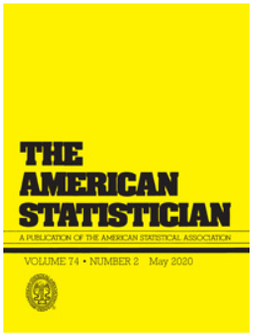

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

Time Parameterizations in Cluster Randomized Trial Planning

Kelsey L. Grantham, Andrew B. Forbes, Stephane Heritier & Jessica Kasza

To cite this article: Kelsey L. Grantham, Andrew B. Forbes, Stephane Heritier & Jessica Kasza (2020) Time Parameterizations in Cluster Randomized Trial Planning, The American Statistician, 74:2, 184-189, DOI: 10.1080/00031305.2019.1623072

To link to this article: https://doi.org/10.1080/00031305.2019.1623072

Accepted author version posted online: 28 May 2019.
Published online: 05 Jul 2019.

THE AMERICAN STATISTICIAN
2020, VOL. 74, NO. 2, 184–189: Short Technical Note
https://doi.org/10.1080/00031305.2019.1623072

Taylor & Francis
Taylor & Francis Group

SHORT TECHNICAL NOTE

Time Parameterizations in Cluster Randomized Trial Planning

Kelsey L. Grantham, Andrew B. Forbes, Stephane Heritier, and Jessica Kasza

School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

ABSTRACT

Models for cluster randomized trials conducted over multiple time periods should account for underlying temporal trends. However, in practice there is often limited knowledge or data available to inform the choice of time parameterization of these trends, or to anticipate the implications of this choice on trial planning. In this article, we establish a sufficient condition for when the choice of time parameterization does not affect the form of the variance of the treatment effect estimator, thereby simplifying the planning of these trials.

ARTICLE HISTORY
Received December 2018
Revised March 2019

KEYWORDS
Column space; Design matrix; Invariance

1. Introduction

In a cluster randomized trial, groups of subjects, or clusters, such as hospitals, schools, or villages are randomized to intervention or control treatments. When planning a cluster randomized trial conducted over multiple periods, it is now well-recognized that it is necessary to allow for underlying temporal trends by including one or more terms in the statistical model for the likely effect of time on the outcome (Hemming et al. 2015). There are multiple potential choices for the parameterization of time in the model: separate time effects for each treatment period, a linear term for time, or even more complex fractional polynomial or spline forms for time. However, there may be limited data or knowledge available to help guide this choice. In this article, we delineate the situations in which the precise choice of time parameterization does not alter the variance of the treatment effect estimator, a key component of sample size calculations at the planning stage of a trial.

Models for cluster randomized trials tend to treat time as discrete, accounting for broad changes in subjects' outcomes over the treatment periods and allowing for the correlation between subjects' outcomes in a cluster to depend on subjects' periods of measurement (Hooper et al. 2016). Time could be treated as discrete for several possible reasons: the trial measures groups of subjects within each cluster simultaneously at discrete intervals (e.g., the Devon Active Villages Evaluation (DAVE) trial (Solomon et al. 2014) where subjects received questionnaires about their physical activity at discrete times throughout the trial), subjects' outcomes are only available at an aggregate level over each period (Maxwell et al. 2017), or to enable a simple analysis using cluster-period summary data.

In other situations, it may be more natural to consider continuous time such as through the use of time parameterizations that depend on subjects' measurement times or by accounting for subjects' measurement times when modeling the similarity between subjects' outcomes in a cluster, that is, the within-cluster correlation structure. Continuous-time models for cluster randomized trials have not yet been considered extensively in the literature (Grantham et al. 2019), yet the model provides a more general framework and so it is useful to also consider continuous time in this article. This article is organized as follows: in Section 2 we introduce a discrete-time model for cluster randomized trials, some common time parameterizations, and a theorem giving a sufficient condition for when the variance of the treatment effect estimator expression is invariant to the chosen time parameterization. Section 3 outlines the implications for trial planning for several cluster randomized trial designs. We then explain how the model and results would change under a continuous-time framework in Section 4 and provide some concluding remarks in Section 5.

2. Cluster Randomized Trials and Time Parameterizations

2.1. Trial Configuration

Consider a cluster randomized trial on N clusters, where the trial duration is divided into T discrete treatment periods. The allocation of clusters to the intervention or control condition may change with each period, depending on the design chosen. Each possible set of conditions over the trial is a potential treatment sequence, and a trial design is characterized by its particular set of treatment sequences. Let X_{ij} denote the treatment indicator (0 = Control, 1 = Intervention) for cluster $i = 1, \ldots, N$ in period $j = 1, \ldots, T$ and let $\mathbf{X}_i = (X_{i1}, \ldots, X_{iT})'$ be the treatment sequence for cluster i. We assume an equal number of subjects m are measured in each cluster in each period: these subjects may be measured repeatedly (i.e., in multiple treatment periods as in a cohort design (Copas et al. 2015)) or just once. Unless otherwise stated, we will assume that each subject is measured just once, in one period only, and hence that a different set of subjects is measured in each period. We explicitly consider cohort designs in Section 2.3.

