Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *doi: 10.18637/jss.v000.i00*

swdpwr: A SAS Macro and An R Package for Power Calculation in Stepped Wedge Cluster Randomized Trials

Jiachen Chen
Yale University

Xin Zhou
Yale University

Fan Li
Yale University

Donna Spiegelman
Yale University

Abstract

Conditional models and marginal models can both be utilized to account for the power calculations for stepped wedge cluster randomized trials (CRTs). Hussey and Hughes (2007) obtained the design and analysis of this particular type of CRT based on the linear mixed effects model when the outcome is continuous, along with an approximation of this approach for binary outcomes, which was widely implemented but not accurate for binary outcomes. To improve the approximation method for binary outcomes, Zhou, Liao, Kunz, Normand, Wang, and Spiegelman (2020) and Li, Turner, and Preisser (2018b) have recently proposed two new methods for stepped wedge designs (SWDs) with binary outcomes. However, these new methods have not been implemented in publicly available software such as SAS and R. In this work, we have developed a SAS macro `swdpwr` and an R package **swdpwr** to calculate power in SWDs with various settings of both continuous and binary outcomes.

Keywords: sample size estimation, cross-sectional designs, cohort designs, correlation structure, generalized estimating equations, generalized linear mixed models.

1. Introduction

In cluster randomized trials (CRTs), the unit of randomization is the cluster, which can improve administrative convenience and reduce treatment contamination (Murray *et al.* 1998). Traditional clustered designs such as the parallel design and the crossover design may be susceptible to ethical concerns, because not all clusters receive the intervention before the end of the study (Turner, Li, Gallis, Prague, and Murray 2017). In contrast, in stepped wedge designs (SWDs), all clusters start out in the control condition and switch to the intervention condition in a unidirectional and randomly assigned order, and once treated, the clusters

maintain their intervention status until the end of the study. At pre-specified time periods, a random subset of clusters cross over from the control to the intervention condition. Stepped wedge randomization may be preferred for estimating intervention effects when it is logistically more convenient to roll-out intervention in a staggered fashion and when the stakeholders or participating clusters perceive the intervention to be beneficial to the target population (Hussey and Hughes 2007).

Two different SWD sampling schemes have been proposed: the cross-sectional design and the closed cohort design (Copas, Lewis, Thompson, Davey, Baio, and Hargreaves 2015). In a cross-sectional design, different participants are recruited at each time period in each cluster; while in a closed cohort design (which for simplicity will be referred to as a cohort design hereafter), participants are recruited at the beginning of the study and have repeated measures at different time periods (Hemming and Girling 2013). A distinguishing feature of CRTs is that outcomes within the same cluster tend to be correlated with one another (Martin, Girling, Nirantharakumar, Ryan, Marshall, and Hemming 2016a). Because in SWDs outcomes are measured at different time periods, the within-period and inter-period correlation coefficients are likely different and thus should be separately considered in designing SWDs. An additional within-individual correlation should be included when it is a cohort SWD to account for the repeated measures of the same individual over time (Hughes, Granston, and Heagerty 2015). Two statistical models can be used to account for these three levels of intraclass correlation: the conditional model and the marginal model. Conditional models are based on mixed effects models (Pinheiro and Bates 2006; Breslow and Clayton 1993), which accommodate the intraclass correlations via latent random effects. Marginal models describe the population-averaged response values across cluster-periods, and are usually fitted with the generalized estimating equations (GEE) (Liang and Zeger 1986). The interpretations of regression parameters can be different under these two models, with the important exception of the identity and log links when random effects and the covariates are independent, as is typically assumed (Ritz and Spiegelman 2004). The design and analysis of SWDs have been mostly based on conditional models, for instance, Hussey and Hughes (2007); Woertman, de Hoop, Moerbeek, Zuidema, Gerritsen, and Teerenstra (2013); Hemming, Lilford, and Girling (2015); Hooper, Teerenstra, de Hoop, and Eldridge (2016); Li, Turner, and Preisser (2018a). As marginal models carry a straightforward population-averaged interpretation assuming equivalence of study random effects with target population random effects, Li *et al.* (2018b) proposed methods for the design and analysis of SWDs using marginal models.

Binary outcomes are frequently seen in cluster randomized trials as endpoints. However, existing methods for sample size calculation of SWDs have been almost exclusively focused on continuous outcomes. Hussey and Hughes (2007) proposed an approach based on linear mixed effects models and estimated by weighted least squares for continuous outcomes, and obtained an approximation of this approach for binary outcomes. Systematic reviews indicated that the majority of SWDs with binary outcomes used this approximation method (Hemming and Taljaard 2016; Martin, Taljaard, Girling, and Hemming 2016b), which may either overestimate or underestimate the power in different scenarios (Zhou *et al.* 2020). To improve this approximation, Zhou *et al.* (2020) developed a maximum likelihood method for power calculations of SWDs with binary outcomes based on the mixed effects model and Li *et al.* (2018b) proposed a method for binary outcomes within the framework of GEE that employed a block exchangeable within-cluster correlation structure with three correlation parameters.

