CRTFASTGEEPWR: A **SAS** Macro for Power of Generalized Estimating Equations Analysis of Multi-Period Cluster Randomized Trials with Application to Stepped Wedge Designs

Ying Zhang ⦿
The University of North Carolina
at Chapel Hill

John S. Preisser
The University of North Carolina
at Chapel Hill

Fan Li ⦿
Yale School of Public Health

Elizabeth L. Turner ⦿
Duke University

Paul J. Rathouz ⦿
University of Texas at Austin

Abstract

Multi-period cluster randomized trials (CRTs) are increasingly used for the evaluation of interventions delivered at the group level. While generalized estimating equations (GEE) are commonly used to provide population-averaged inference in CRTs, there is a gap of general methods and statistical software tools for power calculation based on multi-parameter, within-cluster correlation structures suitable for multi-period CRTs that can accommodate both complete and incomplete designs. A computationally fast, non-simulation procedure for determining statistical power is described for the GEE analysis of complete and incomplete multi-period cluster randomized trials. The procedure is implemented via a SAS macro, **CRTFASTGEEPWR**, which is applicable to binary, count and continuous responses and several correlation structures in multi-period CRTs. The SAS macro is illustrated in the power calculation of two complete and two incomplete stepped wedge cluster randomized trial scenarios under different specifications of marginal mean model and within-cluster correlation structure. The proposed GEE power method is quite general as demonstrated in the SAS macro with numerous input options. The power procedure and macro can also be used in the planning of parallel and crossover CRTs in addition to cross-sectional and closed cohort stepped wedge trials.

Keywords: correlation decay, group randomized trial, intraclass correlation, incremental intervention effects, marginal models.

1. Introduction

Cluster randomized trials (CRTs) are studies designed to evaluate interventions that operate at a group level, manipulate the physical or social environment, or cannot be delivered to individuals (Murray, Varnell, and Blitstein 2004). Regarding the different schedules of recruiting participants, CRTs are classified into cross-sectional, closed-cohort and open-cohort designs (Copas, Lewis, Thompson, Davey, Baio, and Hargreaves 2015). Cross-sectional designs recruit a unique set of individuals in each period, whereas closed-cohort designs follow the same individuals in clusters with repeated observations across periods. The open-cohort design, however, allows the addition of new members to an existing cohort in each period. Moreover, there are different types of CRT designs, including parallel, crossover, and stepped-wedge designs. As are instances of CRTs with outcomes measured in multiple periods, a stepped-wedge cluster randomized trial (SW-CRT) is a type of CRT wherein clusters switch from control condition to treatment at randomly assigned time points (Hussey and Hughes 2007). Logistical and ethical considerations such as the need to deliver the intervention in stages and the desire to implement the intervention in all clusters are factors involved in the choice to use a SW-CRT (Turner, Li, Gallis, Prague, and Murray 2017). SW-CRTs may be preferred over other designs because they may facilitate cluster recruitment or offer increased power over other cluster randomized designs even with a limited number of clusters (Hemming and Taljaard 2020). Most study planning methods are for complete SW-CRTs, where all clusters have outcome data in all periods. However, incomplete stepped wedge designs are increasingly being deployed, whereby some cluster-periods do not record data due to logistical, resource, and patient-centered considerations (Kasza and Forbes 2019). Specifically, researchers may choose not to collect data in a cohort design or enroll new participants in a cross-sectional design during some cluster-periods. Hemming, Lilford, and Girling (2015) described two types of incompleteness in stepped wedge designs, one involving implementation periods and the other staggered study entry or termination of clusters.

Population-averaged analysis models with GEE estimation and inference have several advantages for the design and analysis of CRTs (Preisser, Young, Zaccaro, and Wolfson 2003). In contrast to generalized linear mixed models, the intervention effect from a population-averaged model describes how the average response changes across the subsets of population defined by the treated and control cluster-periods. Additionally, because models for mean and correlation structures are separately specified, the interpretation of the marginal mean regression parameters remains the same regardless of the specification of working correlation model (Preisser, Lu, and Qaqish 2008). The link function is chosen to obtain inference on the target parameter of choice; for binary responses, the target parameters could be the odds ratio via the logit link, the risk ratio via the log link, or the risk difference via the identify link. Another advantage in using GEE for CRTs is that the estimation of mean model parameters is robust to misspecification of correlation structures in large samples. However, the specification of working independence correlation structure may result in efficiency loss that can be substantial when the cluster-period sizes are not all equal (Tian, Preisser, Esserman, Turner, Rathouz, and Li 2022). Furthermore, an over-simplified exchangeable correlation structure may give inaccurate power calculations when there is correlation decay in multi-period CRTs (Li 2020; Kasza, Hemming, Hooper, Matthews, and Forbes 2019). Thus, correlation structures informed by the study design are recommended for both study design and data analysis of stepped wedge and other multi-period CRTs.

Because CRTs are usually less powerful than individually randomized trials, determination of the proper number and allocation of study participants is critically important. A general, non-simulation, computationally fast power method based on the model-based variance was introduced by Rochon (1998). Moreover, as special cases of the fast GEE power method, simple-to-use sample size formulae for continuous responses and non-simulation procedures for binary responses have recently been proposed for complete, cross-sectional and cohort SW-CRTs (Li, Turner, and Preisser 2018). The methods extend earlier sample size formulae for GEE analysis of parallel-groups CRTs, including cross-sectional and cohort CRTs (Preisser *et al.* 2003, 2008) and multi-level CRTs (Reboussin, Preisser, Song, and Wolfson 2012; Teerenstra, Lu, Preisser, Van Achterberg, and Borm 2010; Wang, Turner, Preisser, and Li 2021). Prior work (Li *et al.* 2018; Li 2020) has shown that the analytical fast GEE power for marginal mean (e.g, intervention) parameters in complete SW-CRTs agrees well with simulated power based on GEE with finite-sample sandwich variance estimators for as few as eight clusters. Zhang, Preisser, Turner, Rathouz, Toles, and Li (2023b) shows, for CRTs with incomplete stepped wedge designs, fast GEE power based on the t-test agrees well with simulated power with the finite-sample sandwich variance estimator of (Kauermann and Carroll 2001) and twelve clusters. In scenarios with 6 clusters, it is similar to simulated power with estimations using the correctly specified model-based variance estimator. Those studies focus on the empirical performance of the analytical power rather than software tools to implement the method for different multi-period CRT designs.

This article describes a comprehensive, analytical power method and accompanying SAS (SAS Institute Inc. 2020) macro for both complete and incomplete multi-period CRTs. The proposed GEE power procedure is motivated by the Connect-Home trial, which uses an incomplete, cross-sectional stepped wedge design to test an intervention to improve outcomes for rehabilitation patients transitioning from skilled nursing facilities (SNFs) to home-based care (Toles *et al.* 2021). The primary component of the intervention is an individualized Transition Plan of Care that SNF staff create to support the patient and caregiver at home. The incomplete design with six SNFs (clusters) and four patients per cluster-period (360 patients total), shown in Figure 1, was chosen based on considerations of internal validity and power under restrictions placed by available resources and logistical considerations. The black and orange boxes represent cluster-periods where no patients are enrolled, giving an incomplete design. Staggered enrollment of SNFs (clusters) is used to initiate data collection in stages with limited research staff resulting in the black boxes. The orange boxes represent the implementation phase, where two months are needed to activate the intensive intervention through training nursing home and home health care staff. The Connect-Home study design is distinct in several aspects. First, the number of periods ($J = 22$) is much greater than the number of sequences ($S = 6$), as compared to the standard SW-CRT in which $J = S + 1$ Hussey and Hughes (2007); Hemming *et al.* (2015); Li *et al.* (2018). Next, the incompleteness of the design adds to the complexity of power calculation.

Thus, the Connect-Home trial inspires an extension of the computationally fast, non-simulation procedures for determining sample size and statistical power for GEE analysis from complete (Li *et al.* 2018) to incomplete SW-CRTs. Specifically, we introduce a SAS macro **CRTFASTGEEPWR** as a computationally efficient, non-simulation based routine for determining the statistical power in multi-period CRTs by further accommodating incompleteness at the design stage. Ouyang, Li, Preisser, and Taljaard (2022) reviewed 18 sample size calculators for stepped wedge CRTs and identified the SAS macro **CRTFASTGEEPWR**

Figure 1: The trial diagram of the Connect Home trial: the blue, orange and green cells denote control, implementation and intervention cluster-periods, respectively.

and the R (R Core Team 2023) package **swdpwr** (Chen, Zhou, Li, and Spiegelman 2022) as two of the most flexible approaches for non-simulation GEE power and sample size calculation for marginal models. Whereas **swdpwr** provides power calculation based on the average intervention effects mean model with categorical period effects for continuous and binary outcomes, exchangeable-type correlation structures, and equal cluster sizes for complete designs, the SAS macro **CRTFASTGEEPWR** additionally offers three options for coding intervention effects, a choice of linear effects or categorical period effects, count outcomes, time-decaying correlation structures, and unequal cluster-period sizes for both complete and incomplete multi-period CRT designs. The SAS macro **CRTFASTGEEPWR** is developed to accommodate binary, count and continuous responses in stepped wedge and other multi-period CRTs with a collection of commonly-seen multilevel intra-cluster correlation structures for incomplete cross-sectional and cohort designs.

The remainder of this article is organized as follows. Section 2 describes the population-averaged models of interest with special consideration of correlation structures suitable for multi-period CRTs. Section 3 summarizes the general power procedure for GEE analysis with adaptation to stepped wedge designs. Section 4 presents the SAS macro details and four examples for complete and incomplete SW-CRTs.