CONTACT Jessica Kasza jessica.kasza@monash.edu School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Rd, Melbourne 3004, Australia.

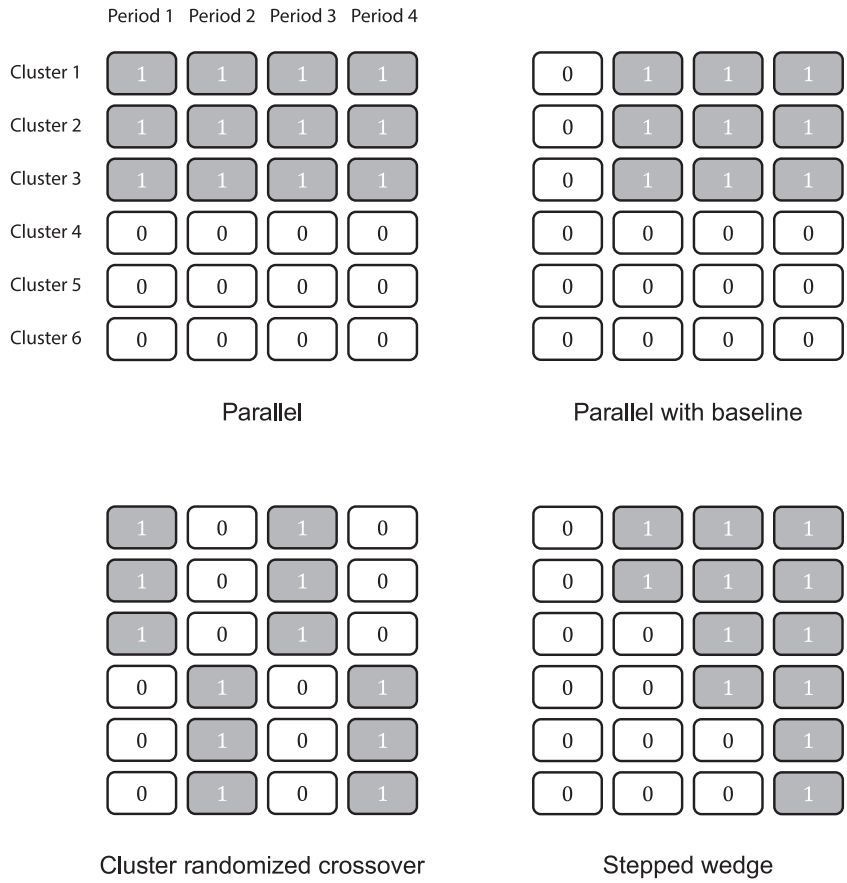

Figure 1. Sets of treatment sequences for cluster randomized trial designs with six clusters and four periods. The rows of each block are the treatment sequences assigned to each cluster (0 = Control, 1 = Intervention) and the columns mark the periods of the trial.

Common cluster randomized trial designs include the parallel, parallel with baseline, cluster randomized crossover, and stepped wedge designs. Figure 1 depicts example schematics for each of these four trial designs. In a parallel design, half of the clusters implement the control condition and the other half implement the intervention condition for the entire trial. A parallel with baseline design is a variation of the parallel design whereby all clusters initially implement the control condition for at least one period, before half of the clusters switch to implementing the intervention condition. A cluster randomized crossover design typically has treatment sequences that alternate between the intervention and control with each subsequent period, with half of the clusters implementing the control in the first period and the other half implementing the intervention in the first period. In a classic stepped wedge design, all clusters implement the control condition in the first period, and clusters switch over to the intervention condition in a staggered manner until the last period, in which all clusters implement the intervention condition.

2.2. Discrete-Time Model and Variance Expression

We consider continuous outcomes, and assume the measured outcome Y_{ijk} from cluster $i = 1, \ldots, N$ in time period $j = 1, \ldots, T$ corresponding to subject $k = 1, \ldots, m$ can be represented by the following model

$$Y_{ijk} = \mathbf{Z}_j\boldsymbol{\beta} + X_{ij}\theta + C_{ij} + e_{ijk}, \qquad (1)$$

$$\mathbf{C}_i = (C_{i1}, C_{i2}, \ldots, C_{iT})' \sim N(0, \mathbf{W}), \quad e_{ijk} \sim N(0, \sigma_e^2),$$

where $\boldsymbol{\beta}$ is a p-dimensional column vector of fixed time effects and $\mathbf{Z}_j$ is a p-dimensional row vector specifying the form for the fixed effects associated with period j, θ is the treatment effect (this is the term we are interested in estimating), $\mathbf{C}_i$ is a T-dimensional vector of cluster-period random effects for cluster i in period j with $T \times T$ covariance matrix $\mathbf{W}$, assumed to be constant across all clusters, and assumed to be independent of e_{ijk}, the subject-specific random error. Let $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_T)'$ be the $T \times p$ design matrix for the time effects for a cluster, assumed common across all clusters. Note that $\mathbf{W}$ and σ_e^2 represent (co)variation around the fixed effects $\mathbf{Z}$ and $\mathbf{X}_i$ in the model, however, we suppress this dependence in our notation. We return to this point in Section 3.1.