These new methods have been recently proposed in the statistical literature, and have not

yet been implemented in publicly available software such as SAS and R, making it difficult for applied researchers to apply these new methods to design their studies in a rigorous fashion. Additionally, existing software for SWDs focused on limited settings and did not have accurate power calculations for binary outcomes. For example, Hemming and Girling (2014) developed a Stata menu-driven program `steppedwedge` based on Hussey and Hughes (2007), and the **swCRTdesign** (Hughes, Hakhu, and Voldal 2019) package in R as well as a spreadsheet (`http://faculty.washington.edu/jphughes/pubs.html`) was published to implement Hussey and Hughes (2007), both of which utilized the linear mixed effects model for continuous outcomes and binary proportions considering only the within-period correlation for cross-sectional designs. Hemming, Kasza, Hooper, Forbes, and Taljaard (2020) proposed the Shiny CRT Calculator programmed in R (`https://github.com/karlahemming/Cluster-RCT-Sample-Size-Calculator`) using linear mixed effects models with three intraclass correlations that accommodated cross-sectional and cohort designs (Hooper *et al.* 2016) for continuous outcomes and also regarded binary outcomes as proportions to fit into the linear mixed effects model for approximation. All the existing software did not implement accurate power calculations for SWDs with binary outcomes. In an effort to make the new methods more accessible and incorporate as various settings of SWDs as possible, we have developed user-friendly software based on the methods proposed by Zhou *et al.* (2020) and Li *et al.* (2018b) to implement power calculations for SWDs with binary outcomes as well as continuous outcomes. The core of the software is developed in Fortran and is built into a SAS macro `swdpwr` and an R package **swdpwr**, respectively.

2. Methods

Throughout this article, the regression parameter β encodes the treatment effect. For testing the treatment effect, we consider the following hypothesis:

$$\mathrm{H}_0 : \beta = 0 \qquad \text{versus} \qquad \mathrm{H}_A : \beta \neq 0, \tag{1}$$

where β_A is the true value of β under the alternative hypothesis that $\beta_A \neq 0$. In this software, power is calculated based on a two-sided Wald-type test given by:

$$\Phi\left(\frac{|\beta_A|}{\sqrt{\mathrm{Var}(\hat{\beta})}} - Z_{1-\alpha/2}\right) + \Phi\left(-\frac{|\beta_A|}{\sqrt{\mathrm{Var}(\hat{\beta})}} - Z_{1-\alpha/2}\right), \tag{2}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, α is the significance level, and $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution. The variance of $\hat{\beta}$ is either defined by asymptotic theory for maximum likelihood estimation (MLE) in a conditional model, or by the theory of estimating equations in the marginal model.

A SWD is defined by I clusters and J time periods, each including K_{ij} individuals at time period j for cluster i (i in $1, \cdots, I$; j in $1, \cdots, J$). For a cross-sectional SWD, the size of cluster i is given by $N_i = \sum_{j=1}^{J} K_{ij}$; for a cohort SWD, assuming that $K_{ij} = K_i$ for the whole study period, the size of cluster i is $N_i = K_i$. Let Y_{ijk} be the response corresponding to individual k at time period j from cluster i (k in $1, \cdots, K_{ij}$), which can be continuous or binary outcomes (for example: success or failure of a surgery, getting the disease or not). For both types of outcomes, we first present the general models with three correlation parameters

and then give more details about the cases implemented in this software. As the design and analysis of SWDs have been mostly based on conditional models, the general models are described by conditional models.

2.1. Models with binary outcomes

Due to the correlated binary outcomes in SWDs, a general model can be fitted utilizing the conditional method (a mixed-effects model):

$$g(p_{ijk}) = \mu + X_{ij}\beta + \gamma_j + b_i + c_{ij} + \pi_{ik} \tag{3}$$

where $g(\cdot)$ is a link function, X_{ij} is a binary treatment assignment (1=intervention; 0=standard of care) in cluster i at time period j, μ is the baseline effect in the control group, γ_j is the fixed time effect corresponding to time period j (j in $1, \cdots, J$, and $\gamma_1 = 0$ for identifiability), and β is the parameter of interest in this study, interpreted as the treatment effect. We assume that b_i is the cluster random effect distributed by $N(0, \sigma_b^2)$, c_{ij} is the cluster-by-time interaction random effect distributed by $N(0, \sigma_c^2)$, and π_{ik} is the random effect for repeated measures of one individual distributed by $N(0, \sigma_\pi^2)$. We also assume that b_i, c_{ij} and π_{ik} are independent of each other. Let $p_{ijk} = P(Y_{ijk} = 1|X_{ij}, b_i, c_{ij}, \pi_{ik})$ be the probability of the outcome for individual k conditioned on the random effects and design allocation.