2. GEE analysis of multi-period CRTs

2.1. Population-averaged model of multi-period CRTs

A unifying population-averaged model framework is described below for the design and statistical analysis of multi-period CRTs. The following notations apply to both cross-sectional and cohort multi-period CRTs, where there are J periods, S sequences, I clusters and I_s clusters in sequence s, such that $I = \sum_{s=1}^{S} I_s$. Let y_{ijk} denote the response of the kth individual from cluster i during period j for $i = 1, \ldots, I, j = 1, \ldots, J_i$, and $k = 1, \ldots, N_{ij}$, noting that $J_i \leq J$ is the number of observed periods (i.e., with data collection) for cluster i, and N_{ij} is the cluster-period size. Let μ_{ijk} denote the marginal mean response of y_{ijk}, which is related to the intervention effect δ and jth categorical period effect β_j with link function $g(.)$ via the marginal mean model

$$g(\mu_{ijk}) = \beta_j + u_{ij}\delta \tag{1}$$

Alternatively, categorical period effects in Equation 1 can be replaced with linear period effects:

$$g(\mu_{ijk}) = \beta_0 + \beta_1(t_{ij} - 1) + u_{ij}\delta \tag{2}$$

where β_0 is the intercept, and $\{t_{ij} : j = 1, \ldots, J_i\}$ are integer-valued calendar periods from the study design such that β_1 is the increment in the mean response on the scale of the link function for a unit increase in calendar period, and u_{ij} is the treatment status in cluster i at period j. Linear period effects are particularly useful for, but are not limited to, multi-period CRTs with fewer clusters than periods ($I < J$), such as in the Connect-Home study.

Three types of intervention effect models are implemented in the SAS macro, the widely used *average intervention effects model* (Hussey and Hughes 2007; Hemming *et al.* 2015; Li *et al.* 2018), the *incremental intervention effects model* (Hughes, Granston, and Heagerty 2015) and the *extended incremental intervention effects model*. In the average intervention effects model, u_{ij} is the period-specific treatment indicator (1 for intervention and 0 for control) for cluster i and δ is the intervention effect, irrespective of time on treatment, on the link function scale. With the specification of u_{ij}, population-averaged models could be used for different types of multi-period CRTs. In multi-period CRTs with parallel designs, $u_{ij} = 0$ for all clusters under baseline period j, and $u_{ij} = 0, 1$ in subsequent post-baseline periods depending on the treatment status. In the case of complete SW-CRTs, clusters switch from control condition to intervention at different periods. Thus, $u_{ij} = 0$ in the control period, $j = 1, \ldots, b_i$ and b_i is the total periods under the control condition in cluster i, whereas, in intervention periods, $u_{ij} = 1, j = b_i + 1, \ldots, J_i$. Conversely, the incremental intervention effects model assumes a gradual uptake of the intervention such that its effect depends on time-on-treatment. Assuming a complete SW-CRT is specified with the incremental intervention effects model, the treatment status is $u_{ij} = 0$ in the control period and $u_{ij} = (j - b_i)/q, j = b_i + 1, \ldots, J_i$ in the intervention period, where $q > 0$ is chosen to scale the intervention effect δ according to user specification. In the Connect-Home trial (Figure 1), $q = 10$ so that δ is defined as the full intervention effect on the link function scale after 10 periods, which corresponds to the number of intervention periods for the first SNF. Finally, an extended incremental intervention effects model is considered for designs additionally having a maintenance phase after the active intervention phase with q periods. In SW-CRTs with a maintenance phase, one research question relates to whether the patient benefit from the intervention as captured by the outcome is maintained after the active intervention period has ceased. The treatment status is $u_{ij} = (j - b_i)/q, j = b_i + 1, \ldots, b_i + q$ in the active intervention periods and $u_{ij} = 1$ in the maintenance phase, for $j > b_i + q$.

Specification of the marginal model is completed with the covariance structure of all individual responses in each cluster. Variance of the individual-level response is $\mathsf{VAR}(y_{ijk}) = v_{ijk}\phi$ where v_{ijk} is the variance function and ϕ is the dispersion parameter. For binary responses, $v_{ijk} = \mu_{ijk}(1 - \mu_{ijk})$ and $\phi = 1$, while for continuous responses following typical normal model assumptions, $v_{ijk} = 1$ and ϕ is the constant variance. For count outcomes with Poisson distribution, $v_{ijk} = \mu_{ijk}$ and ϕ is the dispersion parameter. The macro assumes GEE is used to estimate the marginal mean parameters $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_p, \delta)^\top$ as well as the working correlation matrix parameters; common multilevel correlation structures for multi-period CRTs will be presented in Section 4.1 during the SAS macro description.

2.2. Variance in GEE analysis of multi-period CRTs

Using GEE analysis, parameters in the marginal mean model are estimated through the estimating equations (Liang and Zeger 1986): $\boldsymbol{U} = \sum_{i=1}^{I} \boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0}$, where the response vector is $\boldsymbol{Y}_i = (y_{i11}, \ldots, y_{iJ_iN_{iJ_i}})^\top$ and its expected mean vector is $\boldsymbol{\mu}_i = (\mu_{i11}, \ldots, \mu_{iJ_iN_{iJ_i}})^\top$. Let $\boldsymbol{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\theta}^\top}$ and $\boldsymbol{V}_i = \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i \boldsymbol{A}_i^{1/2}$, in which $\boldsymbol{R}_i$ is the working

Design	Correlation structure	Label	$j = j'$	$j \neq j'$	
Cross sectional	Nested exchangeable	NE	α_1	α_2	–
	Exponential decay	ED	α_0	$\alpha_0 r_0^{|j-j'|}$	–
			$j = j'$	$j \neq j', k \neq k'$	$j \neq j', k = k'$
Cohort	Block exchangeable	BE	α_1	α_2	α_3
	General proportional decay	PD	α_0	$\alpha_0 r_0^{|j-j'|}$	$r_1^{|j-j'|}$

Table 1: Intra-cluster correlation: $\mathrm{corr}\left(y_{ijk}, y_{ij'k'}\right)$ under different correlation structures for SW-CRTs for cluster i, period j and individual k).

correlation with correlation parameters $\boldsymbol{\alpha}$ and $\boldsymbol{A}_i = \phi \, \mathrm{diag}\{v_{i11}, \ldots, v_{iJ_iN_{iJ_i}}\}$. Commonly, the model-based variance estimator for $\mathsf{VAR}_{\mathrm{MB}}(\hat{\boldsymbol{\theta}}) = \left(\sum_{i=1}^{I} \boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{D}_i\right)^{-1}$ is used in the power calculation.

The choice of working correlation structure usually depends on design features of the CRT. Four within-cluster correlation structures for multi-period CRTs are summarized in Table 1. Specifically, there are two distinct correlation structures for each of the cohort and cross-sectional design. Each structure incorporates the usual ICC, which measures the correlation between responses from different individuals within the same cluster during the same period: $\mathrm{corr}\left(y_{ijk}, y_{ijk'}\right) = \alpha_0$ or α_1, $k \neq k'$. For cross-sectional designs, the *nested exchangeable* correlation structure additionally specifies a correlation parameter α_2 for observation pairs collected from different periods. Alternatively, *exponential decay* assumes the between-period correlation between responses from different individuals within the same cluster in the j_{th} and j'_{th} periods decays over time as $\alpha_0 r_0^{|j-j'|}$. For cohort designs, the *block exchangeable* correlation structure distinguishes between-period correlations for pairs of individuals, α_2, from a constant intra-individual correlation for repeated observations, α_3 (Li *et al.* 2018; Preisser *et al.* 2003). On the other hand, the *general proportional decay* correlation structure (Gallis, Wang, Rathouz, Preisser, Li, and Turner 2022) allows for correlation decay over time, where the intra-individual correlation $\mathrm{corr}\left(y_{ijk}, y_{ij'k}\right) = r_1^{|j-j'|}$, $j \neq j'$ has a first-order auto-regressive structure decay rate r_1 and the between-period correlation among responses from different individuals within the same cluster is $\alpha_0 r_0^{|j-j'|}$, $j \neq j', k \neq k'$. When $r_0 = r_1$, it reduces to the proportional decay correlation structure (Li 2020; Lefkopoulou, Moore, and Ryan 1989). Note that the nested exchangeable correlation for cross-sectional designs is a special case of block exchangeable correlation when $\alpha_2 = \alpha_3$. Finally, the simple exchangeable correlation (Hussey and Hughes 2007) additionally specifies that within- and between-period correlations are equal $\alpha_1 = \alpha_2$.

Rochon (1998) proposed a power calculation procedure using the model-based variance, which depends on the design matrix and specification of the working correlation structure. Typically, simulations are used to determine power for population-averaged models with incomplete multi-period CRTs with complex correlation structures. However, there may be substantial computational burdens in generating empirical power based on GEE with a large number of simulation scenarios. To lessen computational burden, Zhang *et al.* (2023b) adapted the non-simulation based fast GEE power method (Rochon 1998) multi-period complete and incomplete SW-CRTs, which we implement in the SAS macro **CRTFASTGEEPWR**.

3. Fast GEE power for multi-period CRTs

3.1. Overview of the fast GEE power method

The two-sided test of the intervention effect H_0: $\delta = 0$ vs H_1: $\delta \neq 0$ is based upon the asymptotic normal distribution of $\sqrt{I}(\hat{\delta} - \delta)$ with mean zero and variance determined by the (p, p)-th element of $\mathsf{VAR}(\sqrt{I}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}))$, when I is sufficiently large (Li *et al.* 2018). In turn, the Wald-test statistic $\hat{\delta}/\mathsf{VAR}(\hat{\delta})$ has an asymptotically standard normal distribution under the null hypothesis. Thus, power to detect an intervention effect of size $\delta \neq 0$ with a nominal type I error rate α is $\Phi\left(z_{\alpha/2} + |\delta|/\sqrt{\mathsf{VAR}(\hat{\delta})}\right)$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $z_{\alpha/2}$ is the normal quantile such that $\Phi(z_{\alpha/2}) = \alpha/2$. However, for CRTs with a small number of clusters, the t test is a good alternative with power modified as $\Phi_{t,I-p}\left(t_{\alpha/2,I-p} + |\delta|/\sqrt{\mathsf{VAR}(\hat{\delta})}\right)$ where $t_{\alpha/2,I-p}$ is the $\alpha/2\%$ quantile of the t distribution with $I - p$ degrees of freedom. Typically, the degrees of freedom of the t statistic in CRTs is set to $I - p$, where $p = \dim(\boldsymbol{\theta})$ is the number of estimated marginal mean model parameters in Equation 1; some authors have used $I - 2$, which is sometimes preferred for multi-period CRTs with fewer number of clusters than periods (Li 2020; Ford and Westgate 2020). The model-based variance of the intervention effect $\mathsf{VAR}(\hat{\delta})$ in the determination of power is defined as the (p, p)-th element in the model-based variance matrix $\mathsf{VAR}_{\mathrm{MB}}(\hat{\boldsymbol{\theta}})$. We refer to this general analytical power method (Rochon 1998) as "fast GEE power" because it is a computationally fast power calculation procedure for CRTs with GEE analysis.

