POWER ANALYSIS FOR STEPPED WEDGE TRIALS WITH TWO TREATMENTS

A PREPRINT

Phillip T Sundin*
Department of Biostatistics
University of California, Los Angeles
Los Angeles, CA 90049
phillip1492@ucla.edu

Catherine M Crespi
Department of Biostatistics
University of California, Los Angeles
Los Angeles, CA 90049
ccrespi@ucla.edu

October 5, 2020

ABSTRACT

Stepped wedge designs (SWDs) are designs for cluster randomized trials that feature staggered, unidirectional cross-over, typically from a control to a treatment condition. Existing literature on statistical power for SWDs primarily focuses on designs with a single treatment. However, SWDs with multiple treatments are being proposed and conducted. We present a linear mixed model for a SWD with two treatments, with and without an interaction between them. We derive closed form solutions for the standard errors of the treatment effect coefficients for such models along with power calculation methods. We consider repeated cross-sectional designs as well as open and closed cohort designs and different random effect structures. Design features are examined to determine their impact on power for main treatment and interaction effects.

Keywords stepped wedge design · factorial design · power calculation · clinical trial · design · cluster randomized trial

1 Introduction

Cluster randomized trials (CRTs) are trials in which groups of individuals, called clusters, are randomized to treatment arms. CRTs can use parallel designs, in which clusters receive only one treatment, or crossover designs, in which clusters are assigned to a sequence of treatments. A variation on the CRT crossover design is the stepped wedge design (SWD) [1]. The most common SWDs involve unidirectional crossover in which clusters transition from a control condition to a treatment condition at different pre-specified times. Clusters are randomized to the pre-determined sequences. Typically, all clusters are in the treatment condition in the last period [2]. An example of such a SWD is shown in Figure 1.

SWDs have several advantages over parallel and crossover CRT designs. SWDs allow comparisons both within cluster and across cluster, which can yield efficiency gains [3]. It may be less costly and logistically easier to roll out the intervention over time instead of all at once, as would occur in many parallel CRTs [4]. Guaranteeing treatment for all clusters may make clusters more willing to participate or alleviate some ethical concerns [3].

*Corresponding author

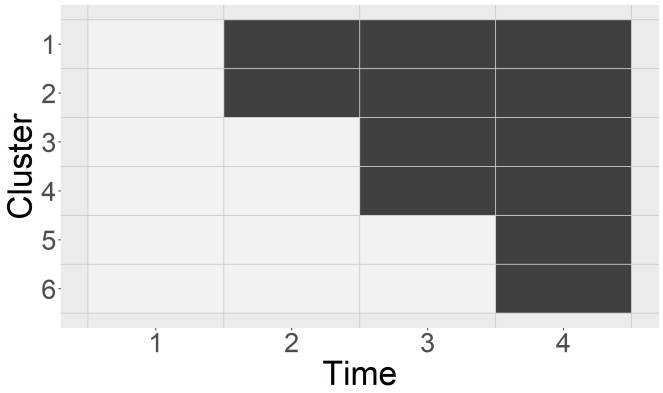


Figure 1: Example of a stepped wedge design with six clusters, four time periods, and one treatment. There are three distinct sequences with two clusters in each sequence. A white cell indicates the control condition and a gray cell represents the treatment condition.

Most research on the design and analysis of stepped wedge trials has focused on the SWD with one intervention contrasted with a control condition. There is a small but growing body of literature on SWDs involving multiple interventions. Grayling et al. [5] focus on studies in which there is a nested natural order of D interventions such that intervention d consists of intervention $d-1$ plus some additional factor. The authors discuss the optimization of treatment sequence allocations and focus on optimal design for such trials. Variances for treatment effect estimates in SWDs with nested interventions have also been studied [6]. The limited existing literature does not include modeling of a potential interaction between treatments [7].

SWDs with multiple interventions are being implemented in practice despite a lack of methodological literature. For example, a recently funded trial of the comparative effectiveness of two interventions to promote human papillomavirus (HPV) vaccination is using a stepped wedge design implemented at seven clinics over a four-year period [8]. One intervention is a parent reminder sent via text message, and the other is a multi-component clinic-based intervention comprised of staff education, workflow modifications, and regular auditing. The study combines the stepped wedge design with a 2×2 factorial design, with cluster-periods in a control condition, single intervention conditions for both treatments, and a combined condition in which clinics receive both interventions. In another example, the STARSHIP study examined two treatments to assess their efficacy in reducing hyperbilirubinaemia in infants [9]. The study employed a SWD with two treatments and a combined condition where a clinic receives both treatments simultaneously. The STARSHIP study assumed additive treatment effects and did not consider an interaction effect. Lyons et al. identified additional studies that employed SWDs with more than one treatment [7]. One such example is the SSTP-ACT study that compared two behavioral health interventions with the goal of reducing behavioral problems in children with cerebral palsy [10]. Similar to the STARSHIP study, the SSTP-ACT study had a combined condition where clusters can receive both treatments simultaneously but did not include an interaction effect.

There are also examples in the literature of studies that conduct two one-treatment SWD trials simultaneously. The FallDem study examined two interventions for improving the lives of dementia patients [7, 11]. Durovni et al. conducted two SWD trials for tuberculosis screening [12, 13]. It is unclear whether the investigators of these studies considered

combining the two separate SWDs into one larger SWD with two treatments. This might have allowed for more precise estimates of treatment effects, direct comparisons of the treatments, and a possible reduction in the number of clusters required to achieve the desired level of power.

In this paper, we consider SWDs with two treatments, with and without a combined condition, and with and without an interaction effect. We call trials that include a combined condition stepped wedge factorial designs (SWFDs). As in other 2×2 factorial designs, these trials involve two factors, each with two levels, with four treatment combinations in total. Factorial designs have several well-known advantages, such as potentially increased efficiency and the ability to study interaction effects [14]. We derive analytical results for conducting power analysis for such trials and examine factors that influence power.

The paper is organized as follows. Section 2 introduces the models for the SWD with two treatments including the SWFD model and develops power analysis methods. Section 3 uses illustrative examples to examine the influence of different design features on power. Section 4 discusses our findings, possible extensions, limitations and future work.

2 Methods

We first present models for SWDs with two treatments, with and without an interaction term, assuming a repeated cross-sectional study design in which outcomes are assessed on different individuals within each cluster in each time period. We then discuss cohort designs, in which outcomes are measured on some or all of the same individuals in a cluster across time periods, and the nested exchangeable model, which relaxes some assumptions. We standardize the models to allow power calculations to be conducted by specifying intracluster correlation (ICC) and autocorrelation values instead of variance components, which can be difficult to specify. Power for both main effects and an interaction effect are considered. The section concludes with an overview of the derivation of closed form solutions for the standard errors of the estimated treatment effect coefficients, with the full derivation found in Appendix A.

2.1 Cross-Sectional Design

We begin with a model for a single binary treatment factor and a repeated cross-sectional design, as originally proposed by Hussey and Hughes [1]. For a design with I clusters observed at T times, and N different individuals per time per cluster, let Y_{ijk} be a continuous outcome for for individual k in cluster i at time j. The model for Y_{ijk} is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta_1 + e_{ijk} \tag{1}$$

where μ is an intercept, $\alpha_i \sim N(0, \sigma_\alpha^2)$ is a random intercept for cluster i, β_j is a fixed effect for time j, X_{ij} is a {0,1} indicator for whether cluster i at time j receives treatment, θ_1 is the treatment effect, and $e_{ijk} \sim N(0, \sigma_e^2)$. The total variance of an individual level outcome is $\sigma_y^2 = \sigma_\alpha^2 + \sigma_e^2$.

This model can easily be expanded to accommodate multiple binary treatment factors. Assuming additive treatment effects, the model with two treatment conditions is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta_1 + W_{ij}\theta_2 + e_{ijk}, \tag{2}$$

where W_{ij} is a {0,1} indicator of whether cluster i at time j receives treatment 2 and θ_2 is the treatment effect for treatment 2. Adding an interaction effect to accommodate trials in which a cluster can receive both treatments simultaneously, the model becomes

$$Y_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta_1 + W_{ij}\theta_2 + X_{ij}W_{ij}\theta_3 + e_{ijk}, \tag{3}$$

where θ_3 is the interaction effect.

In these models, the correlation between outcomes of two different individuals k and k' in the same cluster i at the same time j, Corr($y_{ijk}, y_{ijk'}$), is commonly termed the intracluster correlation (ICC). We denote this ICC as ρ_w. It can be shown that $\rho_w = \sigma^2_\alpha/\sigma^2_Y$. These models also imply that the correlation between two observations on different individuals in the same cluster and same time period is equal to the correlation between two observations in the same cluster but different time periods, i.e., Corr($y_{ijk}, y_{ijk'}$) = Corr($y_{ijk}, y_{ij'k'}$). We denote the latter correlation as ρ_a and note that the repeated cross-sectional model assumes $\rho_w = \rho_a$.

To standardize one of these models, one can divide it through by σ_y. The cluster random intercept α_i now has standardized variance ρ_w and the error term has standardized variance $1 - \rho_w$. Standardization is convenient for power calculations because only ρ_w needs to be specified rather than both σ^2_e and σ^2_α.

To derive standard errors, it is convenient to work with cluster-level outcomes. Let $\overline{Y}_{ij\cdot} = \frac{1}{N}\sum_{k=1}^{N} Y_{ijk}$ be the mean outcome of cluster i at time j across N individuals. The model for cluster-period means is

$$\overline{Y}_{ij\cdot} = \mu + \alpha_i + \beta_j + X_{ij}\theta_1 + W_{ij}\theta_2 + X_{ij}W_{ij}\theta_3 + e_{ij\cdot}, \tag{4}$$

where $e_{ij\cdot} = \frac{1}{N}\sum_{k=1}^{N} e_{ijk} \sim N(0, \sigma^2_c)$ and $\sigma^2_c = \frac{\sigma^2_e}{N}$. Under this model, the variance of a cluster-period mean is Var($\overline{Y}_{ij\cdot}$) $= \sigma^2_c + \sigma^2_\alpha$. Define the outcome vector $\boldsymbol{Y} = (\overline{Y}_{11\cdot}, \ldots, \overline{Y}_{iT\cdot}, \ldots, \overline{Y}_{I1\cdot}, \ldots, \overline{Y}_{IT\cdot})'$. Assuming clusters are independent, the variance-covariance matrix of $\boldsymbol{Y}$ is a $IT \times IT$ matrix of the form

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V_1} & 0 & 0 & 0 \\ 0 & \boldsymbol{V_2} & \ldots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \ldots & \boldsymbol{V_I} \end{bmatrix},$$

with each $T \times T$ matrix $\boldsymbol{V_i}$ having structure

$$\boldsymbol{V_i} = \begin{bmatrix} \sigma_\alpha^2 + \sigma_c^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_c^2 & \dots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 + \sigma_c^2 \end{bmatrix}.$$

In the standardized model, $\boldsymbol{V_i}$ will have diagonal elements $\rho_w + (1 - \rho_w)/N$ and off-diagonal elements ρ_w. Now we turn to the design matrix of the fixed effects. Setting $\beta_T = 0$ for identifiability [1], the $(T + 3) \times 1$ regression coefficient vector for the fixed effects is

$$\boldsymbol{\eta} = \begin{bmatrix} \mu & \beta_1 & \dots & \beta_{T-1} & \theta_1 & \theta_2 & \theta_3 \end{bmatrix}'.$$

Then, the full $IT \times (T + 3)$ design matrix $\boldsymbol{Z}$ becomes

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z_1} & \boldsymbol{Z_2} & \dots & \boldsymbol{Z_I} \end{bmatrix}'.$$

Each matrix $\boldsymbol{Z_i}$ has dimension T $\times$ (T+3) and takes the form

$$\boldsymbol{Z_i} = \begin{bmatrix} \boldsymbol{1_T} & \begin{matrix} \boldsymbol{I_{T-1}} \\ \boldsymbol{0'_{T-1}} \end{matrix} & \boldsymbol{X_i} & \boldsymbol{W_i} & \boldsymbol{(XW)_i} \end{bmatrix}.$$

The elements of the vector $\boldsymbol{X_i} = (X_{i1}, X_{i2}, \dots, X_{iT})'$ are indicators of whether cluster i at time j receives treatment 1, the elements of $\boldsymbol{W_i} = (W_{i1}, W_{i2}, \dots, W_{iT})'$ are indicators of receipt of treatment 2, and $\boldsymbol{(XW)_i}$ is the Hadamard product of $\boldsymbol{X_i}$ and $\boldsymbol{W_i}$, with a value of 1 if cluster i receives both treatments at time j and 0 otherwise. The matrix $\boldsymbol{I_{T-1}}$ contains indicators for each time j from $1, \dots, (T - 1)$. The vector $\boldsymbol{0'_{T-1}}$ corresponds to time T.