To account for the correlation of outcomes in each cluster, a correlation structure with three levels of correlation is employed: (1) α_0, the within-period correlation, which measures the similarity between responses from different individuals within the same cluster during the same time period ($\text{corr}(Y_{ijk}, Y_{ijk'}) = \alpha_0$ for $k \neq k'$); (2) α_1, the inter-period correlation, which measures the similarity between responses from different individuals within the same cluster but across time periods ($\text{corr}(Y_{ijk}, Y_{ij'k'}) = \alpha_1$ for $j \neq j', k \neq k'$); (3) α_2, the within-individual correlation, which measures the similarity between responses from the same individual across time periods ($\text{corr}(Y_{ijk}, Y_{ij'k}) = \alpha_2$ for $j \neq j'$).

This framework can accommodate three link functions $g(\cdot)$: identity, log and logit links. In a cross-sectional design, the correlation structure may depend only on α_0 and α_1 since usually different sets of individuals are assessed for each cluster at different time periods, and α_2 is no longer required. However, in this circumstance, α_1 and α_2 just accord with each other based on their definitions as actually different individuals are assessed at each time period for cross-sectional settings. In other words, we can equate $\alpha_2 = \alpha_1$ to obtain a two-level exchangeable correlation model to accommodate the cross-sectional design (Li *et al.* 2018b). Otherwise, it is a cohort design.

This conditional model 3 accommodates cases of both cross-sectional and cohort SWDs. These general cases can also be fitted by the marginal model with the same correlation structure but without terms of random effects. In our software, we implemented specific cases with binary outcomes under both the conditional and the marginal model:

Conditional model based on GLMM Under the conditional model, Zhou *et al.* (2020) considered a cross-sectional SWD which is modelled by a generalized linear mixed model (GLMM) with similar notations of the general model:

$$g(p_{ijk}) = \mu + X_{ij}\beta + \gamma_j + b_i, \tag{4}$$

where $g(\cdot)$ is a link function, X_{ij} is a binary treatment assignment (1=intervention; 0=standard of care) in cluster i at time period j, μ is the baseline effect in the control group, γ_j is the fixed time effect corresponding to time period j (j in $1, \cdots, J$, and $\gamma_1 = 0$ for identifiability), and b_i is the random effect for cluster i; β is the parameter of interest in this study, interpreted as the treatment effect; $p_{ijk} = P(Y_{ijk} = 1|X_{ij}, b_i)$ is the probability of the outcome for individual k conditioned on the cluster random effect and design allocation at time period j from cluster i, interpreted as the conditional mean response of individuals. We assume a normal distribution for random effects, $b_i \sim N(0, \tau^2)$, although Zhou *et al.* (2020) showed that results are relatively insensitive to the assumed form of the distribution of the random effects. The correlation structure for this case is $\alpha_0 = \alpha_1 = \alpha_2 = \frac{Var(E(Y_{ij}|X_{ij}=0,b_i))}{Var(Y_{ij}|X_{ij}=0)}$. Particularly, under identity link: $\alpha_0 = \alpha_1 = \alpha_2 = \frac{\tau^2}{\tau^2+\mu(1-\mu)}$ (Hussey and Hughes 2007).

According to model (4), the outcome for individual k at time period j from cluster i follows a Bernoulli distribution with the probability of

$$p_{ijk} = g^{-1}(\mu + X_{ij}\beta + \gamma_j + b_i), \tag{5}$$

where $g^{-1}(\cdot)$ is the inverse link function. In this work, we consider three link functions for $g(\cdot)$: identity, log and logit. Then we derive the likelihood of the observed outcomes based on the conditional probability under the specific link function. Because the likelihood involves integrating the unobserved random effects b_i, we use the Gaussian quadrature for numerical integration (McCulloch 1997). The calculation of the large-sample variance and power is then based on the theory of MLE.

We also consider cases of SWDs with no time effects (all $\gamma_j = 0$). The model is

$$g(p_{ijk}) = \mu + X_{ij}\beta + b_i. \tag{6}$$

This model is similar to the one considered in Zhou, Liao, and Spiegelman (2017), and is relevant when the time effects are expected to be minimum. The derivations of the likelihood formula and the calculation for variance and power are similar to the procedures with time effects.