The effect of time may be modeled differently depending on the chosen forms for $\mathbf{Z}_j$ and $\boldsymbol{\beta}$. One common choice is to assume separate effects for each time period, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)'$ and $\mathbf{Z}_j$ a T-dimensional row vector consisting of 1 in the jth position and 0s elsewhere so that $\mathbf{Z}_j\boldsymbol{\beta} = \beta_j$. Another possibility is a linear effect of time over the trial periods, with $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ and $\mathbf{Z}_j = (1, j)$ a two-dimensional row vector. As these parameterizations depend on the trial periods $j = 1, \ldots, T$, the dimension of $\mathbf{Z}_j$ will be less than or equal to T.

Several different forms for the within-cluster covariance matrix $\mathbf{W}$ could be considered: one in which the correlation between any pair of subjects' outcomes in a cluster is the same (Hussey and Hughes 2007) or one in which the correlation

depends on subjects' periods of measurement. Those depending on subjects' periods of measurement assume that the correlation between pairs of subjects' outcomes decays in some manner: either subjects' outcomes measured in different periods are assumed to be less highly correlated than those measured in the same period (Hooper et al. 2016), or the correlation decays as the time between subjects' periods of measurement increases (Kasza et al. 2019). All of these models treat time as discrete and so can be collapsed to the cluster-period mean level without loss of information, with cluster-period means $\bar{Y}_{ij} = \frac{1}{m}\sum_{k=1}^{m} Y_{ijk}$ (Hooper et al. 2016; Kasza et al. 2019).

Model (1) can be represented at the cluster-period mean level in matrix form as

$$\begin{pmatrix} \bar{\mathbf{Y}}_1 \\ \vdots \\ \bar{\mathbf{Y}}_N \end{pmatrix} = \begin{pmatrix} \mathbf{Z} & \mathbf{X}_1 \\ \vdots & \vdots \\ \mathbf{Z} & \mathbf{X}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \theta \end{pmatrix} + \begin{pmatrix} \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_N \end{pmatrix} + \begin{pmatrix} \bar{\mathbf{e}}_1 \\ \vdots \\ \bar{\mathbf{e}}_N \end{pmatrix},$$

where $\bar{\mathbf{Y}}_i = (\bar{Y}_{i1}, \bar{Y}_{i2}, \ldots, \bar{Y}_{iT})'$ is the T-dimensional vector of mean outcomes for subjects measured in cluster i in period j and $\bar{\mathbf{e}}_i = (\bar{e}_{i1}, \bar{e}_{i2}, \ldots, \bar{e}_{iT})'$ is the T-dimensional vector of mean random errors for cluster i in period j. Since we assume $\text{cov}(\mathbf{C}_i) = \text{cov}(\mathbf{C}_{i'}) = \mathbf{W}$ for all i, i' and a constant cluster-period size m, the covariance matrix for a cluster $\text{cov}(\bar{\mathbf{Y}}_i) = \mathbf{V} = \mathbf{W} + \frac{\sigma_e^2}{m}\mathbf{I}_T$ for all i, where $\mathbf{I}_T$ is the $T \times T$ identity matrix.

The variance of the treatment effect estimator at the cluster-period mean level obtained from generalized least squares (Kasza et al. 2019) is given by

$$\text{var}(\hat{\theta}) = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_i - \frac{1}{N} \left(\sum_{i=1}^{N} \mathbf{X}_i \right)' \mathbf{V}^{-1} \mathbf{Z} \left(\mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z} \right)^{-1} \mathbf{Z}' \mathbf{V}^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i \right) \right)^{-1}. \quad (2)$$

Note that the design matrix for the time effects, $\mathbf{Z}$, appears only in the second component of the above expression. Also note that sample size calculations at the planning stage of a trial are fundamentally based on this variance expression.

2.3. Cohort Designs

A model for a cohort design could include an additional random effect term to account for the repeated measures on each subject, such as s_{ik} where $s_{ik} \sim N(0, \sigma_s^2)$ (Hooper et al. 2016). The covariance matrix for a cluster under a cohort design, $\mathbf{V}_c$, would then have the form $\mathbf{V}_c = \mathbf{V} + \frac{\sigma_s^2}{m}\mathbf{J}$ where $\mathbf{V}$ is the covariance matrix from Section 2.2 and $\mathbf{J}$ is a $T \times T$ matrix of 1s. Note that the theorem we introduce in Section 2.5 also holds for a cohort design and the proof follows, with $\mathbf{V}$ replaced by $\mathbf{V}_c$.