2.2 Cohort Design

When an individual can be observed at more than one time period, we must account for the correlation between observations from the same individual. This can be accomplished by adding a random intercept $\psi_{ik} \sim N(0, \sigma_\psi^2)$ for individual k in cluster i to the cross-sectional model (3), yielding

$$Y_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta_1 + W_{ij}\theta_2 + X_{ij}W_{ij}\theta_3 + \psi_{ik} + e_{ijk}. \tag{5}$$

The total variance of an individual level outcome in this model is $\sigma_y^2 = \sigma_\alpha^2 + \sigma_\psi^2 + \sigma_e^2$. The correlation between two observations in the same cluster and same time period, Corr($y_{ijk}, y_{ijk'}$), $k \neq k'$, is $\rho_w = \sigma_\alpha^2/\sigma_Y^2$, and the correlation

between two observations on the same individual k at times j and j', Corr($y_{ij'k}, y_{ijk}$), $j \neq j'$, is $\rho_a = (\sigma_\alpha^2 + \sigma_\psi^2)/\sigma_Y^2$. Note that $\rho_a \geq \rho_w$.

Individual auto-correlation (IAC) is defined as the proportion of the individual-level variance (which in this model is $\sigma_\psi^2 + \sigma_e^2$) that is time-invariant. In this model, the IAC is $\pi = \sigma_\psi^2/(\sigma_\psi^2 + \sigma_e^2)$. Setting $\pi = 0$ yields a repeated cross-sectional design. By allowing π to vary between 0 and a specified maximum value for individuals followed throughout the entire study, this model can handle studies with a mixture of cross-sectional and cohort observations for power calculations [15, 16], including open cohort designs in which individuals can enter and/or depart during the study period [17].

Dividing the model by σ_Y, we obtain the standardized variance of α_i as $\rho_w = \sigma_\alpha^2/\sigma_Y^2$. Using the definition of the IAC and the fact that the standardized variances sum to 1, we obtain the standardized variance of ψ_{ik} as $\pi(1-\rho_w)$ and the standardized variance of the error term as $1 - \rho_w - \pi(1-\rho_w)$. Thus we can specify all standardized variances in terms of ρ_w and π, which is convenient for power calculations.

In this model, the cluster-period means are equal to

$$\overline{Y}_{ij\cdot} = \mu + \alpha_i + \beta_j + X_{ij}\theta_1 + W_{ij}\theta_2 + X_{ij}W_{ij}\theta_3 + \psi_i + e_{ij}, \tag{6}$$

where $\psi_i = \frac{1}{N}\sum_{k=1}^{N}\psi_{ik} \sim N(0, \frac{\sigma_\psi^2}{N})$. The variance-covariance matrix of the cluster-period means for cluster i is

$$\boldsymbol{V_i} = \begin{bmatrix} \sigma_\alpha^2 + \sigma_c^2 + \frac{\sigma_\psi^2}{N} & \sigma_\alpha^2 + \frac{\sigma_\psi^2}{N} & \dots & \sigma_\alpha^2 + \frac{\sigma_\psi^2}{N} \\ \sigma_\alpha^2 + \frac{\sigma_\psi^2}{N} & \sigma_\alpha^2 + \sigma_c^2 + \frac{\sigma_\psi^2}{N} & \dots & \sigma_\alpha^2 + \frac{\sigma_\psi^2}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 + \frac{\sigma_\psi^2}{N} & \sigma_\alpha^2 + \frac{\sigma_\psi^2}{N} & \dots & \sigma_\alpha^2 + \sigma_c^2 + \frac{\sigma_\psi^2}{N} \end{bmatrix}.$$

In the standardized model, the matrix $\boldsymbol{V_i}$ has diagonal elements $\rho_w + (1-\rho_w)/N$ and off-diagonal elements $\rho_w + \pi(1-\rho_w)/N$.

2.3 Nested Exchangeable Correlation Model

The repeated cross-sectional model (3) assumes the correlation between two observations in the same cluster and same time period is equal to the correlation between two observations in the same cluster but different time periods, i.e., $\rho_w = \rho_a$. Using the approach in Hemming et al. [2], one can allow these value to differ by adding a random effect ν_{ij} for cluster i in time j, yielding the model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta_1 + W_{ij}\theta_2 + X_{ij}W_{ij}\theta_3 + \nu_{ij} + e_{ijk}, \tag{7}$$

in which $\nu_{ij} \sim N(0, \sigma_\nu^2)$ is assumed independent of e_{ijk} and α_i. In this model, $\sigma_y^2 = \sigma_\nu^2 + \sigma_\alpha^2 + \sigma_e^2$. This approach divides the cluster-level variance into time-varying and time invariant components, namely, σ_ν^2 and σ_α^2. The within-period ICC is now $\rho_w = (\sigma_\nu^2 + \sigma_\alpha^2)/\sigma_y^2$ and the across-period ICC is $\rho_a = \sigma_\alpha^2/\sigma_y^2$. Let the cluster auto-correlation (CAC) be the proportion of cluster level variance that is time-invariant [15, 16]. Then we have CAC $= \sigma_\alpha^2/(\sigma_\nu^2 + \sigma_\alpha^2) = \rho_a/\rho_w$.

The standardized variances of α_i, ν_{ij} and e_{ijk} are ρ_a, $\rho_w - \rho_a$ and $1 - \rho_w$, respectively. Thus all variance components can be expressed in terms of ρ_w and ρ_a. Note that we must have $\rho_w \geq \rho_a$.

Under this model, cluster-period means take the form

$$\overline{Y}_{ij\cdot} = \mu + \alpha_i + \beta_j + X_{ij}\theta_1 + W_{ij}\theta_2 + X_{ij}W_{ij}\theta_3 + \nu_{ij} + e_{ij\cdot} \tag{8}$$

and the variance-covariance matrix for cluster i is

$$\boldsymbol{V_i} = \begin{bmatrix} \sigma_\alpha^2 + \sigma_c^2 + \sigma_\nu^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_c^2 + \sigma_\nu^2 & \dots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 + \sigma_c^2 + \sigma_\nu^2 \end{bmatrix}.$$

In the standardized model, $\boldsymbol{V_i}$ has diagonal elements $\rho_w + (1 - \rho_w)/N$ and off-diagonal elements ρ_a.

2.4 Power Analysis

Inference for fixed effects in linear mixed models can be conducted using Wald tests or likelihood ratio tests. We focus on Wald tests. For hypotheses of the form $H_0 : \eta = 0$, where η is a fixed effects coefficient, the Wald test statistic takes the form $\hat{\eta}/\sqrt{\text{Var}(\hat{\eta})}$, where $\hat{\eta}$ is the estimated coefficient, and has an approximate standard normal distribution when the null is true [18]. The power to reject H_0 for a specific true value of η, denoted η_a, with type I error rate α is approximately

$$P\left(\left| \frac{\eta_a}{\sqrt{\text{Var}(\hat{\eta})}} \right| \geq \; z_{1-\frac{\alpha}{2}} \;\middle|\; \eta = \eta_a \right)$$

where $z_{1-\frac{\alpha}{2}}$ is the corresponding $(1 - \frac{\alpha}{2})^{th}$ percentile of the standard normal distribution.

To calculate power, we need an expression for Var$(\hat{\eta})$. We derive expressions for Var$(\hat{\eta})$ for each of the three models using the cluster-period mean models in (4), (6) and (8). Given that these are linear mixed models, the variance-covariance matrix of the estimated fixed effect coefficients takes the form $\boldsymbol{C} = (\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}$, where $\boldsymbol{Z}$ is the fixed effects design matrix and $\boldsymbol{V}$ is the variance-covariance matrix of the outcome vector. We are most interested in finding expressions for the variances and covariances of treatment effect coefficient estimates, $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_3$. We find an expression for $\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}$ then invert it to get the elements of $(\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}$ corresponding to these variances and

covariances. We focus on the repeated cross-sectional model here; the derivation generalizes to the cohort and nested exchangeable models by using the appropriate structure of the matrix $\boldsymbol{V}$.

Let $\boldsymbol{Z}$ be the $IT \times (T+3)$ design matrix and $\boldsymbol{V}$ be the $IT \times IT$ variance-covariance matrix of the cluster-level outcomes. Assuming clusters are independent, the matrix $\boldsymbol{V}$ has block diagonal structure with elements $\boldsymbol{V_i} = \sigma_c^2 \boldsymbol{I_T} + \sigma_\alpha^2 \boldsymbol{1_T} \boldsymbol{1'_T}$, where $\boldsymbol{I_T}$ is a $T \times T$ identity matrix and $\boldsymbol{1_T}$ is a $T \times 1$ vector of 1's. Using the Sherman-Morrison formula [19, 20], we can obtain its inverse as

$$\boldsymbol{V_i}^{-1} = \frac{1}{\sigma_c^2(\sigma_c^2 + T\sigma_\alpha^2)} \left[(\sigma_c^2 + T\sigma_\alpha^2)\boldsymbol{I_T} - \sigma_\alpha^2 \boldsymbol{1_T} \boldsymbol{1_T}' \right].$$

This matrix has off-diagonal elements $\frac{-\sigma_\alpha^2}{\sigma_c^2(T\sigma_\alpha^2+\sigma_c^2)}$ and diagonal elements $\frac{(T-1)\sigma_\alpha^2+\sigma_c^2}{\sigma_c^2(T\sigma_\alpha^2+\sigma_c^2)}$. Due to the block diagonal structure of $\boldsymbol{V}$, we have

$$\boldsymbol{Z'V}^{-1}\boldsymbol{Z} = \sum_{i=1}^{I} \boldsymbol{Z_i}'\boldsymbol{V_i}^{-1}\boldsymbol{Z_i},$$

where $\boldsymbol{Z_i}$ is the $T \times (T+3)$ part of the design matrix corresponding to cluster i. We can then rewrite

$$\boldsymbol{Z_i}'\boldsymbol{V_i}^{-1}\boldsymbol{Z_i} = \frac{1}{\sigma_c^2(\sigma_c^2 + T\sigma_\alpha^2)} \left[(\sigma_c^2 + T\sigma_\alpha^2)\boldsymbol{Z_i}'\boldsymbol{Z_i} - \sigma_\alpha^2 \boldsymbol{Z_i}'\boldsymbol{1_T}\boldsymbol{1_T}'\boldsymbol{Z_i} \right]. \tag{9}$$

We then use block matrix inversion techniques to solve for the submatrix corresponding to the coefficients corresponding to the treatment and interaction effects. A full derivation is provided in Appendix A.

3 Illustrative Examples

Although we derived closed form solutions for the variance and covariances of the treatment and interaction effect estimates, there is no simple, intuitive expression for the standard errors nor for power. We therefore use numerical examples to illustrate how power for detecting treatment and interaction effects is affected by design features for SWDs with two treatments. All calculations were performed in R version 3.6.1 [21] with code available on Github at https://github.com/phillipsundin/SWFD.

