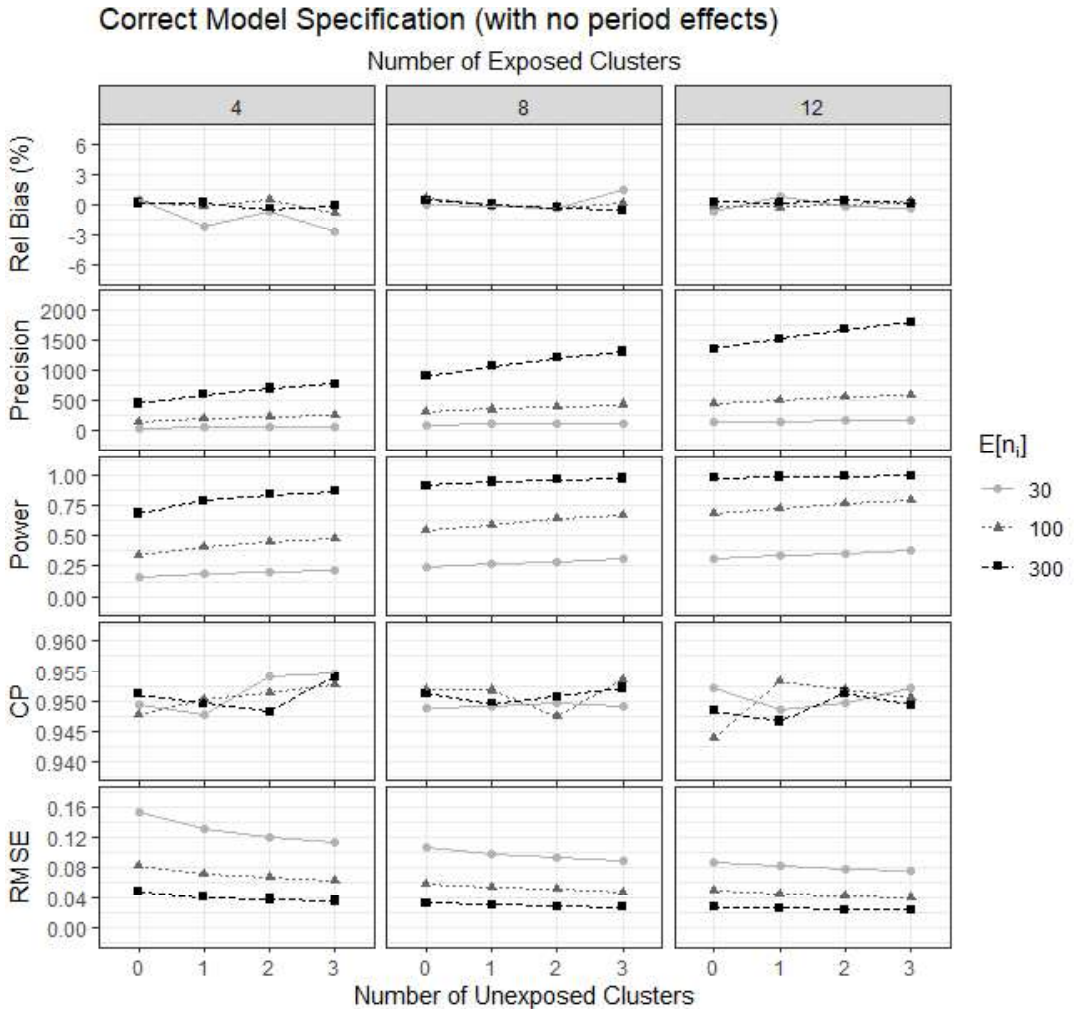

Figure 5. The impact of including unexposed clusters on properties of the fixed effects LSDV intervention effect estimator when there are no period effects ($\phi_j = 0$) in the data generating process, presented across number of exposed clusters, unexposed clusters, and average cluster size $\mathrm{E}[n_i]$, with fixed true intervention effect $\delta = -0.1$ and between-cluster variance $\tau_\alpha^2 = 1/19$.