3.2. Adaption of the fast GEE power for incomplete SW-CRTs

The fast GEE power procedure has been previously investigated for complete SW-CRTs for the average intervention effects marginal mean model with categorical period effects (Li *et al.* 2018). Motivated by Connect-Home, we define a class of incomplete designs for which Connect-Home is an archetype allowing for implementation periods and/or staggered entry/termination. Let b_{i0} and b_{i1} denote the first and last calendar periods of data collection for cluster i in the control condition (e.g., $b_{20} = 2$ and $b_{21} = 7$ for the second sequence, $s = 2$, in Figure 1), such that there are $b_i = b_{i1} - b_{i0} + 1$ total periods in the control condition; q_{i0} and q_{i1} as the first and last calendar periods of data collection for cluster i in the intervention condition (e.g., $q_{20} = 10$ and $q_{21} = 18$, $s = 2$) such that there are $q_i = q_{i1} - q_{i0} + 1$ total periods in the intervention condition; and c_i implementation periods occurring in calendar periods $b_{i1} + 1, \ldots b_{i1} + c_i$ where $c_i = q_{i0} - b_{i1} - 1$.

A key step in applying the fast GEE power computation to incomplete SW-CRTs is the generation of cluster-level design matrices, consisting of covariates in Equation 1 or 2. For incomplete SW-CRTs, we specify a Design Pattern (DP) matrix by the notation of $\{b_{i0}, b_{i1}, q_{i0}, q_{i1}, i = 1, \ldots, I\}$ to represent the experimental design analogous to the power analysis of continuous responses in linear mixed models (Hemming *et al.* 2015). Each element in the DP matrix corresponds to a representative cluster-period in the SW-CRT design with entries of 0 for control condition, 1 for intervention condition and 2 for cluster-periods in sequences without data collection. To illustrate the specification of DP matrix, an example of an incomplete SW-CRT (Hemming *et al.* 2015) is used here with $S = 2$ treatment sequences, $T = 4$ periods, and an implementation period that occurs in period 2 for the first

sequence and in period 3 for the second sequence. The incomplete design is specified by $(b_{10}, b_{11}, q_{10}, q_{11}) = (1, 1, 3, 4)$ for $s = 1$ and $(b_{20}, b_{21}, q_{20}, q_{21}) = (1, 2, 4, 4)$ for $s = 2$. Then

$$\mathrm{DP} = \left(\begin{array}{cccc} 0 & 2 & 1 & 1 \\ 0 & 0 & 2 & 1 \end{array} \right)$$

The DP matrix serves to modify the design matrix obtained under the complete design to more accurately determine $\mathsf{VAR}_{\mathrm{MB}}(\hat{\boldsymbol{\theta}})$ and thus $\mathsf{VAR}(\hat{\delta})$ in the power calculation for CRTs with incomplete design.

4. The **SAS** macro details

4.1. Input arguments in the macro

A SAS macro **CRTFASTGEEPWR** that implements the fast GEE power method is developed for multi-period CRTs with complete and incomplete designs and available at `https://www.bios.unc.edu/~preisser/personal/software.html`.

Table 2 provides required and optional arguments in the macro, which are classified into three aspects: describing characteristics of the multi-period CRT, parameterizing the marginal mean model and choosing the working correlation structure. All required arguments are in boldface in the first column of Table 2.

First, users are required to describe characteristics of the multi-period CRT through a design pattern matrix, specified by `DesignPattern`, containing the number of sequences and periods, numerical indicators for treatment status, and the incompleteness in the design. The number of clusters in sequences are specified by a vector `M` to allow varying number of clusters across sequences. The SAS macro applies to cross-sectional and closed-cohort multi-period CRT designs but not to open cohort designs. Specifically, the SAS macro allows a varying number of participants across cluster periods for cross-sectional designs through the specification of `CP_size_matrix`. For closed-cohort designs, cluster-period sizes within rows (sequences) of the `CP_size_matrix` should be non-increasing. Equal cluster-period sizes within rows imply no dropout, whereas decreasing numbers are assumed to be due to dropout only (not to intermittent missing data) so that the design matrices for clusters are unambiguous. An example of decaying cohort size due to dropout in a closed-cohort CRT using **CRTFASTGEEPWR** is in Appendix C.1.

Marginal mean model options include binary, count, and continuous responses with three link functions, specified by `dist` and `link`, respectively. Note that the default link function is the canonical link. Meanwhile, the categorical period effects model or the linear period effects model is selected by specifying `period_effect_type`. The user also needs to choose one of three intervention effects models introduced in Section 2, specified by `intervention_effect_type`. For the incremental intervention effects model, `max_intervention_period` should be filled with the number of periods at which the full treatment effect δ is reached. For the extended incremental intervention effects model, the `max_intervention_period` is the number of periods under active intervention phase and there should be at least one maintenance period in each sequence. The intervention effect

Macro variable	Input	Description
`DesignPattern`	Variable	An $S \times T$ matrix with 0 representing control periods, 1 for intervention periods, and 2 for periods without data collection.
`CP_size_matrix`	Variable	Number of cluster-period sizes, a matrix vector with dimension $S \times T$
`M`	Variable	Number of clusters in each sequence, a vector with dimension $S \times 1$
`dist`	BINARY POISSON NORMAL	The distribution for the outcomes
`link`	LOGIT LOG IDENTITY	Link function for the outcomes. For binary, count, and continuous responses, the default link is logit, log and identity, respectively
`phi`	Variable	The dispersion parameter
`intervention_effect_type`	AVE INC INC_EX	Average intervention effects model Incremental intervention effects model Extended incremental intervention effects model
`period_effect_type`	CAT LIN	Categorical period effects model Linear period effects model
`delta`	Variable	The parameters of intervention
`beta_period_effects`	Variable	The parameters for period effects: A $T \times 1$ vector for categorical period effect A 2×1 vector for continuous period effect
`corr_type`	NE/ED BE/PD	NE: Nested Exchangeable, ED: Exponential decay, BE: Block exchangeable, PD: General proportional decay
`alpha0`	Variable	The within-period correlation in ED and PD correlation structures
`r0`	Variable	Between-subject correlation decay rate over time in ED and PD correlation structures
`r1`	Variable	Within-subject correlation decay rate over time in the PD correlation structure
`alpha1`	Variable	The within-period correlation in NE and BE correlation structures
`alpha2`	Variable	The inter-period correlation in NE and BE correlation structures
`alpha3`	Variable	The within-subject correlation in the BE correlation structure
`max_intervention_period`	Variable	The number of intervention periods to reach full intervention effects in incremental intervention effects model and extended incremental intervention effects model
`alpha`	Variable	Significance level with two sided test, default at 0.05
`df_choice`	1	Degrees of freedom, $df = I - p$ is default; p is number of mean model parameters
	2	$df = I - 2$

Table 2: Input arguments in the SAS macro **CRTFASTGEEPWR**.

size and period effects at the scale of link function are all required with the choice of specific intervention and period effects model through `delta` and `beta_period_effects`.

The working correlation structure is specified through the specification of `corr_type` and corresponding intra-cluster correlations (ICCs). The macro does not have an option for cross-sectional versus cohort design, but rather the type of study design should inform the user's choice of correlation structure as suggested in Table 1. The significance level for two-sided tests, specified by `alpha` is optional with a default value of 0.05. In the degrees of freedom determination, the number of clusters minus the number of marginal mean parameters, $df = I - p$, is the default formula. Another degrees of freedom determination $df = I - 2$ in Li (2020) is also available, specified by `df_choice`.

There are some consistency checks in SAS macro **CRTFASTGEEPWR** to ensure parameters are reasonably specified for the power calculation. First, if there are cluster periods in the design pattern matrix with no data collection (i.e., 2 in `DesignPattern`), the corresponding locations in the `CP_size_matrix` need to be 0. Second, the length of `beta_period_effects` should match the selected period effects model. For the linear period effects model, there are two period effects parameters. While for the categorical period effects model, the number of period effects parameters is equal to the column size of the Design pattern matrix `DesignPattern`. Finally, the marginal mean outcome μ_{ijk} needs to be within the rational range based on the specific outcome type, such as μ_{ijk} within (0,1) for binary outcomes and μ_{ijk} are non-negative for count outcomes. Specifically, for binary outcomes, the Frèchet bounds are also checked based on the specification of working correlation structures and the marginal means to ensure their compatibility (Qaqish 2003).

4.2. Multi-period CRT examples of **SAS** macro **CRTFASTGEEPWR**

In this section, we focus on illustrating the power calculation of two complete and two incomplete stepped wedge cluster randomized trials using the SAS macro **CRTFASTGEEPWR**, with different outcome types, specification of marginal mean models, and correlation structures.