2.4. Time Parameterizations

Many different time parameterizations may be plausible for a cluster randomized trial. One requirement is that the time parameterization does not constrain any of the cluster-period cells to zero; that is, it must be possible to obtain the mean outcome for the control group from the chosen parameterization, such as with separate period effects or an overall intercept term with effects for all but one period. As discrete time parameterizations involve some formulation of the T periods of the trial, the time parameterization matrix $\mathbf{Z}$ will have maximum dimension T. The time parameterization matrix specifying categorical fixed effects for each period may take the form $\mathbf{Z}_{\text{cat}} = \mathbf{I}_T$, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)'$. Note that this could be equivalently parameterized with an overall intercept term μ so that $\boldsymbol{\beta} = (\mu, \beta_2, \ldots, \beta_T)'$ (with $\beta_1 = 0$ for identifiability) and

$$\mathbf{Z}_{\text{cat}} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 1 \end{pmatrix}.$$

Alternatively, the time parameterization specifying a linear term for time across periods would take the form

$$\mathbf{Z}_{\text{lin}} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{pmatrix}.$$

Other parameterizations could also be used such as splines, fractional polynomials, or higher order polynomials, as long as a linear term is also included.

2.5. Effect of Time Parameterization on the Variance of the Treatment Effect Estimator

The following theorem gives a sufficient condition for when the variance of the treatment effect estimator is invariant to the form of the time parameterization. Note that this theorem involves the column space of a matrix (Strang 2006, p. 71). (A vector $\mathbf{b}$ is in the column space of a matrix $\mathbf{A}$ if there exists a vector $\mathbf{v}$ for which $\mathbf{Av} = \mathbf{b}$, that is, if there is a linear combination of the columns of $\mathbf{A}$ which yields $\mathbf{b}$.)

Theorem 1. Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{iT})'$ denote the treatment sequence for cluster i and $\mathbf{Z}$ denote the design matrix for the effect of time, assumed to be identical across clusters. If $\sum_{i=1}^{N} \mathbf{X}_i = (\sum_{i=1}^{N} X_{i1}, \ldots, \sum_{i=1}^{N} X_{iT})'$ is in the column space of $\mathbf{Z}$ then $\text{var}(\hat{\theta})$ is invariant to the choice of $\mathbf{Z}$.

Proof. Suppose $\sum_{i=1}^{N} \mathbf{X}_i$ is in the column space of $\mathbf{Z}$. Let $\mathbf{P} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{V}^{-1} = \mathbf{ZQ}$. Note that $\mathbf{Q}$ is a weak inverse of $\mathbf{Z}$ since $\mathbf{ZQZ} = \mathbf{Z}$ and hence $\mathbf{P}$ is an oblique projection onto the column space of $\mathbf{Z}$ (Seber 2008, p. 126). Then $\mathbf{P}\left(\sum_{i=1}^{N} \mathbf{X}_i\right) = \left(\sum_{i=1}^{N} \mathbf{X}_i\right)$ and so (2) becomes

$$\text{var}(\hat{\theta}) = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_i - \frac{1}{N} \left(\sum_{i=1}^{N} \mathbf{X}_i \right)' \mathbf{V}^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i \right) \right)^{-1}.$$

Therefore, $\text{var}(\hat{\theta})$ is invariant to the choice of $\mathbf{Z}$. □

Hence, for any choice of time parameterization $\mathbf{Z}$ for which the sum of the treatment sequences across clusters $\sum_{i=1}^{N} \mathbf{X}_i$ is

in the column space of $\mathbf{Z}$, the variance of the treatment effect estimator $\text{var}(\hat{\theta})$ will remain the same, all else being equal. In the next section, we will outline the implications for common trial designs with an arbitrary number of clusters and periods. To establish whether a sum of the treatment sequences vector is in the column space of a time parameterization for a particular trial configuration in practice, we refer the reader to the Appendix where we provide an additional useful result and example code to perform this check.

3. Implications for Cluster Randomized Trial Planning

3.1. Effect of Time Parameterization on Trial Planning

We have given a sufficient condition for when the time parameterization does not alter the variance of the treatment effect estimator. However, the time parameterization still affects trial planning through the estimates of the variance components we assume: the estimates of the variance components obtained from a model are conditional on the parameterization of time in that model. As a result, the appropriateness of the model specification used affects whether the variance components are correctly estimated. For example, if the effect of time were underspecified, such as choosing a linear term for time when a more appropriate specification would have also included a quadratic term, then the variance components may be overestimated since the time components in the model would not have adequately accounted for the variation in the outcome attributable to time. When obtaining estimates of the variance components from previously published studies or by fitting a model to administrative data, it is important that the time parameterization adequately captures temporal effects.