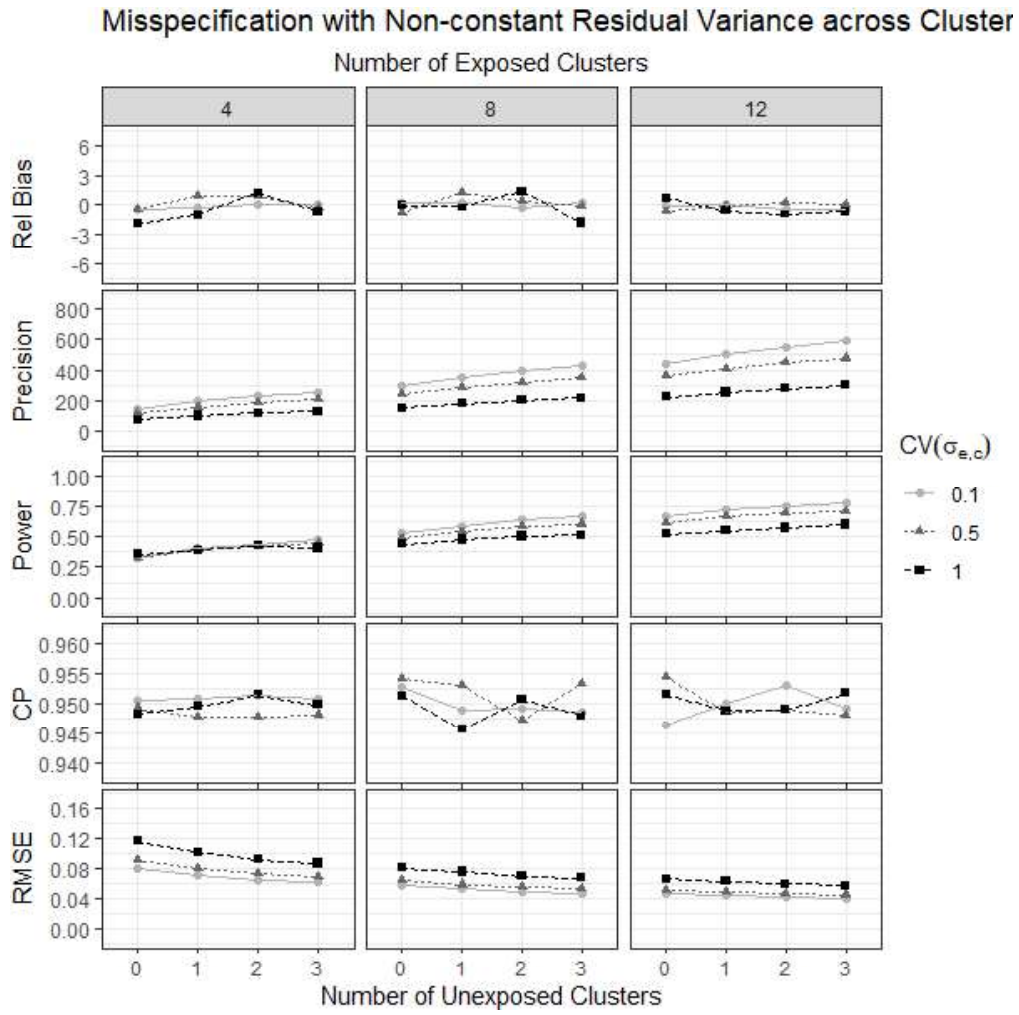

Figure 6. The impact of including unexposed clusters on properties of the fixed effects LSDV intervention effect estimator when clusters have unique within-cluster variability $\sigma_{e,c}$ in the residual errors. Properties of the intervention effect estimator are presented across number of exposed clusters, unexposed clusters, and coefficient of variation of the within-cluster variability CV($\sigma_{e,c}$), with fixed true treatment effect $\delta = -0.1$, between-cluster variability $\tau_\alpha^2 = 1/19$, and average cluster size $E[n_i] = 100$.