Marginal model based on GEE The marginal model can also account for cohort SWDs with with three correlation parameters, where individuals from each cluster are enrolled at the start of the trial and followed up for repeated measurements (Li *et al.* 2018b). With a slight abuse of notations, the marginal model based on GEE is:

$$g(p_{ijk}) = \mu + X_{ij}\beta + \gamma_j. \tag{7}$$

Here we denote $p_{ijk} = P(Y_{ijk} = 1|X_{ij})$, interpreted as the marginal mean response of individuals; μ is the baseline effect in the control group, γ_j is the fixed time effect; β is the parameter of interest in this study, interpreted as the marginal treatment effect. We also consider two settings: with time effects (j in $1, \cdots, J$, and $\gamma_1 = 0$ for identifiability) and without time effects (all $\gamma_j = 0$). When assuming $K_{ij} = K_i$ for each j, the block exchangeable working correlation matrix for cluster i is written as (Li *et al.* 2018b):

$$\mathrm{R}_i = (1 - \alpha_0 + \alpha_1 - \alpha_2)\mathrm{I}_{JK_i} + (\alpha_2 - \alpha_1)\mathrm{J}_J \otimes \mathrm{I}_{K_i} + (\alpha_0 - \alpha_1)\mathrm{I}_J \otimes \mathrm{J}_{K_i} + \alpha_1 J_{JK_i}. \tag{8}$$

where J_u is a $u*u$ matrix with all elements of 1, I_k is a $k*k$ identity matrix. Li *et al.* (2018b) showed that R_i has four distinct eigenvalues and all of them should have positive values, in order to ensure a positive-definite correlation structure. We assume $\eta = (\mu, \gamma_2, \gamma_3, ..., \gamma_J, \beta)'$ to be the vector of parameters in model 7, and $\eta = (\mu, \beta)'$ to be the parameter vector without time effects. Let $Y_i = (Y_{i11}, Y_{i12}..., Y_{iJK_i})'$ and $p_i = (p_{i11}, p_{i12}..., p_{iJK_i})'$. The GEE estimator $\hat{\eta}$ is obtained by solving $\sum_i \mathrm{D}_i' \mathrm{V}_i{}^{-1}(Y_i - p_i) = 0$, where $\mathrm{D}_i = \partial p_i / \partial \eta'$, $\mathrm{V}_i = {A_i}^{1/2} R_i {A_i}^{1/2}$, and A_i is the JK_i-dimensional diagonal matrix with elements $\phi v(p_{ijk})$, with ϕ representing the dispersion parameter ($\phi = 1$ for binary outcomes) and the variance function $v(p_{ijk}) = p_{ijk}(1 - p_{ijk})$. Assuming the correlation structure is correctly specified, $\hat{\eta}$ is approximately multivariate normal with mean η and covariance estimated by the model-based estimator $(\sum_i \mathrm{D}_i'(\hat{\eta}) \mathrm{V}_i^{-1}(\alpha) \mathrm{D}_i(\hat{\eta}))^{-1}$. Hence, we can calculate the variance of $\hat{\beta}$.

Cross-sectional SWDs can also be designed under the marginal method with appropriate specification of correlation parameters $\alpha_1 = \alpha_2$. Then individuals enrolled in each cluster at each time period will be different.

For binary outcomes, the marginal means limit the ranges of correlation (Qaqish 2003; Ridout, Demetrio, and Firth 1999), thus we need additional restrictions for $\alpha_0, \alpha_1, \alpha_2$ beyond those required to ensure a positive definite R_i and that the correlations are between 0 and 1. To ensure valid probability between 0 and 1 under different link functions, we also need restrictions for parameters including treatment effect, baseline effect and time effects. For example, $\mu + \beta + \gamma_j$ should not exceed 0 under log link function. When the input parameters for power calculations are out of range, an error message will return in the software.

2.2. Models with continuous outcomes

Hooper *et al.* (2016) assumed a linear mixed effects model for the correlated continuous outcomes:

$$Y_{ijk} = \mu + X_{ij}\beta + \gamma_j + b_i + c_{ij} + \pi_{ik} + \epsilon_{ijk} \tag{9}$$

where X_{ij} is a binary treatment assignment (1=intervention; 0=standard of care) in cluster i at time period j, β is the treatment effect, μ is the baseline effect in control groups, γ_j is the fixed time effect corresponding to time period j (with $\gamma_1 = 0$). We assume that b_i is the cluster random effect distributed by $N(0, \sigma_b^2)$, c_{ij} is the cluster-by-time interaction random effect distributed by $N(0, \sigma_c^2)$, π_{ik} is the random effect for repeated measures of one individual distributed by $N(0, \sigma_\pi^2)$, and $\epsilon_{ijk} \sim N(0, \sigma_e^2)$. We also assume that b_i, c_{ij}, π_{ik} and ϵ_{ijk} are independent of each other, and the total variance of Y_{ijk} is $\sigma_t^2 = \sigma_b^2 + \sigma_c^2 + \sigma_\pi^2 + \sigma_e^2$. To account for the correlation of outcomes in each cluster, a correlation structure with three levels are employed: (1) α_0, the within-period correlation, which measures the similarity between responses from different individuals within the same cluster during the same time period ($\mathrm{corr}(Y_{ijk}, Y_{ijk'}) = \alpha_0 = \frac{(\sigma_b^2 + \sigma_c^2)}{\sigma_t^2}$ for $k \neq k'$); (2) α_1, the inter-period correlation, which measures the similarity between responses from different individuals within the same cluster but across time periods ($\mathrm{corr}(Y_{ijk}, Y_{ij'k'}) = \alpha_1 = \frac{\sigma_b^2}{\sigma_t^2}$ for $j \neq j', k \neq k'$); (3) α_2, the within-individual correlation, which measures the similarity between responses from the same individual across time periods ($\mathrm{corr}(Y_{ijk}, Y_{ij'k}) = \alpha_2 = \frac{(\sigma_b^2 + \sigma_\pi^2)}{\sigma_t^2}$ for $j \neq j'$). The correlation structure R_i and the specification for cross-sectional and cohort designs are the same with those in Section 2.1.