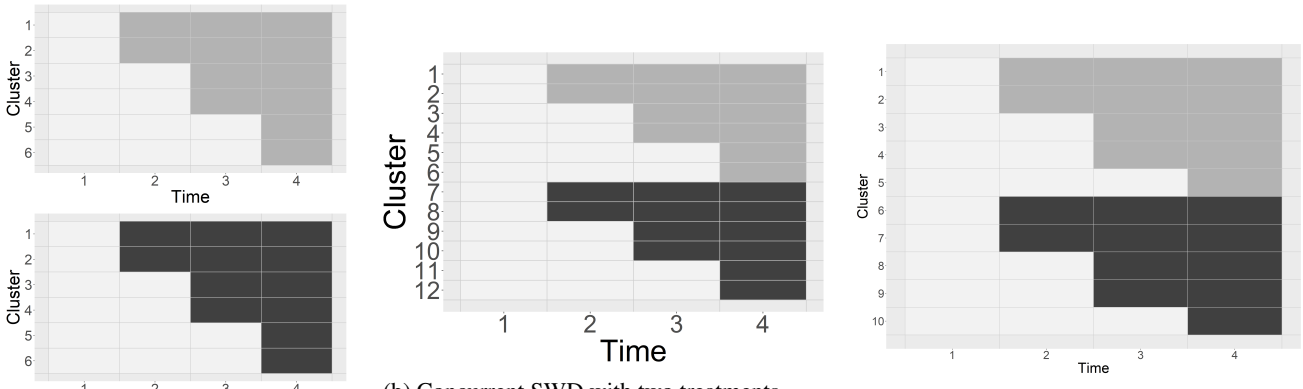
Our main concerns are how power is impacted by the assignment of cluster-periods to conditions and by the factorial aspect of designs. We focus attention on cluster-period assignments that are likely to be used in practice. For example, we only consider designs in which a combined condition follows a control condition or a single treatment condition. We do not consider designs in which a cluster can transition from a combined to a single treatment condition, or from one single treatment to the other single treatment, because such designs are vulnerable to contamination and therefore would not be commonly used. We also focus on designs with relatively few clusters, which is common in practice [22].

For simplicity, we use the repeated cross-sectional design for our examples. Power for repeated cross-sectional versus cohort designs is not a focus of our paper and has been addressed by others [16]. In general, cohort designs have higher power than cross-sectional designs [23, 24]. Often, ICCs are reported as the within-time, within-cluster correlation, ρ_w; the across-time, within-cluster correlation ρ_a and individual auto-correlation π are not always reported. Given this lack of information, it may be sensible to make the simplifying assumption that $\rho_w = \rho_a$, which corresponds to the repeated cross-sectional design.

3.1 Two-Treatment Concurrent SWD vs Two Separate Single Treatment SWDs

Several studies have conducted two separate one-treatment SWD trials [11, 12]. We explore how power might be increased or the required number of clusters reduced by combining two one-treatment trials into one trial with two treatments.

Consider the two one-treatment SWD trials, each with six clusters and four time periods, illustrated in Figure 2a. The design in Figure 2b combines these into a single 12-cluster trial; such a design has been called a concurrent design [7]. Figure 2c shows a similar concurrent design with only 10 clusters. Let δ_1 and δ_2 denote the standardized effect sizes for treatments 1 and 2, respectively. We set $\delta_1 = \delta_2 = 0.4$, representing medium effect sizes [25]. Note that within each design, the power to detect treatment effects 1 and 2 is the same due to symmetry. We specify $N = 15$ individuals per cluster-period in a repeated cross-sectional design. Power for detecting a treatment effect in one of the one-treatment SWDs is calculated using model (1); for the concurrent SWDs with two treatments, power is calculated using model (2).



(a) Two separate single treatment SWDs

(b) Concurrent SWD with two treatments, 12 clusters

(c) Concurrent SWD with two treatments, 10 clusters

Figure 2: Examples of two one-treatment SWDs versus concurrent SWDs with two treatments. White cells indicate cluster-periods in the control condition. Light and dark gray cells indicate treatment conditions for treatments 1 and 2, respectively.

Figure 3 displays power for either treatment effect for the three designs as a function of ρ_w. Combining the two one-treatment SWDs into a single concurrent design while maintaining the same number of clusters yields a large gain in power, ranging from 0.14 to 0.20 for the values of ρ_w considered. The concurrent design with only ten clusters reduces the sample size by about 17% and has moderately higher power, ranging from 0.08 to 0.11. Further reducing

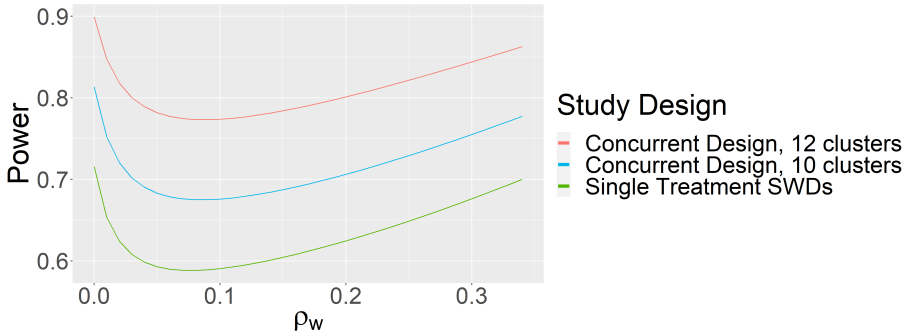


Figure 3: Comparing power for either treatment effect for two simultaneous one-treatment SWDs, a concurrent SWD with two treatments with 12 clusters, and a concurrent SWD with two treatments with 10 clusters.

the number of clusters while maintaining symmetry would yield lower power than the single treatment SWDs in this example.

In addition to power gains and reduced sample size requirements, another advantage of including two treatments in one SWD is the capability to directly compare the two interventions. This can be accomplished using tests of linear contrasts of treatment effects. A closed form solution for the standard error of such contrasts, which includes a covariance term, can be computed using our methods. Suppose we wish to detect a standardized difference of 0.4 units between the two treatments. Figure 4 displays power for the linear contrast as a function ρ_w. The relationship between ρ_w and power for the linear contrast is similar in shape to the relationship between ρ_w and power for detecting main effects. However, the nadir of the power curve is at a higher value of ρ_w (about 0.12) for the linear contrast in this example.

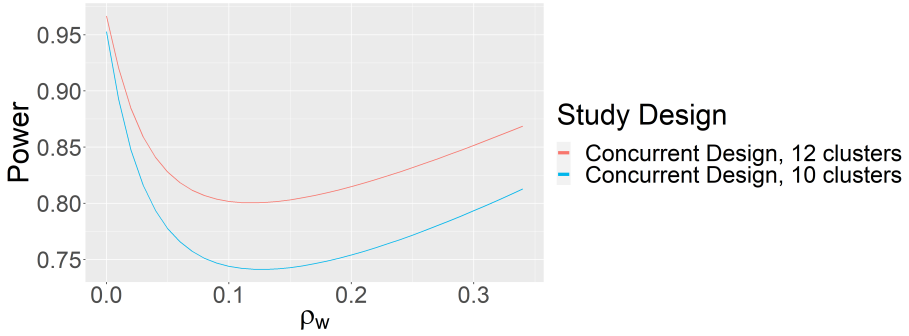
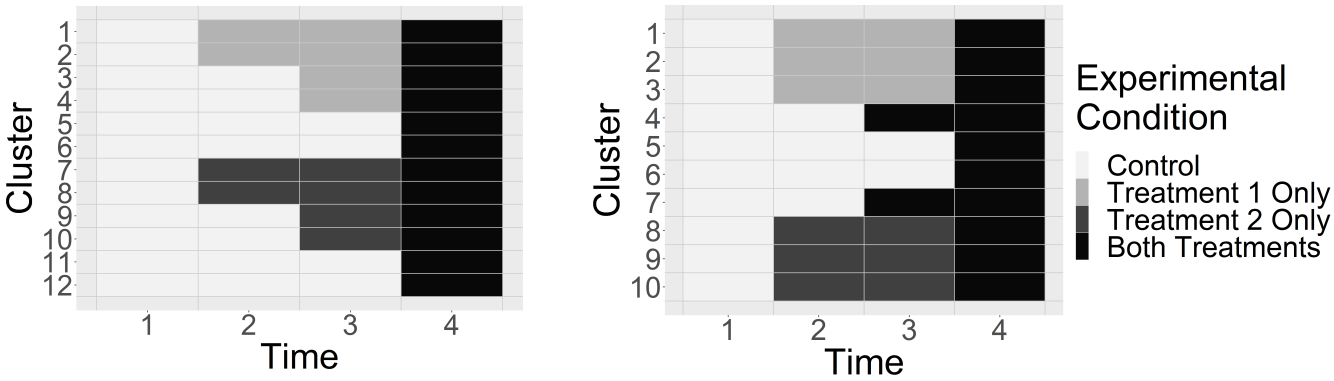


Figure 4: Comparing power for detecting a difference between two treatments for a 12-cluster concurrent SWD and a 10-cluster concurrent SWD.

3.2 Stepped Wedge Factorial Designs with Additive Treatment Effects

We now consider using designs with a combined condition to estimate additive treatment effects. We contrast the 12-cluster concurrent design in Figure 2b with SWFDs that feature some cluster-periods receiving both treatments simultaneously. While many designs are possible, for illustrative purposes we consider the two designs in Figure 5. Figure 5a shows a 12-cluster “late" factorial design in which all clusters transition to the combined condition for the last period. Figure 5b shows an “earlier" factorial design with only ten clusters that introduces the combined condition

earlier. Both designs feature six cluster-periods in each single treatment condition and twelve cluster-periods in the combined condition, representing equal intervention resources. Treatment effects are assumed to be additive. Other design aspects, including repeated cross-sectional observations, moderate effect sizes ($\delta_1 = \delta_2 = 0.4$) and $N = 15$ for each cluster-period, are assumed to be the same.



(a) "Late" factorial design with 12 clusters in which every cluster receives the combined condition in the last (fourth) period

(b) "Earlier" factorial design with 10 clusters, in which two clusters begin receiving the combined condition in the third period

Figure 5: Stepped Wedge Factorial Design Examples.

Figure 6 compares power for the three designs (Figures 2b, 5a and 5b) as a function of ρ_w. For $\rho_w < 0.02$, the 12-cluster concurrent design and the 10-cluster “earlier" factorial design have similar power while the 12-cluster “late" factorial design has the lowest power. For $\rho_w > 0.02$, the power of the 10-cluster “earlier" design is highest; the 12-cluster “late" design continues to have the lowest power.

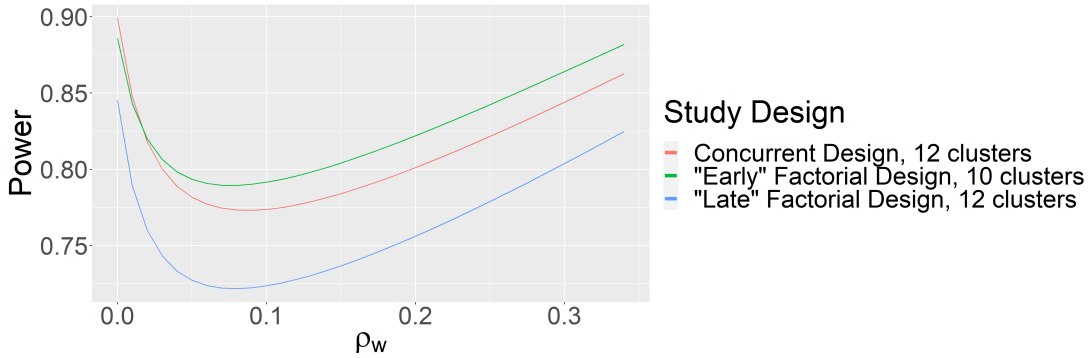


Figure 6: Comparing power for a 12-cluster concurrent design, a factorial design with 10 clusters, and a factorial design with 12 clusters.