The first example illustrates power calculation based on the Connect-Home trial design (Figure 1) with linear period effects for a continuous outcome, patient preparedness for home care (a scale with range 0 to 100) assessed 7 days after discharge from the SNF. There are 6 sequences with 22 periods in the study, having 7 periods without patient enrollment and 15 periods with patient enrollment in each cluster. In the power calculation, there is 1 cluster in each sequence and 360 patients in the trial, in which 4 patients enrolled in each non-missing cluster period. In `CP_size_matrix`, 0 means no patients enrolled in the specific sequence and period, which corresponds to locations with 2 in the design pattern matrix. We assume the baseline patient preparedness score as $\beta_0 = 68$ and a gently increasing linear period effect such that $\beta_1 = 0.1$ for $J = 1, \ldots, 22$ with common variance $\phi = 64$ (standard deviation = 8). The full effect size is reached at 10 months on intervention condition ($q = 10$) for the incremental intervention effects model with $\delta = 10$. ICCs are specified with $(\alpha_1, \alpha_2) = (0.03, 0.015)$ under the nested exchangeable correlation structure to indicate a small within-cluster correlation for the cross-sectional design. The power is calculated using the z test with normal approximation and t test with $df = 3$. From the results, the power using the z-test is much greater than the t test. Simulation studies have shown that the z test is too optimistic and tends to have an inflated test size in SW-CRTs with a small number of clusters (Li *et al.* 2018). Thus, we recommend calculating power with the t test for the Connect Home study.

```
%CRTFASTGEEPWR(alpha = 0.05, m = %str(J(6, 1, 1)), corr_type = NE,
   alpha1 = 0.03, alpha2 = 0.015, intervention_effect_type = INC,
   delta = 10, period_effect_type = LIN, max_intervention_period = 10,
   beta_period_effects = %str({68, 0.1}), dist = normal, phi = 64,
   CP_size_matrix = %str({4 4 4 4 4 0 0 4 4 4 4 4 4 4 4 4 4 0 0 0 0 0,
                          0 4 4 4 4 4 4 4 0 0 4 4 4 4 4 4 4 4 4 0 0 0 0 0,
                          0 0 4 4 4 4 4 4 4 4 0 0 4 4 4 4 4 4 4 4 0 0 0 0,
                          0 0 0 4 4 4 4 4 4 4 4 4 0 0 4 4 4 4 4 4 4 0 0 0,
                          0 0 0 0 4 4 4 4 4 4 4 4 4 4 0 0 4 4 4 4 4 4 0,
                          0 0 0 0 0 4 4 4 4 4 4 4 4 4 4 4 0 0 4 4 4 4 4}),
   DesignPattern = %str({ 0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2,
                          2 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 1 2 2 2 2,
                          2 2 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 2 2 2,
                          2 2 2 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 2 2,
                          2 2 2 2 0 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 2,
                          2 2 2 2 2 0 0 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1}));


 The fast GEE power of NORMAL outcomes with nested exchangeable correlation
                  structure and (alpha1,alpha2):(0.03, 0.015)
         Under incremental intervention effects model and delta = 10
```

T	S	clusters	df	theta	totaln	Dist	LINK	stddel	zpower	tpower
22	6	6	3	68	360	NORMAL	IDENTITY	3.9139	0.9746	0.7413
				0.1						
				10						

The second example aims to calculate the power for a count outcome based on the Connect-Home trial design (Figure 1), which is the number of days of acute care use for patients within 60 days after discharge from the SNF. In this example, an average intervention effect model is specified with the same linear period effect model as in the first example. The design is identical to the first example but the number of clusters increases from 6 to 12 clusters to achieve enough power for the count outcome. For the parameters in the marginal mean model with log link, baseline number of acute care use is assumed to be 1.24 days with an slightly decreasing period effect over time, giving $\beta_0 = \log(1.24) = 0.215$ and $\beta_1 = -0.01$ with the dispersion parameter $\phi = 1.2$. Assuming the intervention reduces the mean number of acute care days by 40%, the intervention effect under the average intervention effects model is $\delta = \log(0.6) = -0.511$. ICCs are specified with $(\alpha_0, r_0) = (0.03, 0.8)$ under exponential decay correlation structure to indicate a small within-cluster correlation for the cross-sectional design. The power is calculated using the z test with normal approximation and the t test with $df = 9$. In the example with 12 clusters, the z test gives slightly higher power compared to the t test, which indicates potentially overly optimistic power calculation as shown in simulation studies (Zhang, Preisser, Li, Turner, Toles, and Rathouz 2023a).

```
%CRTFASTGEEPWR(alpha = 0.05, m = %str(J(6, 1, 2)), corr_type = ED,
   alpha0 = 0.03, R0 = 0.8, intervention_effect_type = AVE, delta = -0.511,
   period_effect_type = LIN, beta_period_effects = %str({0.215, -0.01}),
   dist = poisson, phi = 1.2,
```

Clusters	Period					
	1	2	3	4	5	6
1-8						
9-16						
17-24						
25-32						
33-40						

Figure 2: The trial diagram of the decision-making trial: the study includes 40 clusters and each treatment sequence includes 8 clusters. The blue and green cells denote control and intervention cluster-periods, respectively.

```
CP_size_matrix = %str({4 4 4 4 4 0 0 4 4 4 4 4 4 4 4 4 4 0 0 0 0 0,
                       0 4 4 4 4 4 4 0 0 4 4 4 4 4 4 4 4 4 0 0 0 0 0,
                       0 0 4 4 4 4 4 4 4 0 0 4 4 4 4 4 4 4 4 0 0 0 0,
                       0 0 0 4 4 4 4 4 4 4 4 0 0 4 4 4 4 4 4 4 0 0 0,
                       0 0 0 0 4 4 4 4 4 4 4 4 4 0 0 4 4 4 4 4 4 0,
                       0 0 0 0 0 4 4 4 4 4 4 4 4 4 4 0 0 4 4 4 4 4 4}),
DesignPattern = %str({0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2,
                      2 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 1 2 2 2 2 2,
                      2 2 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 2 2 2,
                      2 2 2 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 2 2,
                      2 2 2 2 0 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 2,
                      2 2 2 2 2 0 0 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1}));
```

```
The fast GEE power of POISSON outcomes with exponential decay correlation
              structure and (alpha0,r0):(0.03, 0.8)
        Under average intervention effects model and delta = -0.511


 T   S clusters  df   theta  totaln    Dist    Link  stddel  zpower  tpower
22   6        12   9   0.215     720 POISSON     LOG  3.1096  0.8749  0.7906
                        -0.01
                       -0.511
```

The third example illustrates power calculation for a cross-sectional SW-CRT to improve pre-operative decision-making, by the use of a patient-driven question prompt list intervention (Taylor *et al.* 2017; Schwarze *et al.* 2020). In the third example, 480 patients enrolled across six periods are clustered within 40 surgeons who are randomized to transition from control (blue cells) to intervention condition (green cells) at one of five randomly assigned sequences (8 surgeons per sequence; Figure 2). We calculate the power for a binary primary outcome regarding whether the patient has a post-treatment regret. We assume a balanced and complete design for the study, 12 patients for each surgeon with two patients in each cluster period. In the marginal mean model for the binary outcome with logit link and average intervention effects model, the control is assumed to have 2.2 times the odds of reporting post-treatment regret compared to the intervention group, given by $\delta = \log(1/2.2) = -0.789$. The average probability of post-treatment regret at baseline is

assumed to be 0.22, such that $\beta_0 = \log(0.22/0.78) = -1.266$ with an consistent increasing period effects $\beta_i = 0.01, i = 1, \ldots, 5$. For the working correlation structure, we used the exponential decay correlation structure with ICCs $(\alpha_0, r_0) = (0.03, 0.8)$. Power using the t test and the z test both reach 80% and are similar to one another considering the moderately large number of clusters.

```
%CRTFASTGEEPWR(alpha = 0.05, m = %str(J(5, 1, 8)), corr_type = ED,
  intervention_effect_type = AVE, delta = -0.789, period_effect_type = CAT,
  beta_period_effects = %str({-1.266, 0.01, 0.01, 0.01, 0.01, 0.01}),
  alpha0 = 0.03, R0 = 0.8, dist = binary, phi = 1,
  CP_size_matrix = %str(J(5, 6, 2)),
  DesignPattern =%str({0 1 1 1 1 1,
                       0 0 1 1 1 1,
                       0 0 0 1 1 1,
                       0 0 0 0 1 1,
                       0 0 0 0 0 1}));


   The fast GEE power of binary outcomes with exponential decay correlation
                 structure and (alpha0,r0):(0.03, 0.8)
        Under average intervention effects model and delta = -0.789


T  S clusters  df  theta   totaln  Dist    Link   stddel  zpower  tpower
6  5       40  33 -1.266      480  BINARY  LOGIT   2.917  0.8307  0.8081
                     0.01
                     0.01
                     0.01
                     0.01
                     0.01
                    -0.789
```

The fourth example is based on the Heart Health NOW (HHN) study with a large number of clusters. We assume the HHN study using a complete, stratified, SW-CRT evaluating the effect of primary care practice support on evidence-based cardiovascular disease (CVD) prevention, organizational change process measures, and patient outcomes, the latter captured by electronic health records (EHR, Weiner *et al.* 2015). Medical practices are randomized to receive the intervention at one of three time points (steps) within two strata defined by high (the first three treatment sequences, Figure 3) or low (last three rows) readiness for change. After four quarters in the intervention phase (green boxes), each practice enters a maintenance phase (gray boxes) for two to six quarters depending upon the allocated treatment sequence. HHN was a quality improvement research project whereby the intervention of practice facilitation aimed to bring about enduring change in patient-centered and organizational outcomes. We consider a combined binary outcome regarding whether there is hospitalization due to stroke, acute myocardial infarction, or angina for patients. We assume that there are 30 medical practices (clusters) in each sequence (practice cohort) with 100 patients enrolled in each cluster period; thus 1100 patients will be enrolled in each cluster. If the baseline probability of hospitalization is 0.05, making $\beta_0 = \log(0.05/0.95) = -2.944$ with an consistent decreasing period effects $\beta_1 = -0.01$. Under the extended incremental intervention effects model, the