3.2. Implications for Specific Designs

3.2.1. Parallel Design

A parallel design typically has half of the clusters implement the treatment sequence $(1, \ldots, 1)'$ and the other half of the clusters implement the sequence $(0, \ldots, 0)'$. The sum of these sequences across clusters is $\sum_{i=1}^{N} \mathbf{X}_i = \frac{N}{2}(1, \ldots, 1)'$. This vector is in the column space of any time parameterization with an intercept term (or any equivalent representation): a parameterization with an intercept term will have a column of 1s and so the vector $\sum_{i=1}^{N} \mathbf{X}_i$ can be easily obtained by multiplying this column by the constant term $\frac{N}{2}$ and the remaining columns by 0. Parameterizations with an intercept term include $\mathbf{Z}_{\text{cat}}$ and $\mathbf{Z}_{\text{lin}}$ in Section 2.4 and so by Theorem 1 all such time parameterizations yield the same $\text{var}(\hat{\theta})$. Hence, sample size expressions do not involve the time parameterization in $\mathbf{Z}$.

3.2.2. Parallel With Baseline Design

The parallel with baseline design has half of the clusters implement the control in the first period followed by the intervention, and the other half implement the control for the entire trial, with treatment sequences $(0, 1, \ldots, 1)'$ and $(0, \ldots, 0)'$, respectively. The sum of the treatment sequences across clusters is therefore $\sum_{i=1}^{N} \mathbf{X}_i = \frac{N}{2}(0, 1, \ldots, 1)'$. Due to the presence of the 0 element in this vector, it is not necessarily in the column space of all time parameterizations with an intercept term unlike the parallel design. This vector is in the column space of the time parameterization with fixed period effects, $\mathbf{Z}_{\text{cat}}$, since $\mathbf{Z}_{\text{cat}}\mathbf{u} = \sum_{i=1}^{N} \mathbf{X}_i$ for $\mathbf{u} = \frac{N}{2}(0, 1, \ldots, 1)'$. However, it is not in the column space of the time parameterization with a linear term for time, $\mathbf{Z}_{\text{lin}}$, since there does not exist a vector $\mathbf{v}$ for which $\mathbf{Z}_{\text{lin}}\mathbf{v} = \sum_{i=1}^{N} \mathbf{X}_i$. Therefore, the choice of time parameterization for such a design may affect $\text{var}(\hat{\theta})$, impacting sample size calculations.

3.2.3. Cluster Randomized Crossover Design

The balanced cluster randomized crossover design with treatment sequences that alternate with each period has half of the clusters implement the treatment sequence $(0, 1, \ldots, 0, 1)'$ and the other half implement the sequence $(1, 0, \ldots, 1, 0)'$. Then $\sum_{i=1}^{N} \mathbf{X}_i = \frac{N}{2}(1, \ldots, 1)'$. As with the parallel design, this vector is in the column space of any time parameterization with an intercept term (or equivalent representation) including $\mathbf{Z}_{\text{cat}}$ and $\mathbf{Z}_{\text{lin}}$ and so any such time parameterization yields the same $\text{var}(\hat{\theta})$.

3.2.4. Stepped Wedge Design

The classic stepped wedge design with an equal allocation of clusters to each treatment sequence has sequences ranging from $(0, \ldots, 0, 1)'$ to $(0, 1, \ldots, 1)'$ and so the sum of the treatment sequences vector $\sum_{i=1}^{N} \mathbf{X}_i = \frac{N}{T-1}(0, 1, \ldots, T-1)'$. This vector is in the column space of both $\mathbf{Z}_{\text{cat}}$ and $\mathbf{Z}_{\text{lin}}$: $\mathbf{Z}_{\text{cat}}\mathbf{u} = \sum_{i=1}^{N} \mathbf{X}_i$ for $\mathbf{u} = \frac{N}{T-1}(0, 1, \ldots, T-1)'$ and $\mathbf{Z}_{\text{lin}}\mathbf{v} = \sum_{i=1}^{N} \mathbf{X}_i$ for $\mathbf{v} = \frac{N}{T-1}(-1, 1)'$. Therefore, either parameterization yields the same $\text{var}(\hat{\theta})$. Note, however, that the sum of the treatment sequences here has a more complex form than for the parallel and cluster randomized crossover designs: not all time parameterizations with an intercept term would necessarily be in the column space of this vector.

4. Continuous-Time Framework

4.1. Continuous-Time Model and Variance Expression

We may also consider a model that allows use of continuous time parameterizations or that can accommodate a within-cluster correlation structure where the correlation between a pair of subjects' outcomes in a cluster depends on subjects' measurement times, for instance decaying as the distance between measurement times increases (Grantham et al. 2019). This model can be represented at the individual level by the following

$$Y_{ijk} = \mathbf{Z}_{jk}\boldsymbol{\beta} + X_{ij}\theta + C_{ijk} + e_{ijk},\ \mathbf{C}_i \sim N(0, \mathbf{S}),\ e_{ijk} \sim N(0, \sigma_e^2), \tag{3}$$

where $\mathbf{Z}_{jk}$ is a p-dimensional row vector specifying the form for the time effect associated with subject k in period j and $\mathbf{C}_i = (C_{i11}, \ldots, C_{i1m}, \ldots, C_{iT1}, \ldots, C_{iTm})'$ is a Tm-dimensional vector of random effects for cluster i with $Tm \times Tm$ covariance matrix $\mathbf{S}$. All other terms have the same interpretation as in model (1).