This example illustrates several points. First, when the treatment effects are additive, including a combined condition has the potential to increase efficiency and reduce the number of clusters required to achieve a desired level of power. However, the timing of transitions to the combined condition matters. Simply assigning all clusters to the combined condition for the last period reduced power. Rather, the combined condition needed to be introduced in an earlier period to realize efficiency gains. Furthermore, the value of the ICC is also important. Efficiency gains from using a SWFD

were only realized when the ICC exceeded a certain value. A similar phenomenon has been found for a one-treatment SWD compared to a parallel cluster randomized trial [26].

We can also compare the two treatments directly using the SWFD. Suppose we wish to detect a standardized difference of 0.4 between the two treatment conditions. Figure 7 shows that is this example, the concurrent design has the highest power for detecting this contrast across all values of ρ_w. For low values of ρ_w, the two factorial designs have similar power, but for $\rho_w > 0.10$, the design with 12 clusters has about 7% higher power compared to the 10-cluster design.

This example shows that for hypotheses comparing treatments, introducing a combined condition is likely to reduce power compared to a concurrent design. Introducing a combined condition decreases both the variances and covariance of the two treatment effect estimates; however, the covariance tends to decrease at a much faster rate than the variances as the number of cluster-periods in the combined condition is increased. Because $\text{Var}(\hat{\theta}_1 - \hat{\theta}_2) = \text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2) - 2\text{Cov}(\hat{\theta}_1, \hat{\theta}_2)$, the higher covariance values in the concurrent design will result in a smaller overall variance for the contrast, and thus higher power, compared to most SWFDs. This result may seem counterintuitive, since factorial designs are associated with efficiency gains for testing linear contrasts in the context of independent units of randomization. However, note that we still see efficiency gains for testing main effects with this factorial design.

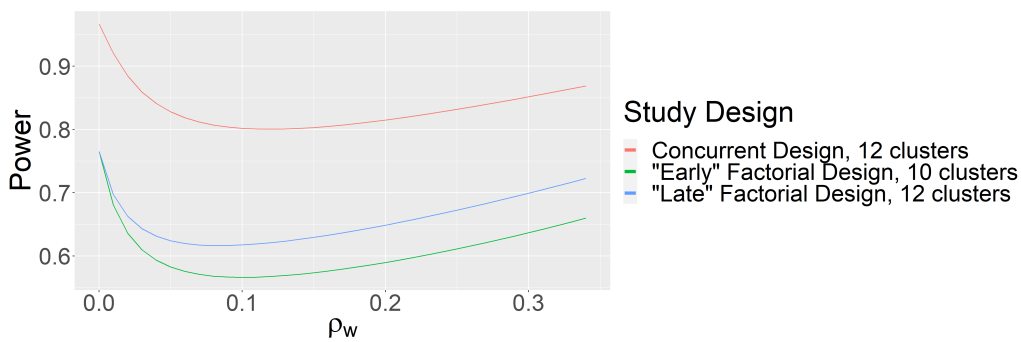


Figure 7: Comparing power for detecting a difference between two treatments for the concurrent and factorial designs.

3.3 Stepped Wedge Factorial Designs with Interaction Effect

We now consider designs in which detecting an interaction is of interest. Consider the four designs in Figure 8. All have eight clusters, as well as seven cluster-periods in treatment 1 only, seven cluster-periods in treatment 2 only, and ten cluster-periods in the combined condition. Design #1 is essentially a concurrent design in which each cluster transitions from control to one condition only during the study, including the combined condition. Design #2 has three clusters transitioning from control to treatment 1 to combined, three clusters transitioning from control to treatment 2 to combined, and two clusters transitioning from control to a single treatment condition relatively late in the study. Design #3 has six of the eight clusters transition to a single treatment in classic stepped wedge style, and all clusters transition to the combined condition. Design #4 is similar but includes two clusters that never transition to the combined condition as well as combined condition clusters that start earlier.

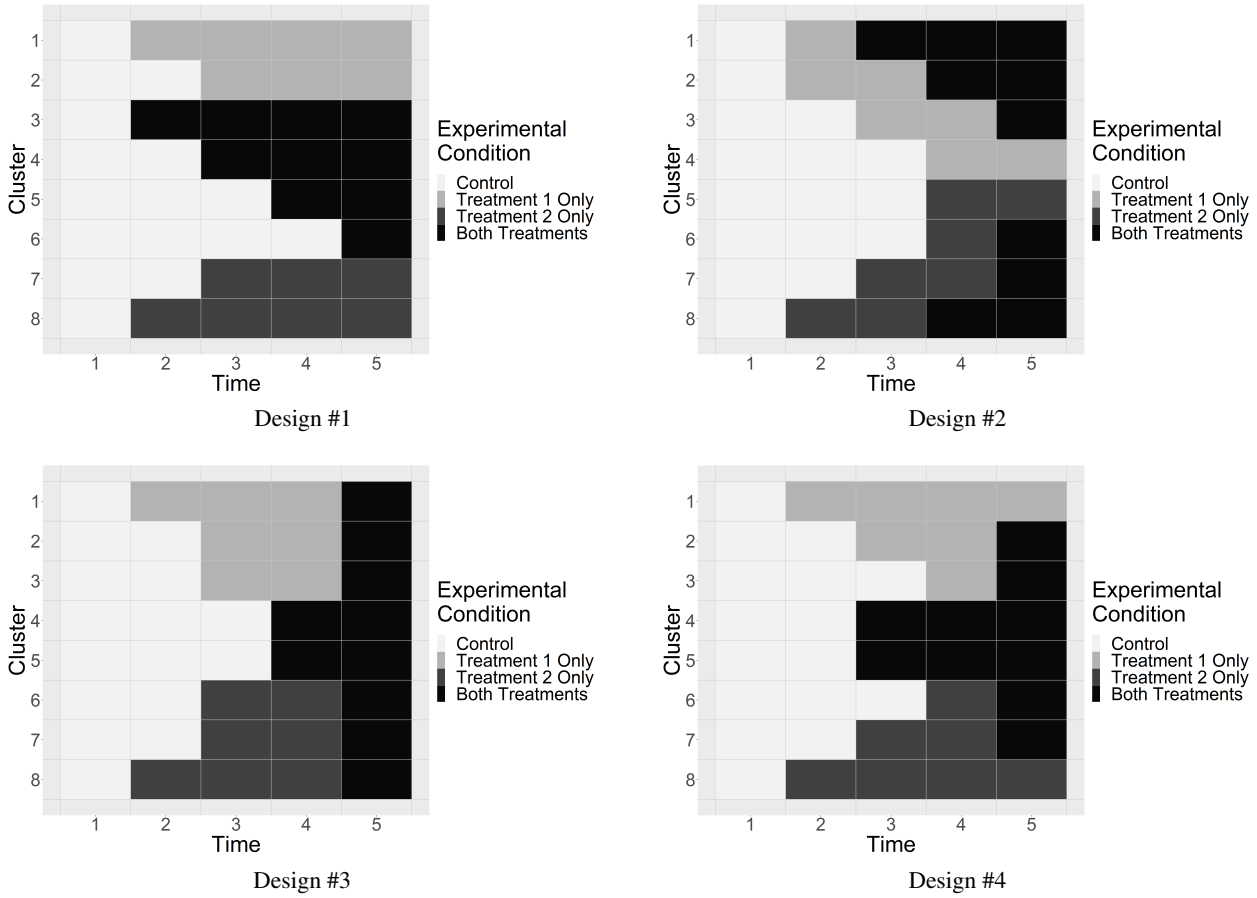


Figure 8: Varying sequences in a stepped wedge factorial design.

We again assume a repeated cross-sectional design with 15 individuals per cluster-period. We calculate power to detect a standardized effect size of 0.60 for each main and interaction effect. Because this represents a relatively large interaction effect, this example illustrates how design choices impact power for an interaction effect.

Figure 9 presents power as a function of ρ_w for the four designs. Note that Designs #1, #3, and #4 are symmetric in treatments 1 and 2 and thus power for these two treatment effects is the same in these designs. Design #2 is close to symmetric, but symmetry may be hard to achieve with a small number of clusters. We focus first on power for main effects, displayed in Figure 9a. For all values of ρ_w, Design #2 has the highest power for detecting main effects. In this design, power for treatment 1 is slightly higher than that for treatment 2. This result is attributable to a more balanced sequencing of treatment 1 over time compared to treatment 2, as the majority of cluster-periods with only treatment 2 are administered later in the study. For values of $\rho_w < 0.06$, Design #4 has the lowest power; for $\rho_w > 0.06$, Design #1 (concurrent) has the lowest power. Focusing on power for the interaction effect, displayed in Figure 9b, Design #2 has by far the highest power, while Design #3 has the lowest power. Design #1 also has relatively low power.

Design #2 is clearly superior for detecting both main and interaction effects. In Design #2, clusters transition between conditions more than any other design. When there are more transitions, within-cluster comparisons are increased, and thus power to detect effects is increased. Design #1 does not have any clusters transition to different treatment

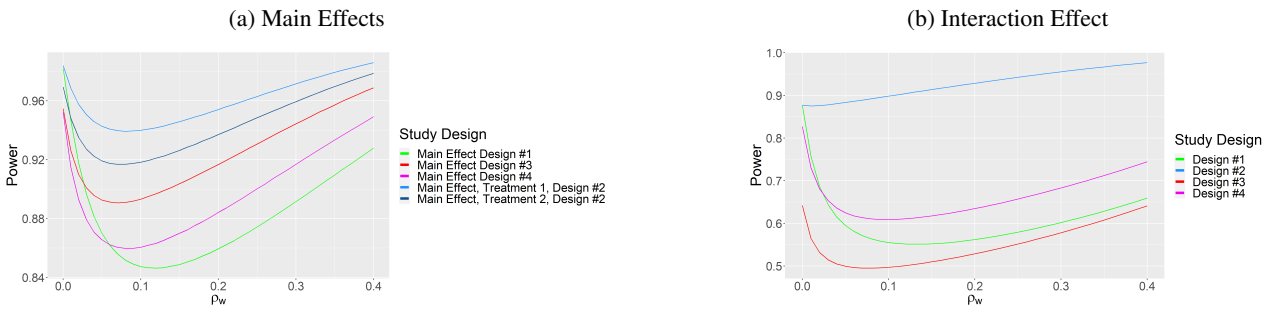


Figure 9: Comparing main and interaction effects for SWFDs with different sequences.

conditions after the first transition from the control condition, and thus suffers from the lowest power for main effects. For the interaction effect, Design #3 suffers from the lowest power. This can be partially attributed to confounding with time, as all of the clusters are in the combined condition during the last time period, and there are only two cluster-periods in the combined condition before the last time period.

This example illustrates that to power on the interaction term, designs should include (1) clusters that transition from control to single treatment to the combined condition and (2) transitions to the combined condition should occur relatively early in the study. Our examples also show that power for both main and interactions effect in SWFDs has a non-monotone relationship with ρ_w, with the lowest power for values of ρ_w between 0 and 0.2, depending on the particular design.

We note that in all of our examples, the power curves have had a non-monotone relationship with ρ_w, such that power is a decreasing function of ρ_w at lower values of ρ_w and an increasing function at higher values. This is consistent with power curves for stepped wedge designs with only one treatment [3, 26, 27].

4 Discussion

In this paper, we present power calculation methods for stepped wedge trials that involve two treatments, with and without a combined condition or interaction. Designs with two treatments are increasingly being used in practice, despite a paucity of literature on their statistical design and analysis.