This general model 9 is described at individual-level responses based on mixed effects model, which agrees with the population-averaged marginal model in Li *et al.* (2018b):

$$\mu_{ijk} = \mu + X_{ij}\beta + \gamma_j \tag{10}$$

where μ_{ijk} is the marginal mean response of Y_{ijk}. We also consider two settings: with time effects (j in $1, \cdots, J$, and $\gamma_1 = 0$ for identifiability) and without time effects (all $\gamma_j = 0$). We assume $\eta = (\mu, \gamma_2, \gamma_3, ..., \gamma_J, \beta)'$ to be the vector of parameters in model 10, and $\eta = (\mu, \beta)'$ to be the parameter vector without time effects.

Li *et al.* (2018b) and Hooper *et al.* (2016) are proposed in the same scenarios that accommodate three correlation parameters, under marginal method and conditional method, respectively. In both models, the covariance is estimated by the model-based estimator $(\sum_i \mathbf{Z_i'V_i^{-1}Z_i})^{-1}$, where $\mathbf{Z_i}$ is the $J * (J+1)$ design matrix corresponds to the parameter vector η under the SWD framework within cluster i and $\mathbf{V_i}$ is the working covariance matrix within cluster i based on a block exchangeable correlation structure, in which $\mathbf{V_i} = \sigma_t^2 R_i$. Our software implements Li *et al.* (2018b) for cross-sectional and cohort SWDs with continuous outcomes.

2.3. Remarks

This software includes procedures of power calculations for SWDs with binary outcomes under identity, logit and log link functions based on Zhou *et al.* (2020) and Li *et al.* (2018b), and also incorporates procedures for continuous outcomes as in Hussey and Hughes (2007), Hooper *et al.* (2016) and Li *et al.* (2018b) with the identity link function. In Appendix A, by proving the equivalence of variance for intervention effect in these models for continuous outcomes, we showed that Li *et al.* (2018b) incorporated the cross-sectional scenarios in Hussey and Hughes (2007) and the cohort scenarios in Hooper *et al.* (2016), and obtained the same power formulae under corresponding scenarios. Hence, we can directly implement just Li *et al.* (2018b) in our software for different scenarios of continuous outcomes.

Regarding the cross-sectional settings under $\alpha_0 = \alpha_1 = \alpha_2$ with binary outcomes, we provide both the conditional (Zhou *et al.* 2020) and the marginal (Li *et al.* 2018b) method. Zhou *et al.* (2020) utilized the maximum likelihood estimator under GLMM and Li *et al.* (2018b) used the generalized least squares (GLS) estimator for GEE, of which the variance for these two estimates are different. As in Ritz and Spiegelman (2004), the treatment effect β is the same for these two methods under identity and log link functions. We observe that the power obtained by the conditional method is uniformly larger than that of the marginal method given the same setting and parameters under the identity and log links for scenarios with and without time effects, except when the treatment effect is zero. This pattern is expected because GLMM is likelihood based, while marginal models rely on quasi-likelihood and unbiased estimating equations.

We also need to emphasize that our software considers both cases with and without time effects. In most literature, see, for instance, Hussey and Hughes (2007), Hemming *et al.* (2015), Hooper *et al.* (2016), and Li *et al.* (2018a,b), they derived their method assuming the existence of time effects However, Zhou *et al.* (2017) argued that SWDs are mostly used in the study of relatively short-term outcomes with short-term interventions as well, thus it is reasonable to assume no time effects in the primary analysis and Zhou *et al.* (2020) proposed their method considering cases with and without time effects. When developing this software,

		Conditional method	Marginal method
Binary outcomes (identity, log, logit links)	Cross-sectional design	Zhou *et al.* (2020)	Li *et al.* (2018b)
	Cohort design	Under development	Li *et al.* (2018b)
Continuous outcomes (identity link)	Cross-sectional design	Li *et al.* (2018b); Hussey and Hughes (2007); Zhou *et al.* (2017)	
	Cohort design	Li *et al.* (2018a,b); Hooper *et al.* (2016)	

Table 1: List of all the methods implemented in **swdpwr**

we include and generalize all the methods to cases with and without time effects. Here we list in Table 1 all the scenarios that are implemented in the software, accommodating cases and methods with and without time effects.