Model (3) can be represented at the individual level in matrix form by

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{pmatrix} = \begin{pmatrix} \mathbf{Z}^* & \mathbf{X}_1^* \\ \vdots & \vdots \\ \mathbf{Z}^* & \mathbf{X}_N^* \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \theta \end{pmatrix} + \begin{pmatrix} \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_N \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_N \end{pmatrix},$$

where $\mathbf{Y}_i = (Y_{i11}, \ldots, Y_{i1m}, \ldots, Y_{iT1}, \ldots, Y_{iTm})'$ is the Tm-dimensional vector of outcomes for subjects in cluster i, $\mathbf{Z}^*$ is the $Tm \times p$ design matrix for the time effects for a cluster, $\mathbf{X}_i^* = \mathbf{X}_i \otimes \mathbf{1}_m$ is the expanded treatment sequence for cluster i where $\mathbf{1}_m$ is an m-dimensional column vector of 1s, and $\mathbf{e}_i$ is the Tm-dimensional vector of random errors for subjects' outcomes in cluster i. Note that $\otimes$ denotes the Kronecker product so that $\mathbf{X}_i \otimes \mathbf{1}_m$ yields a Tm-dimensional vector with each element of $\mathbf{X}_i$ repeated m times. We again assume $\text{cov}(\mathbf{C}_i) = \text{cov}(\mathbf{C}_{i'})$ for all i, i', and so the $Tm \times Tm$ covariance matrix for a cluster $\text{cov}(\mathbf{Y}_i) = \mathbf{V}$ for all i.

Under this model, the vector of cluster-period means is not a sufficient statistic for the treatment effect, and so the expression for the variance of the treatment effect estimator must be at the individual subject level (Grantham et al. 2019), with

$$\text{var}(\hat{\theta}) = \left(\sum_{i=1}^{N} \mathbf{X}_i^{*'} \mathbf{V}^{-1} \mathbf{X}_i^* - \frac{1}{N} \left(\sum_{i=1}^{N} \mathbf{X}_i \otimes \mathbf{1}_m \right)' \right.$$
$$\left. \mathbf{V}^{-1} \mathbf{Z}^* \left(\mathbf{Z}^{*'} \mathbf{V}^{-1} \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*'} \mathbf{V}^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i \otimes \mathbf{1}_m \right) \right)^{-1}.$$

Adapting Theorem 1 to the individual level simply requires replacing $\mathbf{Z}$ with $\mathbf{Z}^*$ and $\sum_{i=1}^{N} \mathbf{X}_i$ with $\sum_{i=1}^{N} \mathbf{X}_i \otimes \mathbf{1}_m$. The proof follows similarly. With such a model, we could still consider discrete time parameterizations, simply represented at the individual level, such as $\mathbf{Z}_{\text{cat}} \otimes \mathbf{1}_m$ or $\mathbf{Z}_{\text{lin}} \otimes \mathbf{1}_m$, or a more detailed time parameterization such as a continuous linear effect of time, with

$$\mathbf{Z}_{\text{lin}}^* = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_{Tm} \end{pmatrix},$$

where t_1 denotes the first subject's measurement time in a cluster and t_{Tm} denotes the last. Note that if $\sum_{i=1}^{N} \mathbf{X}_i$ is in the column space of a discrete time parameterization $\mathbf{Z}$, then we also have that $\sum_{i=1}^{N} \mathbf{X}_i \otimes \mathbf{1}_m$ is in the column space of $\mathbf{Z} \otimes \mathbf{1}_m$. Model (3) and the associated individual-level variance expression are simply more general than those in Section 2.2 and so we will see in the following sections that if discrete time parameterizations were to be considered here then the design-specific results in Section 3 will still hold.

4.2. Implications for Specific Designs

4.2.1. Parallel and Cluster Randomized Crossover Designs

Recall that $\sum_{i=1}^{N} \mathbf{X}_i = \frac{N}{2}(1, \ldots, 1)'$ for both parallel and cluster randomized crossover designs, and so the sum of the treatment sequences at the individual level is simply the Tm-dimensional vector $\sum_{i=1}^{N} \mathbf{X}_i \otimes \mathbf{1}_m = \frac{N}{2}(1, \ldots, 1, \ldots, 1, \ldots, 1)'$. Due to its simple form, this vector is again in the column space of any time parameterization with an intercept term including $\mathbf{Z}_{\text{cat}} \otimes \mathbf{1}_m$, $\mathbf{Z}_{\text{lin}} \otimes \mathbf{1}_m$, and $\mathbf{Z}_{\text{lin}}^*$ (since $\mathbf{Z}_{\text{lin}}^* \mathbf{v} = \frac{N}{2}(1, \ldots, 1, \ldots, 1, \ldots, 1)'$ for $\mathbf{v} = \frac{N}{2}(1, 0)'$), and so any such time parameterization yields the same $\text{var}(\hat{\theta})$.