We focus on studies that include a relatively small number of clusters, which is common for stepped wedge trials [22]. In our examples, it was not possible to explore all possible design options. However, the examples demonstrate several principles. We show that a concurrent design, in which two one-treatment stepped wedge trials are conducted as a single study, is more efficient than two separate one-treatment studies, and we enable power calculations for such studies. In concurrent designs, cluster-periods in the control condition perform “double duty" by serving as controls for both treatment conditions. Such trials are essentially three-arm trials in which two interventions are each compared to a control condition.

When designing trials with two treatment conditions, investigators may wish to consider the need to control the experimentwise type I error rate. We note that a concurrent stepped wedge design is similar to a one-way analysis of variance design in which multiple treatment groups are each compared to a common control group. In this setting, Dunnett's method may be used for experimentwise type I error rate control [28].

Our examples illustrate that factorial stepped wedge designs that include cluster-periods in a combined condition can increase power substantially compared to concurrent designs when treatment effects are additive. However, power may end up being inadequate if an interaction between treatments is manifested in the study but was not taken into account in power calculations. One approach for guarding against this eventuality is to conduct sensitivity analyses that assume some interaction between treatments when designing the study. Our proposed power calculation methods can be used for this purpose.

Our models allow for an interaction between two treatments in a stepped wedge design. In many studies, detecting an interaction effect may be of scientific interest. The presence of an interaction generally decreases power for detecting main effects in factorial designs [29]. In a SWFD where the aims include detecting an interaction effect, treatment sequencing is critical. Our examples show that in general, designs in which clusters transition from control to single treatment to combined treatment will be more powerful than designs in which clusters make only one transition, from control to single treatment or control to combined condition. Such multiple-transition designs allow for more within-cluster comparisons, which are a driving factor in power for stepped wedge trials in general.

In this paper, we focus on SWDs with two treatments. However, if every condition can be represented by a binary indicator, the results are generalizable to any number of treatments and two-way interactions. Consider a model with M main effects and B two-way interaction terms. The variance-covariance matrix of the regression coefficients would be a $(M + B) \times (M + B)$ matrix. The elements of this matrix would have the same form as the elements of the 3×3 matrix in Appendix A for diagonal and off-diagonal elements for both main and interaction effects. Solving for $[(M + B) \times (M + B)]^{-1}$ would yield the variance-covariance matrix for the estimated coefficients. Note that this approach holds for two-way interactions only; higher-order interactions are not considered.

We consider continuous outcomes only. Further development is needed for non-continuous outcomes, including binary, survival, categorical and count outcomes. For the nested exchangeable model in Section 2.3, the cluster autocorrelation is constrained to be the same for cluster means across time periods, regardless of the length of time between observing cluster level outcomes. This may not be an accurate assumption, as cluster means observed closer in time may be more correlated than those that are farther apart [2]. There are models for one treatment SWDs that allow the correlation between cluster means to decay over time [30, 31, 32]. For linear mixed models with a decaying correlation structure, the covariance matrix is a Toeplitz matrix and requires the use of the Trench algorithm to numerically invert [31]. We did not include this feature in our work here as we focused on the derivation of closed form variances and covariances of treatment and interaction effects. In this paper, we only consider complete SWDs. Incomplete designs, in which for some periods, data are not collected from some clusters, have been addressed for stepped wedge trials with a single

treatment [33, 34]. Finally, we have assumed that treatment effects are instantaneous and do not consider any delay in treatment effects, which have been considered for SWDs with a single treatment [1, 32, 35]. Future work could explore how delays in one or both treatment effects may impact power of main and interaction effects.

Appendix A: Derivation of Standard Errors of Treatment Effect Estimates

The variance-covariance matrix of the fixed effects in a linear mixed model is found by taking the inverse of $\boldsymbol{Z}^{'}\boldsymbol{V}^{-1}\boldsymbol{Z}$ in which $\boldsymbol{Z}$ is the fixed effects design matrix and $\boldsymbol{V}$ is the variance-covariance matrix of the outcome. In the SWFD model presented in (4), $\boldsymbol{Z}$ is the $IT \times (T+3)$ design matrix and $\boldsymbol{V}$ is the $IT \times IT$ variance-covariance matrix of the outcome. We limit ourselves to finding closed form expressions only for the variance-covariance matrix of the treatment effect estimates, $\hat{\theta}_1, \hat{\theta}_2,$ and $\hat{\theta}_3$. We first find an expression for $\boldsymbol{Z}^{'}\boldsymbol{V}^{-1}\boldsymbol{Z}$ and then use block matrix inversion techniques to get the desired elements of $(\boldsymbol{Z}^{'}\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}$.

Defining the Precision Matrix

We focus on the repeated cross-sectional model. For the nested exchangeable and cohort models, the appropriate values for elements of $\boldsymbol{V}$ can be substituted.

Assuming that clusters are independent, for the repeated cross-sectional model, the matrix $\boldsymbol{V}$ has block diagonal structure with elements $\boldsymbol{V_i} = \sigma_c^2\boldsymbol{I_T} + \sigma_\alpha^2\boldsymbol{1_T}\boldsymbol{1'_T}$, where $\boldsymbol{I_T}$ is a $T \times T$ identity matrix and $\boldsymbol{1_T}$ is a $T \times 1$ vector of 1's. Using the Sherman-Morrison formula [19, 20] for the inverse of a matrix of this form, we obtain

$$\boldsymbol{V_i}^{-1} = \frac{1}{\sigma_c^2(\sigma_c^2 + T\sigma_\alpha^2)}\left[(\sigma_c^2 + T\sigma_\alpha^2)\boldsymbol{I_T} - \sigma_\alpha^2\boldsymbol{1_T}\boldsymbol{1_T^{'}}\right].$$

Due to the the block diagonal structure of $\boldsymbol{V}$, we can write the precision matrix as

$$\boldsymbol{Z}^{'}\boldsymbol{V}^{-1}\boldsymbol{Z} = \sum_{i=1}^{I}\boldsymbol{Z_i^{'}}\boldsymbol{V_i}^{-1}\boldsymbol{Z_i}.$$

The submatrix $\boldsymbol{Z_i}$ is the $T \times (T+3)$ subset of $\boldsymbol{Z}$ corresponding to cluster i. We then can write

$$\boldsymbol{Z_i^{'}}\boldsymbol{V_i}^{-1}\boldsymbol{Z_i} = \frac{1}{\sigma_c^2(\sigma_c^2 + T\sigma_\alpha^2)}\left[(\sigma_c^2 + T\sigma_\alpha^2)\boldsymbol{Z_i^{'}}\boldsymbol{Z_i} - \sigma_\alpha^2\boldsymbol{Z_i^{'}}\boldsymbol{1_T}\boldsymbol{1_T^{'}}\boldsymbol{Z_i}\right]. \tag{10}$$

In the following, the vectors $\boldsymbol{X}, \boldsymbol{W},$ and $\boldsymbol{XW}$ correspond to the columns of the design matrix corresponding to treatment 1, treatment 2, and the interaction term, respectively. Further, let $\boldsymbol{X_i}$ be the $(T \times 1)$ vector that corresponds to cluster i, and $\boldsymbol{X_{i,-T}}$ be the $(T-1) \times 1$ vector for cluster i that does not include the value of $\boldsymbol{X_i}$ at time T. We use similar notation for $\boldsymbol{W_i}$ and $\boldsymbol{W_{i,-T}}$ for treatment 2, and we use $\boldsymbol{XW_i}$ and $\boldsymbol{XW_{i,-T}}$ for the interaction term. Using the summation of submatrices in (10), we can write the precision matrix $\boldsymbol{Z}^{'}\boldsymbol{V}^{-1}\boldsymbol{Z}$ as a $(T+3) \times (T+3)$ symmetric

matrix whose lower triangular elements are

$$\boldsymbol{Z}^{'}\boldsymbol{V}^{-\mathbf{1}}\boldsymbol{Z} = \begin{bmatrix} Tf & & & & & \\ f\mathbf{1}_{\boldsymbol{T-1}} & (f+gT)\boldsymbol{I}_{\boldsymbol{T-1}} - g\mathbf{1}_{\boldsymbol{T-1}}\mathbf{1}_{\boldsymbol{T-1}}^{'} & & & & \\ y_1 & \sum_{i=1}^{I}\frac{\boldsymbol{X}_{\boldsymbol{i,-T}}^{'}}{\sigma_c^2} - \sigma_\alpha^2 h_1 \mathbf{1}_{\boldsymbol{T-1}}^{'} & l_1 - z_1 & & & \\ y_2 & \sum_{i=1}^{I}\frac{\boldsymbol{W}_{\boldsymbol{i,-T}}^{'}}{\sigma_c^2} - \sigma_\alpha^2 h_2 \mathbf{1}_{\boldsymbol{T-1}}^{'} & q_1 & l_2 - z_2 & & \\ y_3 & \sum_{i=1}^{I}\frac{(\boldsymbol{XW})_{\boldsymbol{i,-T}}^{'}}{\sigma_c^2} - \sigma_\alpha^2 h_3 \mathbf{1}_{\boldsymbol{T-1}}^{'} & q_2 & q_3 & l_3 - z_3 \end{bmatrix}.$$

The term $\mathbf{1}_{\boldsymbol{T-1}}$ is a $(T-1)\times 1$ vector of 1's and $\boldsymbol{I}_{\boldsymbol{T-1}}$ is a $(T-1)$ identity matrix, and we define

$$a = \frac{1}{\sigma_c^2 + T\sigma_\alpha^2}, \quad b = \frac{1}{\sigma_c^2}, \quad c = ab, \quad X^T = \sum_{j=1}^{T} X_{ij}, \quad W^T = \sum_{j=1}^{T} W_{ij}, \quad (XW)^T = \sum_{j=1}^{T} X_{ij}W_{ij},$$

$$X^{IT} = \sum_{i=1}^{I}\sum_{j=1}^{T} X_{ij}, \quad W^{IT} = \sum_{i=1}^{I}\sum_{j=1}^{T} W_{ij}, \quad (XW)^{IT} = \sum_{i=1}^{I}\sum_{j=1}^{T} X_{ij}W_{ij}, \quad f = Ia, \quad g = Ic\sigma_\alpha^2,$$

$$y_1 = aX^{IT}, \quad y_2 = aW^{IT}, \quad y_3 = a(XW)^{IT}, \quad h_1 = cX^{IT} \quad h_2 = cW^{IT}, \quad h_3 = c(XW)^{IT}, \quad z_1 = c\sigma_\alpha^2 \sum_{i=1}^{I}(X^T)^2,$$

$$z_2 = c\sigma_\alpha^2 \sum_{i=1}^{I}\left(W^T\right)^2, \quad z_3 = c\sigma_\alpha^2 \sum_{i=1}^{I}\left((XW)^T\right)^2, \quad l_1 = bX^{IT}, \quad l_2 = bW^{IT}, \quad l_3 = b(XW)^{IT},$$

$$q_1 = l_3 - c\sigma_\alpha^2 \sum_{i=1}^{I}\left(X^T\right)\left(W^T\right), \quad q_2 = l_3 - c\sigma_\alpha^2 \sum_{i=1}^{I}\left((XW)^T\right)\left(X^T\right), \quad q_3 = l_3 - c\sigma_\alpha^2 \sum_{i=1}^{I}\left((XW)^T\right)\left(W^T\right).$$

The terms q_2 and q_3 make use of the relationship $X_{ij}^2 = X_{ij}$ and $W_{ij}^2 = W_{ij}$.