3. Illustrations of the software

As the software accommodates power calculations of SWDs under various settings, each of the arguments might include multiple specifications. Both the marginal model (GEE) and conditional model (GLMM) can be adopted for calculations, including three link functions: linear, log and logit link. The input arguments are the same for R and SAS, except that the data preparation procedures are different for platforms. Different arguments (for example, time effects, baseline effect in control groups, treatment effect, study design, intraclass correlations (ICCs), etc.) can be specified based on the preliminary information obtained in practice.

In this section, we give a detailed description for the implementation in R and SAS, respectively. Examples for using this software under different settings will be discussed in Section 4. We assume balanced cluster-period sizes so that $K_{ij} = K$ for different i and j in the current version of the software. To account for time effects, we assume that the difference of time effects between neighbouring time periods is constant, which indicates that the absolute value of γ_j increases by j uniformly. Hence, the input parameter for time effects γ_J will include all information about time effect at each time period.

3.1. Description of the function in R package

This package **swdpwr** meets the need of statistical power calculation for SWDs with both binary and continuous outcomes through the `swdpower` function. The arguments of `swdpower` are

```
swdpower(I, J, K, dataset, response = 2, model = 2, link = 1, mu, beta,
+ gammaJ = 0, sigma2 = 0, alpha = 0.05, ICC0 = 0.1, ICC1 = ICC0/2, ICC2)
```

with details provided in Table 2 along with defaults if any. The input argument `dataset` is a

matrix generated in R with elements of 0 (control) or 1 (intervention) for users to define the detailed study design, in which each row represents a cluster and each column represents a time period. The input arguments `I` and `J` should correspond to the correct number of rows and columns of the matrix in the generated data set. The argument `sigma2` is a nuisance parameter for binary outcomes but is necessary for continuous outcomes. The objects returned by `swdpower` function are the design matrix, summary features of the design as well as the power in this scenario.

3.2. Description of the SAS macro

The SAS macro `swdpwr` plays the same role as the function `swdpower` in R package **swdpwr** described above, and this macro also needs a pre-prepared data set of the study design matrix generated in SAS. For illustration, a toy data set `design` is created, which contains the allocation of control (0) or intervention (1) for each cluster at different time periods as well as the number of clusters for each allocation shown the first column. The step to generate this data set in SAS is as follows, in which each type of allocation is conducted in 3 clusters.

```
data design;
input numofclusters time1 time2 time3 time4;
cards;
3 0 1 1 1
3 0 0 1 1
;
run;
```

The procedure of data set generation is the common `DATA` step, but is supposed to include a header with correct time period description.

Sharing the same utility with the function `swdpower` in R package **swdpwr**, the input arguments of `swdpwr` are

```
% macro swdpwr(I = , J = , K = ,  dataset = , response = 2 , model = 2,
+ link = 1 , mu = , beta = , gammaJ = 0, sigma2 = 0, alpha = 0.05,
+ ICC0 = 0.1, ICC1 = ICC0/2 , ICC2 = );
```

with details given in Table 2 together with defaults if any. The input arguments `I` and `J` should correspond to the correct number of rows and columns of the matrix in the generated data set. This macro will return an output of summary features of the design, as well as the power value under the particular scenario.

4. Examples

The usage of the software is based on platforms of R and SAS, which requires separate illustrations. The following sections are organized according to different scenarios such as continuous and binary outcomes, cross-sectional and cohort settings, different model options, different link functions, different time effects assumptions, etc. Each section will contain examples under both platforms.

Argument	Description	Default
I	number of clusters	
J	number of time periods	
K	number of individuals at each time period in a cluster	
dataset	Data set that describes the study design	
response	1 if continuous outcomes, 2 if binary outcomes	2
model	1 if using conditional method, 2 if using marginal method	2
link	1 for identity link, 2 for log link, 3 for logit link	1
mu	baseline effect in control groups	
beta	treatment effect (the parameter we would like to test)	
gammaJ	time effect at time period J	0
sigma2	marginal variance of the outcome (only by continuous outcomes)	0
alpha	Type I error	0.05
ICC0	within-period correlation α_0	0.1
ICC1	inter-period correlation α_1	ICC0/2
ICC2	within-individual correlation α_2	

Table 2: Arguments of software swdpwr

4.1. Conditional method with binary outcomes for cross-sectional designs

Zhou *et al.* (2020) proposed this maximum likelihood method under cross-sectional settings, which discussed scenarios with time effects and without time effects. The method was firstly developed under the identity link function for binary outcomes, and can be generalized to logit link and log link functions in the same mechanism. As with binary outcomes, the marginal variance argument sigma2 becomes a nuisance parameter. Due to this specific cross-sectional setting, we specify ICC0 = ICC1 = ICC2 in the function as in Section 2.1.