Year	2015	2016				2017				2018	
Quarter (period)	4	1	2	3	4	1	2	3	4	1	2
practice cohort	High Readiness										
1 (n=34)											
2 (n=27)											
3 (n=30)											
	Low Readiness										
4 (n=36)											
5 (n=34)											
6 (n=58)											
		Control periods				Intervention periods			Maintenance periods		

Figure 3: The study design of the HHN study: the blue, green and gray cells denote control, active intervention and maintenance cluster-periods, respectively.

intervention effect is assumed to decrease the odds of hospitalization at the end of 4 quarters by 25% under the active intervention condition, $\delta = \log(0.75) = -0.288$, maintaining the same effect size in the maintenance periods. The working correlation structure for the binary outcome is a nested exchangeable correlation structure with ICCs $(\alpha_1, \alpha_2) = (0.03, 0.015)$. Considering the large cluster size and number of clusters, powers is very similar under the z test and t test, reaching 78%.

```
%CRTFASTGEEPWR(alpha = 0.05, m = %str(J(6, 1, 30)), corr_type = NE,
  intervention_effect_type = INC_EX, period_effect_type = LIN,
  delta = -0.288, max_intervention_period = 4,
  beta_period_effects = %str({-2.944, -0.01}),
  alpha1 = 0.03, alpha2 = 0.015, dist =  binary, phi = 1,
  CP_size_matrix = %str(J(6, 11, 100)),
  DesignPattern = %str({0 1 1 1 1 1 1 1 1 1 1,
                        0 0 1 1 1 1 1 1 1 1 1,
                        0 0 0 1 1 1 1 1 1 1 1,
                        0 0 0 1 1 1 1 1 1 1 1,
                        0 0 0 0 1 1 1 1 1 1 1,
                        0 0 0 0 0 1 1 1 1 1 1}));
```

```
 The fast GEE power of binary outcomes with nested exchangeable correlation
              structure and (alpha1,alpha2):(0.03, 0.015)
  Under extended incremental intervention effects model and delta = -0.288


 T  S  clusters   df  theta totaln    Dist    Link  stddel  zpower  tpower
11  6       180  177 -2.944 198000  BINARY   LOGIT  2.7477  0.7846  0.7801
                     -0.01
                     -0.288
```

In the appendix, we also describe the power calculation of multi-period parallel cluster randomized trials, as well as SAS codes to calculate powers under varying effect sizes using the SAS macro.

5. Discussion

This article proposes a fast GEE power method for binary, count, and continuous responses of complete and incomplete multi-period CRTs, including parallel-arm longitudinal CRTs, cluster randomized crossover trials and SW-CRTs. The fast GEE power approach was illustrated in the planning of four complete and incomplete cross-sectional stepped wedge designs for binary, count and continuous outcomes. The SAS macro **CRTFASTGEEPWR** is novel in several aspects. Through specification of the Design Pattern matrix in the spirit of Hemming *et al.* (2015) and Rochon (1998), the general GEE power method in the SAS macro is implemented for four within-cluster correlation structures proposed in the recent literature for multi-period CRTs. To our knowledge, there is no other statistical softwares that unifies the multi-level correlation structures for designing cross-sectional and cohort CRTs. This proposed software for power of multi-period CRTs based on marginal models fills a gap by adding to the literature of power calculators, which has mostly focused on mixed models (Hughes *et al.* 2015; Hemming, Kasza, Hooper, Forbes, and Taljaard 2020).

In the SAS macro **CRTFASTGEEPWR**, the user needs to choose from among three parameterizations of the intervention effect and two specifications of the period effects. Categorical or linear period effects may be chosen. A linear time trend could be appropriate when the underlying secular trend under control condition increases or decreases at an approximately constant rate over time (e.g., an educated speculation from historic data before the planning of the study). When the secular trend is non-linear or when little prior information is available, categorical period effects might be a more robust specification and with fewer assumption compared to a linear time trend. In the special case of a complete stepped wedge design where the treatment sequences are balanced and the outcome is continuous, Grantham, Forbes, Heritier, and Kasza (2020) has shown that the model-based variance of the intervention effect estimator under linear period effects (Model 1) is identical to that under categorical period effects (Model 2). In more general cases, however, one would expect the variance of the intervention effect estimator to be larger under Model 1 due to the additional cost in degrees of freedom for estimating more period effect parameters. Therefore, whenever feasible, Model 1 can be considered as a more conservative option for power calculation, and the specified degrees of freedom should be concordant with this conservative analysis option.

As the power calculation of the macro **CRTFASTGEEPWR** is based on the t distribution, it offers two choices of degree of freedom based upon the intended GEE analysis, either $I - 2$ or $I - p$, where I is the number of clusters and p is the number of regression coefficients in the marginal mean model. The options are based on previous simulation studies for multi-period CRTs (Li *et al.* 2018; Li 2020; Li, Yu, Rathouz, Turner, and Preisser 2022; Zhang *et al.* 2023b). In general, when the number of clusters is less than the number of periods, the user must set $df = I - 2$ or specify the linear time trend in Model 2 with $df = I - 3$ since otherwise the degrees of freedom becomes negative. As a concrete example, in the Connect-Home trial (Figure 1), $I = 6$ and $J = 22$ and, thus, $I - p = 6 - 23 = -17 < 0$. Therefore, a linear time trend is recommended when using $I - p$ as the degrees of freedom. Zhang *et al.* (2023b) showed that predicted power based on the fast, non-simulation GEE power method with $df = I - p$ has good agreement to the empirical power for testing the intervention effect from simulation studies when the bias-corrected sandwich variance estimator of Kauermann and Carroll (2001) is used in GEE analysis. Others have recommended using $df = I - 2$ (Ford and Westgate 2020; Li 2020; Li *et al.* 2022) based on simulation studies.

Several variance estimators have been widely studied for use in GEE analysis. Based on simulation studies of its finite-sample properties, the bias-corrected variance estimator of Kauermann and Carroll (2001) has been recommended for use in GEE analysis of parallel-groups CRTs (Preisser *et al.* 2003), multi-level CRTs (Teerenstra *et al.* 2010), crossover CRTs (Li, Forbes, Turner, and Preisser 2019), and SW-CRTs (Li *et al.* 2018) with as few as 8 to 10 clusters. For as few as six clusters, recommendations are less certain. In this case, Zhang *et al.* (2023b) suggest that the GEE with the model-based variance matrix may be the best choice. In the case of constructing confidence intervals for ICCs, Preisser *et al.* (2008) and Li *et al.* (2022) recommend the bias-variance estimator of Mancl and DeRouen (2001) and matrix-adjusted estimating equations (MAEE) for point estimates. Evidence in the literature for count outcomes is relatively scarce compared to that for binary and continuous outcomes.

A limitation of the **SAS** macro **CRTFASTGEEPWR** is that the power procedure that it implements is based on the GEE model-based variance estimator, when sandwich variance estimators (Zeger, Liang, and Albert 1988), which validly estimate the true variance under misspecification of the correlation structure, are often used in analysis. As noted in the introduction, under correct specification of the correlation matrix, the accuracy of the fast GEE power method based on the GEE model-based variance matrix has been validated in simulation studies based on GEE with a bias-corrected variance estimator for a few as eight clusters (Kauermann and Carroll 2001; Li *et al.* 2022). Future research could investigate the use of asymptotic sandwich variance formulae in the fast GEE power method under correlation structure mis-specification. For example, if the planned GEE analysis specifies the working independence correlation matrix (Thompson, Hemming, Forbes, Fielding, and Hayes 2021), power could be computed under one or more assumed true non-independence correlation structures. Another adaptation of the fast GEE power method, with the goal of improving Type I error control in CRTs with small number of clusters, would base the power calculation on the sandwich variance formula that is the expectation of a bias-corrected empirical variance estimator (Kauermann and Carroll 2001; Mancl and DeRouen 2001; Fay and Graubard 2001).

Several recent software programs have been developed for the GEE analysis of studies with a small number of clusters. For researchers interested in GEE analysis of binary responses with bias-corrected variance estimators for CRTs, the **SAS** macro **GEECORR** by Shing, Preisser, and Zink (2021) is available. A **SAS** macro **GEEMAEE** developed by the authors for binary, count, and continuous outcomes additionally applies bias-corrections in the estimation of ICCs and their variance estimators, which is desirable for reporting ICC parameters as recommended by the CONSORT statement for stepped wedge trials (Hemming *et al.* 2018). The **GEECORR** and **GEEMAEE** **SAS** macro are available at `https://www.bios.unc.edu/~preisser/personal/software.html`.

Acknowledgments

References

Chen J, Zhou X, Li F, Spiegelman D (2022). "**swdpwr**: A SAS Macro and an R Package for Power Calculations in Stepped Wedge Cluster Randomized Trials." *Computer Methods and Programs in Biomedicine*, **213**, 106522. `doi:10.1016/j.cmpb.2021.106522`.

Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR (2015). "Designing a Stepped Wedge Trial: Three Main Designs, Carry-Over Effects and Randomisation Approaches." *Trials*, **16**, 352. `doi:10.1186/s13063-015-0842-7`.

Fay MP, Graubard BI (2001). "Small-Sample Adjustments for Wald-Type Tests Using Sandwich Estimators." *Biometrics*, **57**(4), 1198–206. `doi:10.1111/j.0006-341x.2001.01198.x`.

Ford WP, Westgate PM (2020). "Maintaining the Validity of Inference in Small-Sample Stepped Wedge Cluster Randomized Trials with Binary Outcomes When Using Generalized Estimating Equations." *Statistics in Medicine*, **39**(21), 2779–2792. `doi:10.1002/sim.8575`.

Gallis JA, Wang X, Rathouz PJ, Preisser JS, Li F, Turner EL (2022). "`power swgee`: GEE-Based Power Calculations in Stepped Wedge Cluster Randomized Trials." *The Stata Journal*, **22**(4), 811–841. `doi:10.1177/1536867x221140953`.

Grantham KL, Forbes AB, Heritier S, Kasza J (2020). "Time Parameterizations in Cluster Randomized Trial Planning." *The American Statistician*, **74**(2), 184–189. `doi:10.1080/00031305.2019.1623072`.