4.2.2. Parallel With Baseline Design

The sum of the treatment sequences at the individual level for a parallel with baseline design is $\sum_{i=1}^{N} \mathbf{X}_i \otimes \mathbf{1}_m = \frac{N}{2}(0, \ldots, 0, 1, \ldots, 1, \ldots, 1, \ldots, 1)'$. As before, the common time parameterizations do not all yield the same $\text{var}(\hat{\theta})$ expression: $\sum_{i=1}^{N} \mathbf{X}_i \otimes \mathbf{1}_m$ is in the column space of $\mathbf{Z}_{\text{cat}} \otimes \mathbf{1}_m$ but not in the column space of $\mathbf{Z}_{\text{lin}} \otimes \mathbf{1}_m$ or $\mathbf{Z}_{\text{lin}}^*$.

4.2.3. Stepped Wedge Design

For the stepped wedge design, the sum of the treatment sequences at the individual level is $\sum_{i=1}^{N} \mathbf{X}_i \otimes \mathbf{1}_m = \frac{N}{T-1}(0, \ldots, 0, 1, \ldots, 1, \ldots, T-1, \ldots, T-1)'$. This vector is in the column space of $\mathbf{Z}_{\text{cat}} \otimes \mathbf{1}_m$ and $\mathbf{Z}_{\text{lin}} \otimes \mathbf{1}_m$, but not $\mathbf{Z}_{\text{lin}}^*$: there does not exist a vector $\mathbf{u}$ for which $\mathbf{Z}_{\text{lin}}^* \mathbf{u} = \frac{N}{T-1}(0, \ldots, 0, 1, \ldots, 1, \ldots, T-1, \ldots, T-1)'$ (except for the situation where $\mathbf{Z}_{\text{lin}}^* = \mathbf{Z}_{\text{lin}} \otimes \mathbf{1}_m$). Therefore, the choice of time parameterization under an individual-level model for a stepped wedge design may affect $\text{var}(\hat{\theta})$.

5. Concluding Remarks

In this article, we have shown that the variance of the treatment effect estimator remains the same when certain time parameterizations are chosen in the cluster randomized trial models we describe. The time parameterizations we consider are among the standard time effect specifications in multiple-period cluster randomized trials. Since sample size calculations take this variance as an input, this information can be used to simplify trial planning.

The theorem we present under a discrete-time framework concerns the vector $\sum_{i=1}^{N} \mathbf{X}_i$, obtained by summing across all N clusters' treatment sequences for a particular trial design, and the design matrix for the effect of time, $\mathbf{Z}$, which we refer to as a time parameterization. We show that the variance of the treatment effect estimator expression reduces to a form that does not involve the time parameterization matrix $\mathbf{Z}$ if the sum of the treatment sequences vector $\sum_{i=1}^{N} \mathbf{X}_i$ can be written as a linear combination of the columns of $\mathbf{Z}$; that is, if $\sum_{i=1}^{N} \mathbf{X}_i$ is in the column space of $\mathbf{Z}$. Parallel and cluster randomized crossover designs such as those depicted in Figure 1 yield a sum of the treatment sequences vector with a simple form, namely a constant times a vector of 1s: $\sum_{i=1}^{N} \mathbf{X}_i = \frac{N}{2}(1, \ldots, 1)'$. As this vector is in the column space of any time parameterization with an intercept term, then by Theorem 1, all such time parameterizations yield the same variance expression for both discrete- and continuous-time model formulations. We cannot make as general a conclusion about time parameterizations for the stepped wedge and parallel with baseline designs, however, as the corresponding vectors $\sum_{i=1}^{N} \mathbf{X}_i$ have a more complex form. For the two specific discrete time parameterizations we consider, one with categorical period effects and the other with a linear trend,

both parameterizations yield the same variance expression for the stepped wedge design, but they do not yield the same variance expression for the parallel with baseline design.

While the reduced variance expression does not contain $\mathbf{Z}$, it still contains the matrix $\mathbf{V}$, the covariance matrix for a cluster. We note in Section 3.1 that if we were in the fortunate position at the trial planning stage of being able to estimate the variance components that make up $\mathbf{V}$ using data related to the trial setting, then the chosen time parameterization in the model we fit to these data would still affect the estimated parameter values and the resulting variance. However, it will often be the case that the variance components will be assumed based on similar past studies, in which case once these values are chosen, $\mathbf{V}$ would remain fixed. Then for any time parameterizations we may consider that satisfy the column space condition, the evaluated variance of the treatment effect estimator values would be the same, yielding identical sample size calculations.

We have made the assumption that the trial data will be sufficient to estimate all model parameters at the analysis stage, yet this may not always be the case. Some time parameterizations we might consider could have a large number of parameters to estimate, for example, categorical period effects for a trial with a large number of periods. Estimating a large number of time effects could be problematic for trials with a modest sample size, such as those with a small number of clusters or small cluster sizes. In these situations, time parameterizations with a smaller number of parameters would be preferred to ensure that all parameters in the model can be reliably estimated.