Blocking the Precision Matrix

The next step is to partition the precision matrix into a 2×2 block matrix. The matrix $(\boldsymbol{Z'V^{-1}Z})_{11}$ is the $T\times T$ submatrix, $(\boldsymbol{Z}^{'}\boldsymbol{V^{-1}Z})_{21} = (\boldsymbol{Z}^{'}\boldsymbol{V^{-1}Z})_{12}^{'}$ is the $T\times 3$ submatrix and $(\boldsymbol{Z'V^{-1}Z})_{22}$ is the 3×3 submatrix corresponding to the precision of the parameters of interest: $\hat{\theta}_1, \hat{\theta}_2,$ and $\hat{\theta}_3$. Using block matrix inversion [36], we can find the inverse of this submatrix as $(\boldsymbol{Z'V^{-1}Z})_{22}^{-1} = \left((\boldsymbol{Z'V^{-1}Z})_{22} - (\boldsymbol{Z'V^{-1}Z})_{21}(\boldsymbol{Z'V^{-1}Z})_{11}^{-1}(\boldsymbol{Z'V^{-1}Z})_{12}\right)^{-1}$. We first obtain $(\boldsymbol{Z'V^{-1}Z}))_{11}^{-1}$ using another variation of block matrix inversion [36] and Schur complements, yielding

$$(\boldsymbol{Z'V^{-1}Z})_{11}^{-1} = \frac{1}{(f+gT)}\begin{bmatrix} \frac{(g+f)}{f} & -\mathbf{1'}_{\boldsymbol{T-1}} \\ -\mathbf{1}_{\boldsymbol{T-1}} & (\boldsymbol{I}_{\boldsymbol{T-1}} + \mathbf{1}_{\boldsymbol{T-1}}\mathbf{1'}_{\boldsymbol{T-1}}) \end{bmatrix}.$$

Let $\boldsymbol{B} = \boldsymbol{I}_{\boldsymbol{T-1}} + \mathbf{1}_{\boldsymbol{T-1}}\mathbf{1}_{\boldsymbol{T-1}}^{'}$ and let $\boldsymbol{M} = (\boldsymbol{Z'V^{-1}Z})_{21}(\boldsymbol{Z'V^{-1}Z})_{11}^{-1}(\boldsymbol{Z'V^{-1}Z})_{12}$. Using matrix multiplication, we can find

$$\boldsymbol{M} = \frac{1}{(f+gT)}\begin{bmatrix} y_1 & \sum_{i=1}^{I}\frac{\boldsymbol{X}_{\boldsymbol{i,-T}}^{'}}{\sigma_c^2} - \sigma_\alpha^2 h_1 \mathbf{1}_{\boldsymbol{T-1}}^{'} \\ y_2 & \sum_{i=1}^{I}\frac{\boldsymbol{W}_{\boldsymbol{i,-T}}^{'}}{\sigma_c^2} - \sigma_\alpha^2 h_2 \mathbf{1}_{\boldsymbol{T-1}}^{'} \\ y_3 & \sum_{i=1}^{I}\frac{(\boldsymbol{XW})_{\boldsymbol{i,-T}}^{'}}{\sigma_c^2} - \sigma_\alpha^2 h_3 \mathbf{1}_{\boldsymbol{T-1}}^{'} \end{bmatrix} \begin{bmatrix} \frac{(g+f)}{f} & -\mathbf{1}_{\boldsymbol{T-1}}^{'} \\ -\mathbf{1}_{\boldsymbol{T-1}} & \boldsymbol{B} \end{bmatrix} \begin{bmatrix} y_1 & \sum_{i=1}^{I}\frac{\boldsymbol{X}_{\boldsymbol{i,-T}}^{'}}{\sigma_c^2} - \sigma_\alpha^2 h_1 \mathbf{1}_{\boldsymbol{T-1}}^{'} \\ y_2 & \sum_{i=1}^{I}\frac{\boldsymbol{W}_{\boldsymbol{i,-T}}^{'}}{\sigma_c^2} - \sigma_\alpha^2 h_2 \mathbf{1}_{\boldsymbol{T-1}}^{'} \\ y_3 & \sum_{i=1}^{I}\frac{(\boldsymbol{XW})_{\boldsymbol{i,-T}}^{'}}{\sigma_c^2} - \sigma_\alpha^2 h_3 \mathbf{1}_{\boldsymbol{T-1}}^{'} \end{bmatrix}^{'}$$

$$= \frac{1}{(f+gT)} \begin{bmatrix} y_1 & \sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}}' \\ y_2 & \sum_{i=1}^{I} \frac{\boldsymbol{W}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_2 \boldsymbol{1}_{\boldsymbol{T-1}}' \\ y_3 & \sum_{i=1}^{I} \frac{(\boldsymbol{XW})_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_3 \boldsymbol{1}_{\boldsymbol{T-1}}' \end{bmatrix} \begin{bmatrix} \frac{g+f}{f} y_1 - \boldsymbol{1}_{\boldsymbol{T-1}}' \left(\sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}} \right) & -y_1 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \\ \frac{g+f}{f} y_2 - \boldsymbol{1}_{\boldsymbol{T-1}}' \left(\sum_{i=1}^{I} \frac{\boldsymbol{W}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_2 \boldsymbol{1}_{\boldsymbol{T-1}} \right) & -y_2 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{\boldsymbol{W}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_2 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \\ \frac{g+f}{f} y_3 - \boldsymbol{1}_{\boldsymbol{T-1}}' \left(\sum_{i=1}^{I} \frac{(\boldsymbol{XW})_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_3 \boldsymbol{1}_{\boldsymbol{T-1}} \right) & -y_3 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{(\boldsymbol{XW})_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_3 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \end{bmatrix}'.$$

Solving for the $-\boldsymbol{1}_{\boldsymbol{T-1}}' \left(\sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}} \right)$ in the second matrix (similar for other elements), we obtain

$$= \left(\sum_{i=j}^{T-1} \sum_{i=1}^{I} \frac{X_{ij}}{\sigma_c^2} - h\sigma_\alpha^2 \boldsymbol{1}_{\boldsymbol{T-1}}' \boldsymbol{1}_{\boldsymbol{T-1}} \right) = \left(\sum_{i=j}^{T-1} \sum_{i=1}^{I} \frac{X_{ij}}{\sigma_c^2} - \sigma_\alpha^2 (T-1) \sum_{i=1}^{I} \sum_{j=1}^{T} \frac{X_{ij}}{\sigma_c^2(\sigma_c^2 + T\sigma_\alpha^2)} \right)$$

$$= \left(\sum_{i=j}^{T} \sum_{i=1}^{I} \frac{\sigma_c^2 X_{ij} + T\sigma_\alpha^2 X_{ij}}{\sigma_c^2(\sigma_c^2 + T\sigma_\alpha^2)} - \frac{\sigma_\alpha^2 T X_{ij} - \sigma_\alpha^2 X_{ij}}{\sigma_c^2(\sigma_c^2 + T\sigma_\alpha^2)} \right) - \sum_{i=1}^{I} \frac{X_{iT}}{\sigma_c^2} = \left(\sum_{i=j}^{T} \sum_{i=1}^{I} \frac{X_{ij}}{(\sigma_c^2 + T\sigma_\alpha^2)} \right) + \left(\sum_{i=j}^{T} \sum_{i=1}^{I} \frac{\sigma_\alpha^2 X_{ij}}{\sigma_c^2(\sigma_c^2 + T\sigma_\alpha^2)} \right) - \sum_{i=1}^{I} \frac{X_{iT}}{\sigma_c^2}$$

$$= y_1 + \sigma_\alpha^2 h_1 - \sum_{i=1}^{I} \frac{X_{iT}}{\sigma_c^2}.$$

Substituting this term back into $\boldsymbol{M}$ yields $\boldsymbol{M} = \frac{1}{(f+gT)} \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$ with elements

$$m_{11} = y_1 \left(\frac{g+f}{f} y_1 + \left(-y_1 - \sigma_\alpha^2 h_1 + \sum_{i=1}^{I} \frac{X_{iT}}{\sigma_c^2} \right) \right) + \left(\sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}}' \right) \left(-y_1 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \right)$$

$$m_{12} = y_1 \left(\frac{g+f}{f} y_2 + \left(-y_2 - \sigma_\alpha^2 h_2 + \sum_{i=1}^{I} \frac{W_{iT}}{\sigma_c^2} \right) \right) + \left(\sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}}' \right) \left(-y_2 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{\boldsymbol{W}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_2 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \right)$$

$$m_{13} = y_1 \left(\frac{g+f}{f} y_3 + \left(-y_3 - \sigma_\alpha^2 h_3 + \sum_{i=1}^{I} \frac{(XW)_{iT}}{\sigma_c^2} \right) \right) + \left(\sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}}' \right) \left(-y_3 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{(\boldsymbol{XW})_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_3 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \right)$$

$$m_{21} = y_2 \left(\frac{g+f}{f} y_1 - \left(y_1 + \sigma_\alpha^2 h_1 - \sum_{i=1}^{I} \frac{X_{iT}}{\sigma_c^2} \right) \right) + \left(\sum_{i=1}^{I} \frac{\boldsymbol{W}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_2 \boldsymbol{1}_{\boldsymbol{T-1}}' \right) \left(-y_1 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \right)$$

$$m_{22} = y_2 \left(\frac{g+f}{f} y_2 - \left(y_2 + \sigma_\alpha^2 h_2 - \sum_{i=1}^{I} \frac{W_{iT}}{\sigma_c^2} \right) \right) + \left(\sum_{i=1}^{I} \frac{\boldsymbol{W}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_2 \boldsymbol{1}_{\boldsymbol{T-1}}' \right) \left(-y_2 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{\boldsymbol{W}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_2 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \right)$$

$$m_{23} = y_2 \left(\frac{g+f}{f} y_3 - \left(y_3 + \sigma_\alpha^2 h_3 - \sum_{i=1}^{I} \frac{(XW)_{iT}}{\sigma_c^2} \right) \right) + \left(\sum_{i=1}^{I} \frac{\boldsymbol{W}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_2 \boldsymbol{1}_{\boldsymbol{T-1}}' \right) \left(-y_3 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{(\boldsymbol{XW})_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_3 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \right)$$

$$m_{31} = y_3 \left(\frac{g+f}{f} y_1 - \left(y_1 + \sigma_\alpha^2 h_1 - \sum_{i=1}^{I} \frac{X_{iT}}{\sigma_c^2} \right) \right) + \left(\sum_{i=1}^{I} \frac{\boldsymbol{XW}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_3 \boldsymbol{1}_{\boldsymbol{T-1}}' \right) \left(-y_1 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{\boldsymbol{X}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_1 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \right)$$

$$m_{32} = y_3 \left(\frac{g+f}{f} y_2 - \left(y_2 + \sigma_\alpha^2 h_2 - \sum_{i=1}^{I} \frac{W_{iT}}{\sigma_c^2} \right) \right) + \left(\sum_{i=1}^{I} \frac{\boldsymbol{XW}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_3 \boldsymbol{1}_{\boldsymbol{T-1}}' \right) \left(-y_2 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{\boldsymbol{W}_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_2 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \right)$$

$$m_{33} = y_3 \left(\frac{g+f}{f} y_3 - \left(y_3 + \sigma_\alpha^2 h_3 - \sum_{i=1}^{I} \frac{(XW)_{iT}}{\sigma_c^2} \right) \right) + \left(\sum_{i=1}^{I} \frac{\boldsymbol{XW}_{\boldsymbol{i},-\boldsymbol{T}}'}{\sigma_c^2} - \sigma_\alpha^2 h_3 \boldsymbol{1}_{\boldsymbol{T-1}}' \right) \left(-y_3 \boldsymbol{1}_{\boldsymbol{T-1}} + \boldsymbol{B} \left(\sum_{i=1}^{I} \frac{(\boldsymbol{XW})_{\boldsymbol{i},-\boldsymbol{T}}}{\sigma_c^2} - \sigma_\alpha^2 h_3 \boldsymbol{1}_{\boldsymbol{T-1}} \right) \right).$$