Hemming K, Kasza J, Hooper R, Forbes A, Taljaard M (2020). "A Tutorial on Sample Size Calculation for Multiple-Period Cluster Randomized Parallel, Cross-Over and Stepped-Wedge Trials Using the Shiny CRT Calculator." *International Journal of Epidemiology*, **49**(3), 979–995. `doi:10.1093/ije/dyz237`.

Hemming K, Lilford R, Girling AJ (2015). "Stepped-Wedge Cluster Randomised Controlled Trials: A Generic Framework Including Parallel and Multiple-Level Designs." *Statistics in Medicine*, **34**(2), 181–96. `doi:10.1002/sim.6325`.

Hemming K, Taljaard M (2020). "Reflection on Modern Methods: When Is a Stepped-Wedge Cluster Randomized Trial a Good Study Design Choice?" *International Journal of Epidemiology*, **49**(3), 1043–1052. `doi:10.1093/ije/dyaa077`.

Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, Dixon-Woods M, Aldcroft A, Doussau A, Grayling M, Kristunas C, Goldstein CE, Campbell MK, Girling A, Eldridge S, Campbell MJ, Lilford RJ, Weijer C, Forbes AB, Grimshaw JM (2018). "Reporting of Stepped Wedge Cluster Randomised Trials: Extension of The CONSORT 2010 Statement with Explanation and Elaboration." *BMJ*, **363**, k1614. `doi:10.1136/bmj.k1614`.

Hughes JP, Granston TS, Heagerty PJ (2015). "Current Issues in the Design and Analysis of Stepped Wedge Trials." *Contemporary Clinical Trials*, **45**(Pt A), 55–60. `doi:10.1016/j.cct.2015.07.006`.

Hussey MA, Hughes JP (2007). "Design and Analysis of Stepped Wedge Cluster Randomized Trials." *Contemporary Clinical Trials*, **28**(2), 182–191. `doi:10.1016/j.cct.2006.05.007`.

Kasza J, Forbes AB (2019). "Information Content of Cluster-Period Cells in Stepped Wedge Trials." *Biometrics*, **75**(1), 144–152. `doi:10.1111/biom.12959`.

Kasza J, Hemming K, Hooper R, Matthews J, Forbes AB (2019). "Impact of Non-Uniform Correlation Structure on Sample Size and Power in Multiple-Period Cluster Randomised Trials." *Statistical Methods in Medical Research*, **28**(3), 703–716. `doi:10.1177/0962280217734981`.

Kauermann G, Carroll RJ (2001). "A Note on the Efficiency of Sandwich Covariance Matrix Estimation." *Journal of the American Statistical Association*, **96**(456), 1387–1396. `doi:10.1198/016214501753382309`.

Kristunas C, Morris T, Gray L (2017). "Unequal Cluster Sizes in Stepped-Wedge Cluster Randomised Trials: A Systematic Review." *BMJ Open*, **7**, e017151. `doi:10.1136/bmjopen-2017-017151`.

Lefkopoulou M, Moore D, Ryan L (1989). "The Analysis of Multiple Correlated Binary Outcomes: Application to Rodent Teratology Experiments." *Journal of the American Statistical Association*, **84**(407), 810–815. `doi:10.1080/01621459.1989.10478842`.

Li F (2020). "Design and Analysis Considerations for Cohort Stepped Wedge Cluster Randomized Trials with a Decay Correlation Structure." *Statistics in Medicine*, **39**(4), 438–455. `doi:10.1002/sim.8415`.

Li F, Forbes AB, Turner EL, Preisser JS (2019). "Power and Sample Size Requirements for GEE Analyses of Cluster Randomized Crossover Trials." *Statistics in Medicine*, **38**(4), 636–649. `doi:10.1002/sim.7995`.

Li F, Turner EL, Preisser JS (2018). "Sample Size Determination for GEE Analyses of Stepped Wedge Cluster Randomized Trials." *Biometrics*, **74**(4), 1450–1458.

Li F, Yu H, Rathouz PJ, Turner EL, Preisser JS (2022). "Marginal Modeling of Cluster-Period Means and Intraclass Correlations in Stepped Wedge Designs with Binary Outcomes." *Biostatistics*, **23**(3), 772–788. `doi:10.1093/biostatistics/kxaa056`.

Liang KY, Zeger SL (1986). "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, **73**(1), 13–22.

Mancl LA, DeRouen TA (2001). "A Covariance Estimator for GEE with Improved Small-Sample Properties." *Biometrics*, **57**(1), 126–34. `doi:10.1111/j.0006-341x.2001.00126.x`.

Murray DM (1998). *Design and Analysis of Group-Randomized Trials*. Oxford University Press.

Murray DM, Varnell SP, Blitstein JL (2004). "Design and Analysis of Group-Randomized Trials: A Review of Recent Methodological Developments." *American Journal of Public Health*, **94**(3), 423–32. `doi:10.2105/ajph.94.3.423`.

Ouyang Y, Li F, Preisser JS, Taljaard M (2022). "Sample Size Calculators for Planning Stepped-Wedge Cluster Randomized Trials: a Review and Comparison." *International Journal of Epidemiology*, **51**(6), 2000–2013. `doi:10.1093/ije/dyac123`.

Preisser JS, Lu B, Qaqish BF (2008). "Finite Sample Adjustments in Estimating Equations and Covariance Estimators for Intracluster Correlations." *Statistics in Medicine*, **27**(27), 5764–5785. `doi:10.1002/sim.3390`.

Preisser JS, Young ML, Zaccaro DJ, Wolfson M (2003). "An Integrated Population-Averaged Approach to the Design, Analysis and Sample Size Determination of Cluster-Unit Trials." *Statistics in Medicine*, **22**, 1235–54.

Qaqish BF (2003). "A Family of Multivariate Binary Distributions for Simulating Correlated Binary Variables with Specified Marginal Means and Correlations." *Biometrika*, **90**(2), 455–463. `doi:10.1093/biomet/90.2.455`.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Reboussin BA, Preisser JS, Song EY, Wolfson M (2012). "Sample Size Estimation for Alternating Logistic Regression Analysis of Multilevel Randomized Community Trials of Underage Drinking." *Journal of the Royal Statistical Society A*, **175**(3), 691–712. `doi:10.1111/j.1467-985x.2011.01003.x`.

Rochon J (1998). "Application of GEE Procedures for Sample Size Calculations in Repeated Measures Experiments." *Statistics in Medicine*, **17**(14), 1643–1658. `doi:10.1002/(sici)1097-0258(19980730)17:14<1643::aid-sim869>3.0.co;2-3`.

SAS Institute Inc (2020). *The SAS System, Version 15.2*. SAS Institute Inc., Cary. URL `https://www.sas.com/`.

Schwarze ML, Buffington A, Tucholka JL, Hanlon B, Rathouz PJ, Marka N, Taylor LJ, Zimmermann CJ, Kata A, Baggett ND, Fox DA, Schmick AE, Berlin A, Glass NE, Mosenthal AC, Finlayson E, Cooper Z, Brasel KJ (2020). "Effectiveness of a Question Prompt List Intervention for Older Patients Considering Major Surgery: A Multisite Randomized Clinical Trial." *JAMA Surgery*, **155**(1), 6–13. `doi:10.1001/jamasurg.2019.3778`.

Shing TL, Preisser JS, Zink RC (2021). "**GEECORR**: A SAS Macro for Regression Models of Correlated Binary Responses and Within-Cluster Correlation Using Generalized Estimating Equations." *Computer Methods and Programs in Biomedicine*, **208**, 106276. `doi:10.1016/j.cmpb.2021.106276`.

Taylor LJ, Rathouz PJ, Berlin A, Brasel KJ, Mosenthal AC, Finlayson E, Cooper Z, Steffens NM, Jacobson N, Buffington A, Tucholka JL, Zhao Q, Schwarze ML (2017). "Navigating High-Risk Surgery: Protocol for a Multisite, Stepped Wedge, Cluster-Randomised Trial of a Question Prompt List Intervention to Empower Older Adults to Ask Questions That Inform Treatment Decisions." *BMJ Open*, **7**(5), e014002. `doi:10.1136/bmjopen-2016-014002`.

Teerenstra S, Lu B, Preisser JS, Van Achterberg T, Borm GF (2010). "Sample Size Considerations for GEE Analyses of Three-Level Cluster Randomized Trials." *Biometrics*, **66**(4), 1230–7. `doi:10.1111/j.1541-0420.2009.01374.x.`

Thompson JA, Hemming K, Forbes A, Fielding K, Hayes R (2021). "Comparison of Small-Sample Standard-Error Corrections for Generalised Estimating Equations in Stepped Wedge Cluster Randomised Trials with a Binary Outcome: A Simulation Study." *Statistical Methods in Medical Research*, **30**, 425–439. `doi:10.1177/0962280220958735.`

Tian Z, Preisser JS, Esserman D, Turner EL, Rathouz PJ, Li F (2022). "Impact of Unequal Cluster Sizes for GEE Analyses of Stepped Wedge Cluster Randomized Trials with Binary Orutcomes." *Biometrical Journal*, **64**, 419–439. `doi:10.1101/2021.04.07.21255090.`

Toles M, Colón-Emeric C, Hanson LC, Naylor M, Weinberger M, Covington J, Preisser JS (2021). "Transitional Care from Skilled Nursing Facilities to Home: Study Protocol for a Stepped Wedge Cluster Randomized Trial." *Trials*, **22**(1), 1–15. `doi:10.1186/s13063-021-05068-0.`

Turner EL, Li F, Gallis JA, Prague M, Murray DM (2017). "Review of Recent Methodological Developments in Group-Randomized Trials: Part 1-Design." *American Journal of Public Health*, **107**(6), 907–915. `doi:10.2105/ajph.2017.303706.`