Appendix A: Checking Column Space Condition in Practice

To determine whether the sum of the treatment sequences vector $\sum_{i=1}^{N} \mathbf{X}_i$ is in the column space of a time parameterization matrix $\mathbf{Z}$ for a particular trial configuration in practice, readers can make use of the following result (Strang 2006, p. 163): $\sum_{i=1}^{N} \mathbf{X}_i$ is in the column space of $\mathbf{Z}$ if and only if $\mathbf{P}_Z \left(\sum_{i=1}^{N} \mathbf{X}_i\right) = \left(\sum_{i=1}^{N} \mathbf{X}_i\right)$, where $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

Then for a given vector $\sum_{i=1}^{N} \mathbf{X}_i$ and matrix $\mathbf{Z}$, simply construct the matrix $\mathbf{P}_Z$ and post-multiply $\mathbf{P}_Z$ by $\sum_{i=1}^{N} \mathbf{X}_i$.

Appendix B: Example R Code to Check Column Space Condition

```
## Trial configuration: Stepped wedge design,
##  30 clusters, 4 periods
## Two different time parameterizations could
##  be appropriate:
##    (1) Categorical period effects or
##    (2) Linear trend
## Setup
# Define sum of the treatment sequences vector
x <- 10*matrix(seq(0,3), 4, 1)
# Define time parameterization option 1:
#   categorical period effects
Z1 <- diag(4)
# Define time parameterization option 2:
#   linear trend
Z2 <- cbind(rep(1,4), seq(1,4))

## Check time parameterization Z1
# Step 1. Construct projection matrix, P_Z1
P_Z1 <- Z1 %*% solve(t(Z1) %*% Z1) %*% t(Z1)
# Step 2. Calculate P_Z1 x
P_Z1 %*% x
# P_Z1 x = x, so x is in the column space of Z1

## Check time parameterization Z2
# Step 1. Construct projection matrix, P_Z2
P_Z2 <- Z2 %*% solve(t(Z2) %*% Z2) %*% t(Z2)
# Step 2. Calculate P_Z2 x
P_Z2 %*% x
# P_Z2 x = x, so x is also in the column space
#    of Z2
```

Acknowledgments

The authors wish to thank the editor, associate editor, and two anonymous reviewers for their helpful comments and suggestions which have improved the clarity of the article.

Funding

The authors gratefully acknowledge funding by the National Health and Medical Research Council of Australia Project Grant ID 1108283 and an Australian Government Research Training Program (RTP) Scholarship.

References

Copas, A. J., Lewis, J. J., Thompson, J. A., Davey, C., Baio, G., and Hargreaves, J. R. (2015), "Designing a Stepped Wedge Trial: Three Main Designs, Carry-Over Effects and Randomisation Approaches," *Trials*, 16, 352. [184]

Grantham, K. L., Kasza, J., Heritier, S., Hemming, K., and Forbes, A. B. (2019), "Accounting for a Decaying Correlation Structure in Cluster Randomized Trials With Continuous Recruitment," *Statistics in Medicine*, 38, 1918–1934. [184,187,188]

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., and Lilford, R. J. (2015), "The Stepped Wedge Cluster Randomised Trial: Rationale, Design, Analysis, and Reporting," *BMJ*, 350, h391. [184]

Hooper, R., Teerenstra, S., de Hoop, E., and Eldridge, S. (2016), "Sample Size Calculation for Stepped Wedge and Other Longitudinal Cluster Randomised Trials," *Statistics in Medicine*, 35, 4718–4728. [184,186]

Hussey, M. A., and Hughes, J. P. (2007), "Design and Analysis of Stepped Wedge Cluster Randomized Trials," *Contemporary Clinical Trials*, 28, 182–191. [185]

Kasza, J., Hemming, K., Hooper, R., Matthews, J. N. S., and Forbes, A. B. (2019), "Impact of Non-uniform Correlation Structure on Sample Size and Power in Multiple-Period Cluster Randomised Trials," *Statistical Methods in Medical Research*, 28, 703–716. [186]

Maxwell, A. E., Parker, R. A., Drever, J., Rudd, A., Dennis, M. S., Weir, C. J., and Al-Shahi Salman, R. (2017), "Promoting Recruitment Using Information Management Efficiently (PRIME): A Stepped-Wedge, Cluster Randomised Trial of a Complex Recruitment Intervention Embedded Within the REstart or Stop Antithrombotics Randomised Trial," *Trials*, 18, 623. [184]

Seber, G. A. F. (2008), *A Matrix Handbook for Statisticians*, Hoboken, NJ: Wiley. [186]

Solomon, E., Rees, T., Ukoumunne, O. C., Metcalf, B., and Hillsdon, M. (2014), "The Devon Active Villages Evaluation (DAVE) Trial of a Community-Level Physical Activity Intervention in Rural South-West England: A Stepped Wedge Cluster Randomised Controlled Trial," *International Journal of Behavioral Nutrition and Physical Activity*, 11, 94. [184]

Strang, G. (2006), *Linear Algebra and Its Applications* (4th ed.), Belmont, CA: Thomson Brooks/Cole. [186,189]