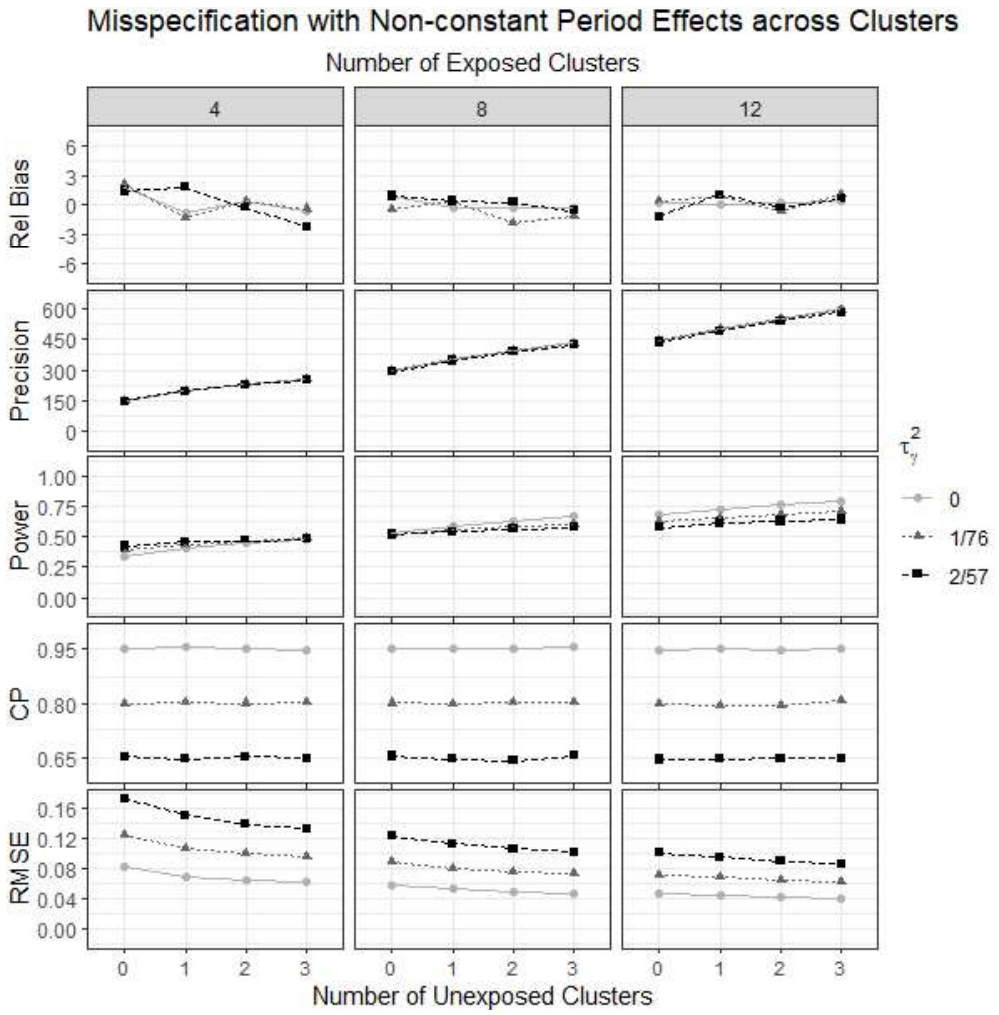

Figure 7. The impact of including unexposed clusters on properties of the fixed effects LSDV intervention effect estimator when period effect varies between different clusters. Properties of the intervention effect estimator are presented across number of exposed clusters, unexposed clusters, and between-period variability τ_γ^2, with fixed true treatment effect $\delta = -0.1$, between-cluster variability $\tau_\alpha^2 = 1/19$, and average cluster size $E[n_i] = 100$.