We use a simulated example to explain the specification of parameters and the use of the software. Fitting the model with simulated input in R:

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1),6),rep(c(0,0,1),6)),12,3,byrow=TRUE)
R> swdpower(I = 12, J = 3, K = 50, dataset, response = 2, model = 1, link = 3,
mu = -0.9, beta= 0.5, gammaJ= 0.2, alpha = 0.05, ICC0 = 0.01, ICC1 = 0.01,
ICC2 = 0.01)
$`design matrix dataset: row-cluster col-time`
      [,1] [,2] [,3]
 [1,]    0    1    1
 [2,]    0    1    1
 [3,]    0    1    1
 [4,]    0    1    1
 [5,]    0    1    1
 [6,]    0    1    1
 [7,]    0    0    1
 [8,]    0    0    1
 [9,]    0    0    1
[10,]    0    0    1
```

```
[11,]    0    0    1
[12,]    0    0    1

$Summary
                          [,1]
I                       12.000
J                        3.000
K                       50.000
total sample size     1800.000
response                 2.000
model                    1.000
link                     3.000
baseline effect mu      -0.900
treatment effect beta    0.500
time effect gamma_J      0.200
ICC0                     0.010
ICC1                     0.010
ICC2                     0.010
Type I error             0.050
Power                    0.735
```

In the example above, 50 individuals are included in each cluster at each time period, with 12 clusters and 3 time periods under the SWD (a total sample size of 1800). The calculation is conducted utilizing the conditional method under the logit link function with time effect set to be 0.2. `ICCs`, here interpreted as the correlation between different individuals within the same cluster, are all set as 0.01 and the Type I error is 0.05. The power obtained from `swdpower` for this scenario is 0.735 for the alternative hypothesis $\beta_A = 0.5$.

The model can also be established under other link functions such as identity link and log link:

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1),6),rep(c(0,0,1),6)),12,3,byrow=TRUE)
R> swdpower(I = 12, J = 3, K = 50, dataset, response = 2, model = 1, link = 1,
mu = 0.3, beta = 0.1, gammaJ = 0, alpha = 0.05, ICC0 = 0.02, ICC1 = 0.02,
ICC2 = 0.02)
$`design matrix dataset: row-cluster col-time`
      [,1] [,2] [,3]
 [1,]    0    1    1
 [2,]    0    1    1
 [3,]    0    1    1
 [4,]    0    1    1
 [5,]    0    1    1
 [6,]    0    1    1
 [7,]    0    0    1
 [8,]    0    0    1
 [9,]    0    0    1
[10,]    0    0    1
```

```
[11,]    0    0    1
[12,]    0    0    1

$Summary
                          [,1]
I                       12.000
J                        3.000
K                       50.000
total sample size     1800.000
response                 2.000
model                    1.000
link                     1.000
baseline effect mu       0.300
treatment effect beta    0.100
time effect gamma_J      0.000
ICC0                     0.020
ICC1                     0.020
ICC2                     0.020
Type I error             0.050
Power                    0.991
```

Different from the previous one, this example fits the conditional method using identity link without time effects. The `ICCs` are 0.02 and the Type I error is 0.05. Calculated from these assumptions, R package gives the power of 0.991 for the alternative hypothesis $\beta_A = 0.1$.

The design in SAS is conducted in the macro `swdpwr`. Similar to the procedures in R, simulated examples are considered for illustration, along with simulated SAS data sets for specifying the study design. To accomplish the power calculation, generating a SAS data set `exmple` as follows is a pre-requisite:

```
data exmple;
input numofclusters time1 time2 time3 time4;
cards;
6 0 1 1 1
6 0 0 1 1
;
run;
```

The macro called to do power calculations would be:

```
%swdpwr(I = 12, J = 4, K = 30, dataset = example, response = 2, model = 1,
 link = 2, mu = -1.2, beta = 0.5, gammaJ = 0.28,  alpha = 0.05,
 ICC0 = 0.03, ICC1 = 0.03, ICC2 = 0.03)
```

The pre-prepared data set `example` should be properly generated as described in Section 3. In this SWD, 30 individuals are included in each cluster at each time period, with 12 clusters and 4 time periods (a total sample size of 1440). Calculations are based on the conditional

Result
I = 12
J = 4
K = 30
Total sample size = 1440
response = 2
model = 1
link = 2
Baseline effect (mu): -1.200
Treatment effect (beta): 0.500
Time effect (gamma J): 0.280
ICC0: 0.03
ICC1: 0.03
ICC2: 0.03
Type I error = 0.05
Power = 0.975

Table 3: Output from macro `swdpwr`

method under log link function with time effects. Table 3 summarizes the output from the macro `swdpwr`. SAS gives the power of 0.975 under this scenario for the alternative hypothesis $\beta_A = 0.5$.