Simplifying the diagonal elements of $\boldsymbol{M}$

The three diagonal elements of $\boldsymbol{M}$ have the same form, so we will solve for the (1,1) element of the matrix corresponding to treatment 1 and apply the form to all three diagonal elements. Let $\eta_1 = y_1+\sigma_\alpha^2 h_1-\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}$, then the diagonal element is:

$$= \frac{1}{(f+gT)}\left(y_1\left(\frac{g+f}{f}y_1-\eta_1\right)+\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_1\boldsymbol{1_{T-1}}\right)\left(-y_1\boldsymbol{1_{T-1}}+\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_1\boldsymbol{1_{T-1}}\right)\right)\right)$$

$$= \frac{1}{(f+gT)}\left(\frac{y_1^2(f+g)}{f}-2y_1\eta_1+\left(\sum_{i=1}^{I}\frac{\boldsymbol{X'_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_1\boldsymbol{1'_{T-1}}\right)\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_1\boldsymbol{1_{T-1}}\right)\right).$$

Further,

$$\left(\sum_{i=1}^{I}\frac{\boldsymbol{X'_{i,-T}}}{\sigma_c^2}-\frac{l_1-y_1}{T}\boldsymbol{1'_{T-1}}\right)\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1-y_1}{T}\boldsymbol{1_{T-1}}\right)$$

$$= \left(\sum_{i=1}^{I}\frac{\boldsymbol{X'_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1'_{T-1}}\right)\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}\right)+\frac{y_1}{T}\boldsymbol{1'_{T-1}}\boldsymbol{B}\frac{y_1}{T}\boldsymbol{1_{T-1}}+2\frac{y_1}{T}\boldsymbol{1'_{T-1}}\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}\right).$$

We solve for each term individually. We will make use of the relationship $\sum_{i=1}^{I}\sum_{j=1}^{T-1}\frac{X_{ij}}{\sigma_c^2} = l_1-\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}$. Solving for the first term, $\left(\sum_{i=1}^{I}\frac{\boldsymbol{X'_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1'_{T-1}}\right)\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}\right)$

$$= \sum_{j=1}^{T-1}\left(\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\right)^2+l_1^2-2l_1\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}+\left(\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}\right)^2-2l_1^2+2l_1\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}+l_1^2-\frac{l_1^2}{T} \quad = \quad \sum_{j=1}^{T}\left(\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\right)^2-\frac{l_1^2}{T}.$$

Letting $w_1 = \sum_{j=1}^{T}\left(\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\right)^2$, the first term becomes $w_1-\frac{l_1^2}{T}$. Similarly for other diagonal elements, let $w_2 = \sum_{j=1}^{T}\left(\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2}\right)^2$ and $w_3 = \sum_{j=1}^{T}\left(\sum_{i=1}^{I}\frac{(XW)_{ij}}{\sigma_c^2}\right)^2$. Solving the second term, $\frac{y_1}{T}\boldsymbol{1'_{T-1}}\boldsymbol{B}\frac{y_1}{T}\boldsymbol{1_{T-1}}$, yields

$$= \frac{y_1^2}{T^2}\boldsymbol{1'_{T-1}}(\boldsymbol{I_{T-1}}+\boldsymbol{1_{T-1}}\boldsymbol{1'_{T-1}})\boldsymbol{1_{T-1}} \;=\; \frac{y_1^2}{T^2}\left((T-1)+(T-1)\boldsymbol{1'_{T-1}}\boldsymbol{1_{T-1}}\right) \;=\; \frac{y_1^2}{T^2}(T-1+T^2-2T+1) \;=\; \frac{y_1^2}{T}(T-1).$$

Solving for the third term, $\frac{2y_1}{T}\boldsymbol{1'_{T-1}}\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}\right)$, yields

$$\frac{2y_1}{T}(\boldsymbol{1'_{T-1}}+(T-1)\boldsymbol{1'_{T-1}})\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}\right) \;=\; 2y_1\boldsymbol{1'_{T-1}}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}\right)$$

$$= 2y_1\left(-\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}-\frac{l_1}{T}+\frac{y_1}{T}-\frac{y_1}{T}\right) = 2y_1\left(\frac{y_1}{T}+\sigma_\alpha^2 h_1-\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}\right).$$

Substituting these terms back in the diagonal element of $\boldsymbol{M}$, we obtain

$$= \frac{1}{(f+gT)}\left(\frac{y_1^2(f+g)}{f}-2y_1\eta_1+\left(\sum_{i=1}^{I}\frac{\boldsymbol{X'_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_1\boldsymbol{1'_{T-1}}\right)\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_1\boldsymbol{1_{T-1}}\right)\right)$$

$$= \frac{1}{(f+gT)}\left(\frac{y_1^2(f+g)}{f} - 2y_1\left(y_1+\sigma_\alpha^2 h_1+\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}\right)+w_1-\frac{l_1^2}{T}+2y_1\left(\frac{y_1}{T}+\sigma_\alpha^2 h_1-\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}\right)+\frac{y_1^2}{T}(T-1)\right)$$

$$= \frac{1}{(f+gT)}\left(y_1^2\left(\frac{f+gT}{fT}\right)+w_1-\frac{l_1^2}{T}\right) \;=\; \frac{y_1^2}{fT}+\frac{1}{(f+gT)}\left(w_1-\frac{l_1^2}{T}\right).$$

The second and third diagonal elements of the matrix $\boldsymbol{M}$ will have the same form.

Simplifying the off-diagonal elements of $\boldsymbol{M}$

To solve for the off-diagonal values, we consider one value, m_{21}, and apply the results to the other off-diagonals. The value of m_{21} is

$$\frac{1}{(f+gT)}\left(y_2\left(\frac{g+f}{f}y_1-y_1-\sigma_\alpha^2 h_1+\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}\right)+\left(\sum_{i=1}^{I}\frac{\boldsymbol{W_{i,-T}}'}{\sigma_c^2}-\sigma_\alpha^2 h_2\boldsymbol{1'_{T-1}}\right)\left(-y_1\boldsymbol{1_{T-1}}+\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_1\boldsymbol{1_{T-1}}\right)\right)\right).$$

We will first simplify the term $y_2\left(\frac{g+f}{f}y_1-y_1-\sigma_\alpha^2 h_1+\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}\right) = y_1y_2\frac{f+g}{f}-y_1y_2-\sigma_\alpha^2 h_1y_2+y_2\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}$. To simplify the second term, $\left(\sum_{i=1}^{I}\frac{\boldsymbol{W_{i,-T}}'}{\sigma_c^2}-\sigma_\alpha^2 h_2\boldsymbol{1'_{T-1}}\right)\left(-y_1\boldsymbol{1_{T-1}}+\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_1\boldsymbol{1_{T-1}}\right)\right)$, we will break up this multiplication into different terms. The first step is multiplying two terms, yielding

$$\left(\sum_{i=1}^{I}\frac{\boldsymbol{W'_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_2\boldsymbol{1'_{T-1}}\right)(-y_1\boldsymbol{1_{T-1}}) \;=\; -y_1\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2}+(T-1)\sigma_\alpha^2 h_2 y_1.$$

The rest of the multiplication $\left(\sum_{i=1}^{I}\frac{\boldsymbol{W_{i,-T}}'}{\sigma_c^2}-\sigma_\alpha^2 h_2\boldsymbol{1'_{T-1}}\right)\left(\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\sigma_\alpha^2 h_1\boldsymbol{1_{T-1}}\right)\right)$ simplifies to

$$= \left(\sum_{i=1}^{I}\frac{\boldsymbol{W_{i,-T}}'}{\sigma_c^2}-\frac{l_2-y_2}{T}\boldsymbol{1'_{T-1}}\right)\left(\left(\boldsymbol{I_{T-1}}+\boldsymbol{1_{T-1}1'_{T-1}}\right)\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1-y_1}{T}\boldsymbol{1_{T-1}}\right)\right)$$

$$= \left(\sum_{i=1}^{I}\frac{\boldsymbol{W_{i,-T}}'}{\sigma_c^2}-\frac{l_2}{T}\boldsymbol{1'_{T-1}}+\frac{y_2}{T}\boldsymbol{1'_{T-1}}\right)\left(\left(\boldsymbol{I_{T-1}}+\boldsymbol{1_{T-1}1'_{T-1}}\right)\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}+\frac{y_1}{T}\boldsymbol{1_{T-1}}\right)\right)$$

$$= \left(\sum_{i=1}^{I}\frac{\boldsymbol{W_{i,-T}}'}{\sigma_c^2}-\frac{l_2}{T}\boldsymbol{1_{T-1}}'\right)\left(\left(\boldsymbol{I_{T-1}}+\boldsymbol{1_{T-1}1'_{T-1}}\right)\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}\right)\right)+$$

$$\frac{y_2}{T}\boldsymbol{1'_{T-1}}(\boldsymbol{I_{T-1}}+\boldsymbol{1_{T-1}1'_{T-1}})\frac{y_1}{T}\boldsymbol{1_{T-1}}+\frac{y_2}{T}\boldsymbol{1'_{T-1}}(\boldsymbol{I_{T-1}}+\boldsymbol{1_{T-1}1'_{T-1}})\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}\right)+$$

$$\left(\sum_{i=1}^{I}\frac{\boldsymbol{W_{i,-T}}'}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1'_{T-1}}\right)(\boldsymbol{I_{T-1}}+\boldsymbol{1_{T-1}1'_{T-1}})(\frac{y_1}{T}\boldsymbol{1_{T-1}}).$$

We break this multiplication up by individual terms, with the first term simplifying to

$$\left(\sum_{i=1}^{I}\frac{\boldsymbol{W_{i,-T}}'}{\sigma_c^2}-\frac{l_2}{T}\boldsymbol{1'_{T-1}}\right)\left(\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2}-\frac{l_1}{T}\boldsymbol{1_{T-1}}\right)\right)$$

$$= \sum_{j=1}^{T-1}\left(\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2}\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\right) + \left(\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2}\right)\left(\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\right) - \frac{l_1}{T}\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2} -$$

$$\frac{l_1(T-1)}{T}\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2} - \frac{l_2}{T}\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2} - \frac{l_2(T-1)}{T}\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2} + \frac{l_1 l_2(T-1)}{T^2} + \frac{l_1 l_2(T-1)^2}{T^2}$$

$$= \sum_{j=1}^{T-1}\left(\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2}\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\right) + \left(\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2}\right)\left(\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\right) - l_1\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2} - l_2\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2} + \frac{(T-1)l_1 l_2}{T}$$

$$= w_{XW} - \sum_{j=1}^{T}\left(\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2}\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\right) - \frac{l_1 l_2}{T}, \text{ where } \quad w_{XW} = \sum_{j=1}^{T}\left(\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2}\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\right).$$

Let $w_{X(XW)} = \sum_{j=1}^{T}\left(\sum_{i=1}^{I}\frac{X_{ij}}{\sigma_c^2}\sum_{i=1}^{I}\frac{(XW)_{ij}}{\sigma_c^2}\right)$ and $w_{W(XW)} = \sum_{j=1}^{T}\left(\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2}\sum_{i=1}^{I}\frac{(XW)_{ij}}{\sigma_c^2}\right)$ for the other off-diagonal elements. Simplifying the remaining terms,

$$\frac{y_2}{T}\boldsymbol{1'_{T-1}}\boldsymbol{B}\frac{y_1}{T}\boldsymbol{1_{T-1}} + \frac{y_2}{T}\boldsymbol{1'_{T-1}}\boldsymbol{B}\left(\sum_{i=1}^{I}\frac{\boldsymbol{X_{i,-T}}}{\sigma_c^2} - \frac{l_1}{T}\boldsymbol{1_{T-1}}\right) + \left(\sum_{i=1}^{I}\frac{\boldsymbol{W_{i,-T}}'}{\sigma_c^2} - \frac{l_2}{T}\boldsymbol{1'_{T-1}}\right)\boldsymbol{B}(\frac{y_1}{T}\boldsymbol{1_{T-1}})$$

$$= \frac{y_1 y_2(T-1)}{T} + y_2\left(\frac{y_1}{T} + \sigma_\alpha^2 h_1 - \sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}\right) + y_1\left(\frac{y_2}{T} + \sigma_\alpha^2 h_2 - \sum_{i=1}^{I}\frac{Y_{iT}}{\sigma_c^2}\right).$$