Wang X, Turner EL, Preisser JS, Li F (2021). "Power Considerations for Generalized Estimating Equations Analyses of Four-Level Cluster Randomized Trials." *Biometrical Journal. Biometrische Zeitschrift*, **64**(4), 663–680. `doi:10.1002/bimj.202100081.`

Weiner BJ, Pignone MP, DuBard CA, Lefebvre A, Suttie JL, Freburger JK, Cykert S (2015). "Advancing Heart Health in North Carolina Primary Care: The Heart Health NOW Study Protocol." *Implementation Sci*, **10**, 160. `doi:10.1186/s13012-015-0348-4.`

Zeger SL, Liang KY, Albert PS (1988). "Models for Longitudinal Data: A Generalized Estimating Equation Approach." *Biometrics*, **44**(4), 1049–1060. `doi:10.2307/2531734.`

Zhang Y, Preisser JS, Li F, Turner EL, Toles M, Rathouz PJ (2023a). "**GEEMAEE**: A SAS Macro for the Analysis of Correlated Outcomes Based on GEE and Finite-Sample Adjustments with Application to Cluster Randomized Trials." *Computer Methods and Programs in Biomedicine*, **230**, 107362. `doi:10.1016/j.cmpb.2023.107362.`

Zhang Y, Preisser JS, Turner EL, Rathouz PJ, Toles M, Li F (2023b). "A General Method for Calculating Power for GEE Analysis of Complete and Incomplete Stepped Wedge Cluster Randomized Trials." *Statistical Methods in Medical Research*, **32**(1), 71–87. `doi:10.1177/09622802221129861.`

A. Standard GEE analysis of multi-period CRTs

For GEE analysis of multi-period CRTs, we consider a general population averaged regression framework. Let μ_{ijk} denote the marginal mean response for a continuous or categorical outcome y_{ijk}, which is the k_{th} subject in the i_{th} cluster and the j_{th} period, for $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J_i$, and $k = 1, 2, \ldots, n_{ij}$. The mean response μ_{ijk} is related to covariates $\boldsymbol{X_{ijk}}$ with the link function $g(.)$ in the following marginal mean model:

$$g(\boldsymbol{\mu_{ijk}}) = \boldsymbol{X_{ijk}^\top \beta}$$

where $\boldsymbol{\beta}$ is the $(p + 1) \times 1$ parameter corresponding to the column of the design matrix and describes how the population averaged response depends on the covariates. (Zeger *et al.* 1988) Denote the vector of response vector as $\boldsymbol{Y_i} = \left(y_{i11}, \ldots, y_{iJ_in_{iJ_i}}\right)^\top$ with marginal mean vector $\boldsymbol{\mu_i} = \left(\mu_{i11}, \ldots, \mu_{iJ_in_{iJ_i}}\right)^\top$. Let the design matrix in the i_{th} cluster be $\boldsymbol{X_i} = (\boldsymbol{X_{i11}}, \ldots, \boldsymbol{X_{iJ_in_{iJ_i}}})^\top$, where $\boldsymbol{X_{ijk}} = \left(x_{1ijk}, \ldots, x_{(p+1)ijk}\right)^\top$ is a $(p+1) \times 1$ vector by combing p exploratory variables and an intercept. The estimating equations for $\boldsymbol{\beta}$ are:

$$\boldsymbol{U_\beta} = \sum_{i=1}^{I} \boldsymbol{D_i^\top V_i^{-1}} (\boldsymbol{Y_i} - \boldsymbol{\mu_i}) = \boldsymbol{0}$$

where $\boldsymbol{D_i} = \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta^\top}}, \boldsymbol{V_i} = \boldsymbol{A_i R_i(\alpha) A_i}, \boldsymbol{R_i(\alpha)}$ is a working variance matrix with parameter $\boldsymbol{\alpha}$, and $\boldsymbol{A_i} = \text{diag}\left(\sqrt{\text{VAR}(y_{i11})}, \sqrt{\text{VAR}(y_{i12})}, \ldots, \sqrt{\text{VAR}\left(y_{iJ_in_{iJ_i}}\right)}\right)$. The variance of the response is denoted as $\text{VAR}(y_{ijk}) = v_{ijk}\phi$ where v_{ijk} is the variance function and ϕ is the dispersion parameter. For binary responses, $v_{ijk} = \mu_{ijk}(1 - \mu_{ijk})$ and $\phi = 1$, while for continuous responses following typical normal model assumptions, $v_{ij} = 1$ and ϕ is the constant variance. For count response, $v_{ijk} = \mu_{ijk}$ and ϕ is the dispersion parameter. Furthermore, the sandwich variance estimator of $\hat{\boldsymbol{\beta}}$ is:

$$\text{VAR}(\hat{\boldsymbol{\beta}}) = \left[\sum_{i=1}^{I} \boldsymbol{D_i^\top V_i^{-1} D_i}\right]^{-1} \left[\sum_{i=1}^{I} \boldsymbol{D_i^\top V_i^{-1}}(\boldsymbol{Y_i} - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y_i} - \hat{\boldsymbol{\mu}}_i)^\top \boldsymbol{V_i^{-1} D_i}\right] \left[\sum_{i=1}^{I} \boldsymbol{D_i^\top V_i^{-1} D_i}\right]^{-1} \tag{3}$$

On the other hand, sandwich variance estimators have the virtue that they provide consistent estimates of the variance matrix for parameter estimates even when the assumed variance structure fails to hold. When the working covariance is believed to be equal to the true correlation structure, then a consistent estimator for variance estimator $\text{VAR}(\hat{\boldsymbol{\beta}})$ is given by $\boldsymbol{\Sigma_1^{-1}} = \left[\sum_{i=1}^{I} \boldsymbol{D_i^\top V_i^{-1} D_i}\right]^{-1}$. Meanwhile, the dispersion parameter could be estimated by the moment estimator proposed in Liang and Zeger (1986), $\hat{\phi} = \frac{1}{N-p}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\sum_{k=1}^{n_{ij}} r_{ijk}^2$, where the Pearson residual $r_{ijk} = (y_{ijk} - \mu_{ijk})/\sqrt{v_{ijk}}$ is calculated from estimated $\hat{\boldsymbol{\beta}}$ and $N = \sum_{i=1}^{I}\sum_{j=1}^{J_i} n_{ij}$ is the number of all participants in the CRT.

B. GEE analysis with finite-sample corrections

In a recent systematic review of stepped wedge CRTs, Kristunas, Morris, and Gray (2017) reported that the median number of clusters is 11 with a range between 6 to 22 clusters. When CRTs have a small number of clusters, there is a greater chance of unequal distributions of potentially confounding factors among the different conditions, necessitating more sophisticated randomization to maintain the study's intrinsic validity Murray (1998). Moreover, a small number of clusters can significantly degrade the efficiency of GEE analysis. The biased sandwich variance estimates in this section would help to reduce the bias of the standard sandwich variance estimator in equation 3.

B.1. Bias-corrected sandwich variance estimators

Mancl and DeRouen (2001) showed that the GEE sandwich variance estimates in Equation 3 tended to underestimate the variance of intervention effects, which results in a greater Wald-type test size than the nominal level with number of clusters less than 40. The estimated tests sizes with the model-based estimator under the true correlation structure are usually closer to the nominal level compared with the robust estimator but are still often inflated. Thus, bias-corrected sandwich variance estimates should be considered under the situation with a small number of clusters. The formula $\hat{\boldsymbol{\Sigma}}_1^{-1}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\hat{\boldsymbol{\Sigma}}_1^{-1}$ unifies the common types bias-corrected sandwich variance estimates for CRTs, where

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} = \sum_{i=1}^{I} \boldsymbol{F}_{1i}\boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{B}_{1i}\left(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i\right)\left(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i\right)' \boldsymbol{B}_{1i}^\top \boldsymbol{V}_i^{-1}\boldsymbol{D}_i\boldsymbol{F}_{1i} \tag{4}$$

which is evaluated at the solution of the GEEs $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$. When both $\boldsymbol{F}_{1i}$ and $\boldsymbol{B}_{1i}$ are identity matrices, Equation 4 reduces to the uncorrected sandwich estimator $\mathsf{VAR}(\hat{\boldsymbol{\beta}})$ (Liang and Zeger 1986) in Equation 3 referred to as BC0. There are mainly three bias-corrected sandwich variance estimates to reduce the finite sample bias of the uncorrected sandwich variance estimator. Setting $\boldsymbol{F}_{1i} = \boldsymbol{I}$ and $\boldsymbol{B}_{1i} = (\boldsymbol{I} - \boldsymbol{H}_{1i})^{-1/2}$ gives the finite-sample correction of Kauermann and Carroll (2001) or referred as BC1, where $\boldsymbol{H}_{1i} = \boldsymbol{D}_i(\sum_{i=1}^{I}\boldsymbol{D}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{D}_i)^{-1}\boldsymbol{D}_i^\top\boldsymbol{V}_i^{-1}$. Next, setting $\boldsymbol{F}_{1i} = \boldsymbol{I}$ and $\boldsymbol{B}_{1i} = (\boldsymbol{I} - \boldsymbol{H}_{1i})^{-1}$ gives the finite-sample correction of Mancl and DeRouen (2001) or referred as BC2. Because the matrix elements of the cluster leverage are between 0 and 1, there is an order of bias-correction BC0 < BC1 < BC2 (Preisser *et al.* 2008). Finally, setting $\boldsymbol{F}_{1i} = \mathrm{diag}\left\{\left(1 - \min\left\{\zeta, \left[\boldsymbol{D}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{D}_i\boldsymbol{\Sigma}_1^{-1}\right]_{jj}\right\}\right)^{-1/2}\right\}$ and $\boldsymbol{B}_{1i} = \boldsymbol{I}$ gives the finite-sample correction of Fay and Graubard (2001) or BC3, where the bound parameter ζ is a user-defined constant (< 1) with a default value 0.75.