Putting all the terms together for the off-diagonals, we get:

$$= y_1 y_2\frac{f+g}{f} - y_1 y_2 - \sigma_\alpha^2 h_1 y_2 + y_2\sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2} - y_1\sum_{j=1}^{T-1}\sum_{i=1}^{I}\frac{W_{ij}}{\sigma_c^2} + (T-1)\sigma_\alpha^2 h_2 y_1 +$$

$$\frac{y_1 y_2(T-1)}{T} + y_2\left(\frac{y_1}{T} + \sigma_\alpha^2 h_1 - \sum_{i=1}^{I}\frac{X_{iT}}{\sigma_c^2}\right) + y_1\left(\frac{y_2}{T} + \sigma_\alpha^2 h_2 - \sum_{i=1}^{I}\frac{Y_{iT}}{\sigma_c^2}\right) + w_{xy} - \frac{l_1 l_2}{T}$$

$$= y_1 y_2\frac{f+gT+fT}{fT} - y_1 l_2 + T\sigma_\alpha^2 h_2 y_1 + w_{xy} - \frac{l_1 l_2}{T} \;=\; y_1 y_2\left(\frac{f+gT}{fT}\right) + w_{xy} - \frac{l_1 l_2}{T}.$$

Multiplying this by the constant $\frac{1}{f+gT}$, the off-diagonal element becomes $\frac{y_1 y_2}{fT} + \frac{1}{f+gT}\left(w_{xy} - \frac{l_1 l_2}{T}\right)$.

Final Algebra

We now have a simplified expression for $\boldsymbol{M}$. To obtain $(\boldsymbol{Z'V^{-1}Z})_{22}^{-1}$, we calculate $(\boldsymbol{Z'V^{-1}Z})_{22} - \boldsymbol{M}$ and take the inverse of this matrix. We have $(\boldsymbol{Z'V^{-1}Z})_{22} - \boldsymbol{M}$ equal to

$$= \begin{bmatrix} l_1 - z_1 - \frac{y_1^2}{fT} - \frac{1}{f+gT}\left(w_1 - \frac{l_1^2}{T}\right) & q_1 - \frac{y_1 y_2}{fT} - \frac{1}{f+gT}\left(w_{XW} - \frac{l_1 l_2}{T}\right) & q_2 - \frac{y_1 y_3}{fT} - \frac{1}{f+gT}\left(w_{X(XW)} - \frac{l_1 l_3}{T}\right) \\ q_1 - \frac{y_1 y_2}{fT} - \frac{1}{f+gT}\left(w_{XW}\frac{l_1 l_2}{T}\right) & l_2 - z_2 - \frac{y_2^2}{fT} - \frac{1}{f+gT}\left(w_2 - \frac{l_2^2}{T}\right) & q_3 - \frac{y_2 y_3}{fT} - \frac{1}{f+gT}\left(w_{W(XW)} - \frac{l_2 l_3}{T}\right) \\ q_2 - \frac{y_1 y_3}{fT} - \frac{1}{f+gT}\left(w_{X(XW)}\frac{l_1 l_3}{T}\right) & q_3 - \frac{y_2 y_3}{fT} - \frac{1}{f+gT}\left(w_{W(XW)}\frac{l_2 l_3}{T}\right) & l_3 - z_3 - \frac{y_3^2}{fT} - \frac{1}{f+gT}\left(w_3 - \frac{l_3^2}{T}\right) \end{bmatrix}.$$

We can now solve for $(\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}_{22}$ by taking the inverse of this 3×3 matrix. If treatment effects are assumed to be additive and an interaction term is not included, we can solve for the upper 2×2 matrix for the variance-covariance matrix of the regression coefficients for two main treatment effects. For a one treatment repeated cross-sectional stepped wedge design, the variance of the treatment effect would be the reciprocal of the diagonal term, $l_1 - z_1 - \frac{y_1^2}{fT} - \frac{1}{f+gT}\left(w_1 - \frac{l_1^2}{T}\right)$, which can be shown to be the same variance as solved for by Hussey and Hughes [1].

Conflict of Interest

No potential conflict of interest was reported by the authors.

References

[1] M. Hussey and J. Hughes, "Design and analysis of stepped wedge cluster randomized trials," *Contemporary Clinical Trials*, vol. 28, pp. 182–191, 2007.

[2] K. Hemming, M. Taljaard, and A. Forbes, "Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples," *Trials*, vol. 18, no. 101, 2017.

[3] W. Woertman, E. de Hoop, M. Moerbeek, S. Zuidemac, D. Gerritsen, and S. Teerenstra, "Stepped wedge designs could reduce the required sample size in cluster randomized trials," *Journal of Clinical Epidemiology*, vol. 66, pp. 752–758, 2013.

[4] M. Grayling, J. Wason, and A. Mander, "Stepped wedge cluster randomized controlled trial designs: a review of reporting quality and design features," *Trials*, vol. 18, p. 33, 2017.

[5] M. Grayling, A. Mander, and J. Wason, "Admissible multiarm stepped-wedge cluster randomized trial designs," *Statistics in Medicine*, vol. 38, pp. 1103–1119, 2018.

[6] P. Zhang, A. Shoben, R. Jackson, and S. Fernandez, "Variance formulae for multiphase stepped wedge cluster randomized trial," *Statistics in Medicine*, 2020.

[7] V. Lyons, L. Li, J. Hughes, and A. Rowhani-Rahbar, "Proposed variations of the stepped-wedge design can be used to accommodate multiple interventions.," *Journal of Clinical Epidemiology*, vol. 86, pp. 160–167, 2017.

[8] Patient-Centered Outcomes Research Institute, *Comparing Strategies for Health Clinics to Increase HPV Vaccinations in Youth*, 2017.

[9] B. van der Geest, J. de Graaf, L. Bertens, M. Poley, E. Ista, R. Kornelisse, I. Reiss, E. Steegers, and J. Been, "Screening and treatment to reduce severe hyperbilirubinaemia in infants in primary care (starship): a factorial stepped-wedge cluster randomised controlled trial protocol," *BMJ Open*, vol. 9, 2019.

[10] K. Whittingham, M. Sanders, L. McKinlay, and R. Boyd, "Interventions to reduce behavioral problems in children with cerebral palsy: An rct," *Pediatrics*, vol. 133, pp. 1249–1257, 2014.

[11] S. Reuther, D. Holle, I. Buscher, O. Dortmann, R. Müller, S. Bartholomeyczik, and M. Halek, "Effect evaluation of two types of dementia-specific case conferences in german nursing homes (falldem) using a stepped-wedge design: study protocol for a randomized controlled trial," *Trials*, vol. 15, no. 319, 2014.

[12] B. Durovni, V. Saraceni, L. Moulton, A. Pacheco, S. Cavalcante, B. King, S. Cohn, A. Efron, R. Chaisson, and J. Golub, "Effect of improved tuberculosis screening and isoniazid preventive therapy on incidence of tuberculosis and death in patients with hiv in clinics in rio de janeiro, brazil: a stepped wedge, cluster-randomised trial," *The Lancet Infectious Disease*, vol. 13, pp. 852–858, 2013.

[13] B. Durovni, V. Saraceni, S. van den Hof, A. Trajman, M. Cordeiro-Santos, S. Cavalcante, A. Menezes, and F. Cobelens, "Impact of replacing smear microscopy with xpert mtb/rif for diagnosing tuberculosis in brazil: a stepped-wedge cluster-randomized trial," *PLoS Med*, vol. 12, 2014.

[14] G. Oelhert, *A First Course in Design and Analysis of Experiments*, ch. Chapter 8, pp. 170–171. W. H. Freeman, 2010.

[15] S. Teerenstra, S. Eldridge, M. Graff, E. de Hoop, and G. Borma, "A simple sample size formula for analysis of covariance in cluster randomized trials," *Statistics in Medicine*, vol. 31, pp. 2169–2178, 2012.

[16] H. Feldman and S. McKinlay, "Cohort versus cross-sectional design in large field trials: precision, sample size, and a unifying model," *Statistics in Medicine*, vol. 13, pp. 61–78, 1994.

[17] A. Copas, J. Lewis, J. Thompson, C. Davey, G. Baio, and J. Hargreaves, "Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches," *Trials*, vol. 16, pp. 95–113, 2015.

[18] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*, ch. 2, pp. 56–57. Springer-Verlag, 2009.

[19] J. Sherman and W. Morrison, "Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix," *Annals of Mathematical Statistics*, vol. 20, no. 4, pp. 620–624, 1949.

[20] M. Bartlett, "An inverse matrix adjustment arising in discriminant analysis," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 107–111, 1951.

[21] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

[22] M. Taljaard, S. Teerenstra, N. Ivers, and D. Fergusson, "Substantial risks associated with few clusters in cluster randomized and stepped wedge designs," *Clinical Trials*, vol. 13, pp. 459–463, 2016.

[23] R. Hooper, S. Teerenstra, E. de Hoop, and S. Eldridge, "Sample size calculation for stepped wedge and other longitudinal cluster randomised trials," *Statistics in Medicine*, 2016.

[24] H. Feldman and S. McKinlay, "Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for designs," *Statistics in Medicine*, vol. 11, pp. 1685–1704, 1992.

[25] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, ch. Chapter 2, pp. 24–27. L. Erlbaum Associates, 1988.

[26] K. Hemming and M. Taljaard, "Sample size calculations for stepped wedge and cluster randomised trials: a unified approach," *Journal of Clinical Epidemiology*, vol. 69, pp. 137–146, 2016.

[27] G. Baio, A. Copas, G. Ambler, J. Hargreaves, E. Beard, and R. Omar, "Sample size calculation for a stepped wedge trial," *Trials*, vol. 16, no. 354, 2015.

[28] C. Dunnett, "A multiple comparison procedure for comparing several treatments with a control," *Biometrics*, vol. 20, pp. 482–491, 1964.

[29] S. Green, P. Liu, and J. O'Sullivan, "Factorial design considerations," *Journal of Clinical Oncology*, vol. 20, pp. 3424–3230, 2002.

[30] F. Li, "Design and analysis considerations for cohort stepped wedgecluster randomized trials with a decay correlation structure," *Statistics in Medicine*, vol. 39, pp. 438–455, 2019.

[31] K. Grantham, J. Kasza, S. Heritier, K. Hemming, and A. Forbes, "Accounting for a decaying correlation structure in cluster randomized trials with continuous recruitment," *Statistics in Medicine*, vol. 38, pp. 1918–1934, 2018.

[32] F. Li, J. Hughes, K. Hemming, M. Taljaard, E. Melnick, and P. Heagerty, "Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview," *Statistical Methods in Medical Research*, vol. 0, pp. 1–28, 2020.

[33] K. Hemming, R. Lilford, and A. J. Girling, "Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple level designs," *Statistics in Medicine*, vol. 34, p. 181–196, 2014.

[34] J. Kasza, M. Taljaard, and A. Forbes, "Information content of stepped-wedge designs when treatment effect heterogeneity and/or implementation periods are present," *Statistics in Medicine*, 2019.

[35] J. Hughes, T. Granston, and P. Heagerty, "Current issues in the design and analysis of stepped wedge trials," *Contemporary Clinical Trials*, vol. 45(Part A), pp. 55–60, 2015.

[36] T. Lu and S. Shiou, "Inverses of 2x2 block matrices," *Computers and Mathematics with Applications*, vol. 43, pp. 119–129, 2002.