The number of clusters and the aim of the finite sample adjustment influence the choice of bias-corrected sandwich variance estimators. Most research is focused on bias-corrected sandwich variance estimators for the intervention effect. In recent work, Li *et al.* (2018) recommended using BC1 or BC3 and MAEE with a t-test for simulated power computation in the design phase for stepped wedge CRTs as few as 8 clusters. Zhang *et al.* (2023b) found that for incomplete stepped wedge CRTs with a small number of clusters, GEE analysis based on BC1 with MAEE is recommended to maintain the suitable type I error rate based on t-test.

C. Examples in applying **CRTFASTGEEPWR** for CRT designs

C.1. Calculate power for a closed-cohort CRT with decaying cluster sizes

We demonstrate the power calculation for CRTs with decreasing cohort size, based on a hypothetical closed-cohort CRT with the same design as the Connect-Home trial. As in Section 4.1, we assume 4 patients are recruited at baseline and 1 patient drops out per cluster after 15 periods after their baseline period. The number of observation per cluster-cluster is shown in the following codes in `CP_size_matrix`. All of the parameters in the marginal mean model are set to the same values as those in Section 4.1. When describing the correlation for patients with repeated measurements in a cluster, we utilize the block exchangeable correlation structure with the within period ICC $\alpha_1 = 0.03$, between-patient and between-period ICC $\alpha_2 = 0.015$, and within-patient and between-period ICC $\alpha_3 = 0.2$. In the software output, there are 348 observations collected from 6 clusters and the power using z test is still above 90%. While the power using t test is 61.5%, much smaller than the z test. Power using t test in the example C.1 is also much smaller than that in Section 4.1, which is caused by the greater ICC due to patients' repeated measure across time and drop-out at the end of the study.

```
%CRTFASTGEEPWR(alpha = 0.05, m = %str(J(6, 1, 1)), corr_type = BE,
  alpha1 = 0.03, alpha2 = 0.015, alpha3 = 0.2,
  intervention_effect_type = INC, delta = 10,
  period_effect_type = LIN, max_intervention_period = 10,
  beta_period_effects = %str({68, 0.1}), dist = normal, phi = 64,
  CP_size_matrix = %str({4 4 4 4 4 0 0 4 4 4 4 4 4 4 4 4 3 3 0 0 0 0 0,
                         0 4 4 4 4 4 4 0 0 4 4 4 4 4 4 4 4 3 3 0 0 0 0,
                         0 0 4 4 4 4 4 4 4 0 0 4 4 4 4 4 4 4 3 3 0 0 0,
                         0 0 0 4 4 4 4 4 4 4 4 0 0 4 4 4 4 4 3 3 0 0,
                         0 0 0 0 4 4 4 4 4 4 4 4 4 0 0 4 4 4 4 3 3 0,
                         0 0 0 0 0 4 4 4 4 4 4 4 4 4 4 0 0 4 4 4 3 3}),
  DesignPattern = %str({ 0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2,
                         2 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 1 1 2 2 2 2,
                         2 2 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 1 2 2 2,
                         2 2 2 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 2 2,
                         2 2 2 2 0 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 2,
                         2 2 2 2 2 0 0 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1}));
```

```
 The fast GEE power of normal outcomes with block exchangeable correlation
        structure and (alpha1,alpha2,alpha3):(0.03, 0.015, 0.2)
        Under incremental intervention effects model and delta = 10
```

T	S	clusters	df	theta	totaln	Dist	LINK	stddel	zpower	tpower
22	6	6	3	68	348	NORMAL	IDENTITY	3.5025	0.9385	0.615
				0.1						
				10						

C.2. Calculate power for a parallel cluster trial with cross-sectional design

We illustrate the use of **SAS** macro **CRTFASTGEEPWR** for a parallel cluster trial, based on the Enforcing Underage Drinking Laws (EUDL) Program (Preisser *et al.* 2003) in the example. The EDUL program funded interventions at the community level to enforce laws related to alcohol use by underage person to reduce the underage drinking. Moreover, the study used a non-randomized trial design because the intervention communities were selected by the administrative units in states. The control communities were selected by the propensity score method to match the intervention communities based on US census data. There are three periods: one baseline assessment and two follow-up assessments for participants in the communities participating in the EUDL program. We will use the design of the EUDL study to calculate the power under the assumption that all confounded covariates were balanced in control and intervention groups. The main outcome is the binary outcome of self-reported last 30-day alcohol use for an underage person. We assume there are 40 clusters in total with 20 clusters per intervention group and 30 participants enrolled in each cluster-period. Assuming the baseline probability of self-reported last 30-day alcohol use for an underage person is 0.6, we set $\beta_0 = \log(0.6/0.4) = 0.405$ with an consistent decreasing period effects $\beta_1 = -0.01$. Under the average intervention effects model, the intervention effect is assumed to decrease the odds of underage drinking by 30% on average, $\delta = \log(0.7) = -0.357$. Moreover, a nested exchangeable correlation structure is used with ICCs $(\alpha_1, \alpha_2) = (0.02, 0.01)$. From the power calculation results, power is close to 90% given the parameters. Thus, this example further illustrates the flexibility of the **SAS** macro in calculating power for multi-period cluster randomized trials with different designs.

```
%CRTFASTGEEPWR(alpha = 0.05, m = %str(J(2, 1, 20)), corr_type = NE,
  alpha1 = 0.02, alpha2 = 0.01 ,intervention_effect_type = AVE,
  delta = -0.357, period_effect_type = CAT,
  beta_period_effects = %str({0.405, -0.01, -0.01}), dist = binary,
  phi=1, CP_size_matrix =  %str(J(2, 3, 30)),
  DesignPattern =%str({0 1 1,
                       0 0 0}));


 The fast GEE power of binary outcomes with nested exchangeable correlation
               structure and (alpha1,alpha2):(0.02, 0.01)
       Under average intervention effects model and delta = -0.357


T S  clusters   df  theta  totaln    Dist     Link  stddel  zpower  tpower
3 2        40   36  0.405    3600  BINARY    LOGIT  3.2624  0.9036  0.8875
                    -0.01
                    -0.01
                    -0.357
```

We provide **SAS** codes to calculate and compare powers using GEE analysis bases on different effect sizes. In the example codes, **CRTFASTGEEPWR** is used to calculate powers under varying effect sizes, reducing the odds of underage drinking in the EDUL study by $(20\%, 25\%, 30\%, 35\%, 40\%)$ on average, leading to $\delta = \log(0.20, 0.25, 0.30, 0.35, 0.40) = (-0.223, -0.288, -0.357, -0.431, -0.511)$. Outputs of the **SAS** codes are attached below the codes.

```
%macro multi_effectsizes(effectsizes);
  %local index value;
  %do index = 1 %to %sysfunc(countw(&effectsizes,%str( )));
  %let value =%scan(&effectsizes,&index,%str( ));
    %CRTFASTGEEPWR(alpha=0.05, m =%str(J(2,1,20)), corr_type = NE,
      alpha1 = 0.02, alpha2 = 0.01, intervention_effect_type = AVE,
      delta = &value, period_effect_type = CAT,
      beta_period_effects = %str({0.405,-0.01,-0.01}), dist = binary,
      phi = 1, CP_size_matrix = %str(J(2, 3, 30)),
      DesignPattern = %str({0 1 1,
                            0 0 0}));
  %end;
%mend;
%multi_effectsizes(-0.223  -0.288 -0.357  -0.431  -0.511);
```

```
 The fast GEE power of binary outcomes with nested exchangeable correlation
              structure and (alpha1,alpha2):(0.02, 0.01)
         Under average intervention effects model and delta = -0.223


T  S  clusters   df  theta  totaln    Dist     Link  stddel  zpower tpower
3  2        40   36  0.405    3600  BINARY    LOGIT  2.0482  0.5352  0.508
                        -0.01
                        -0.01
                        -0.223


 The fast GEE power of binary outcomes with nested exchangeable correlation
              structure and (alpha1,alpha2):(0.02, 0.01)
         Under average intervention effects model and delta = -0.288


T  S  clusters   df  theta  totaln    Dist     Link  stddel  zpower tpower
3  2        40   36  0.405    3600  BINARY    LOGIT  2.6395  0.7516 0.7276
                        -0.01
                        -0.01
                        -0.288


 The fast GEE power of binary outcomes with nested exchangeable correlation
              structure and (alpha1,alpha2):(0.02, 0.01)
         Under average intervention effects model and delta = -0.357


T  S  clusters   df  theta  totaln    Dist     Link  stddel  zpower tpower
3  2        40   36  0.405    3600  BINARY    LOGIT  3.2624  0.9036 0.8875
                        -0.01
                        -0.01
                        -0.357


 The fast GEE power of binary outcomes with nested exchangeable correlation
              structure and (alpha1,alpha2):(0.02, 0.01)
```

```
        Under average intervention effects model and delta = -0.431

T  S  clusters   df  theta  totaln   Dist    Link  stddel  zpower tpower
3  2        40   36  0.405    3600  BINARY   LOGIT  3.9239  0.9752 0.967
                       -0.01
                       -0.01
                      -0.431


 The fast GEE power of binary outcomes with nested exchangeable correlation
                 structure and (alpha1,alpha2):(0.02, 0.01)
        Under average intervention effects model and delta = -0.511

T  S  clusters   df  theta  totaln   Dist    Link  stddel  zpower tpower
3  2        40   36  0.405    3600  BINARY   LOGIT  4.6296  0.9962 0.9933
                       -0.01
                       -0.01
                      -0.511
```

Affiliation:

Ying Zhang, John S. Preisser
Department of Biostatistics
The University of North Carolina at Chapel Hill
Chapel Hill, NC, United States of America
E-mail: zyingunc@gmail.com

Fan Li
Department of Biostatistics
and
Center for Methods in Implementation and Prevention Science
Yale School of Public Health
New Haven, CT, United States of America

Elizabeth L. Turner
Department of Biostatistics and Bioinformatics
Duke Global Health Institute
Duke University
Durham, NC, United States of America

Paul J. Rathouz
Department of Population Health
The University of Texas at Austin
Austin, TX, United States of America