Similarly, the link functions can choose from identity, log and logit links; the other parameters except the nuisance ones can be updated according to different settings.

4.2. Marginal method with binary outcomes for cross-sectional and cohort designs

A marginal method based on GEE was proposed in Li *et al.* (2018b), which employed a block exchangeable correlation structure that can be applied to both cohort designs and cross-sectional designs. This method also accommodates scenarios with and without time effects as well as three link function options. Due to binary outcomes, the marginal variance argument `sigma2` is a nuisance parameter. This cluster randomized design requires the specification of three-level correlations under different designs. The caveat of this method lies in that the range of correlations is additionally restricted by marginal means because of binary outcomes, and this software will return an error message when the input correlation values are out of their plausible range (Qaqish 2003; Ridout *et al.* 1999).

Before giving the detailed examples, we discuss more about the cross-sectional or cohort designs and the specification for the three levels of correlation. As mentioned in Section 2.1, these three correlations represent the within-period, inter-period, and within-individual correlation respectively. For the correlation structure of SWDs, Li *et al.* (2018a) specified four linear eigenvalue constraints so that the correlation values ensure a positive definite correlation matrix. These constraints are enforced in our software for all models. In a cross-sectional design, since different individuals are considered at different time periods in each cluster, α_2 is not required and the block exchangeable correlation structure reduced to the nested exchangeable structure as in Teerenstra, Lu, Preisser, Van Achterberg, and Borm

(2010), with $\alpha_2 = \alpha_1$. Specifically, a single correlation parameter α_1 is used to describe the correlation structure in Zhou *et al.* (2020) such that the working correlation reduced to $\alpha_0 = \alpha_1 = \alpha_2$.

Additional restrictions for α_0, α_1 and α_2 are needed as the marginal means limit the ranges of correlation due to binary outcomes (Qaqish 2003):

$$max(0, \mu_{ijk} + \mu_{ilm} - 1) \leq E[Y_{ijk}Y_{ilm}] \leq min(\mu_{ijk}, \mu_{ilm}) \tag{11}$$

and within each cluster, the following three cases are checked in Equation 11 for all individuals at each time period, where $j \neq l$ and $k \neq m$:

$$E[Y_{ijk}Y_{ijm}] = \mu_{ij}\mu_{ij} + \alpha_0 v_{ij}^{1/2} v_{ij}^{1/2} \tag{12}$$

$$E[Y_{ijk}Y_{ilm}] = \mu_{ij}\mu_{il} + \alpha_1 v_{ij}^{1/2} v_{il}^{1/2} \tag{13}$$

$$E[Y_{ijk}Y_{ilk}] = \mu_{ij}\mu_{il} + \alpha_2 v_{ij}^{1/2} v_{il}^{1/2} \tag{14}$$

where $\mu_{ijk} = \mu_{ij}$ and $v_{ijk} = v_{ij}$ are the mean and variance of the outcome for individual k at time period j from cluster i.

We first simulate a cohort design example to illustrate the use of the software in R for function `swdpower` in package **swdpwr**:

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1,1),4),rep(c(0,0,1,1),4),rep(c(0,0,0,1),4)),
12, 4, byrow=TRUE)
R> swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 2, link = 3,
mu= 0, beta = 0.405, gammaJ = 0.405, alpha = 0.05, ICC0 = 0.03, ICC1 = 0.015,
ICC2 = 0.2)
$`design matrix dataset: row-cluster col-time`
      [,1] [,2] [,3] [,4]
 [1,]    0    1    1    1
 [2,]    0    1    1    1
 [3,]    0    1    1    1
 [4,]    0    1    1    1
 [5,]    0    0    1    1
 [6,]    0    0    1    1
 [7,]    0    0    1    1
 [8,]    0    0    1    1
 [9,]    0    0    0    1
[10,]    0    0    0    1
[11,]    0    0    0    1
[12,]    0    0    0    1

$Summary
                              [,1]
```

```
I                       12.000
J                        4.000
K                      100.000
total sample size     1200.000
response                 2.000
model                    2.000
link                     3.000
baseline effect mu       0.000
treatment effect beta    0.405
time effect gamma_J      0.405
ICC0                     0.030
ICC1                     0.015
ICC2                     0.200
Type I error             0.050
Power                    0.726
```

This example illustrates a cluster randomized cohort SWD including 100 individuals in each cluster, with 12 clusters and 4 time periods (a total sample size of 1200). We conduct the calculation using the marginal model under the logit link with time effects. The three correlation parameters are 0.03, 0.015 and 0.2 and satisfy all restrictions described previously (positive definite correlation structure, marginal means restrictions and the natural range of ICCs). The power calculated given by R for this scenario is around 0.726 for the alternative hypothesis $\beta_A = 0.405$.

Different link functions can be adapted for the calculation. For instance, similar to the previous example, the calculation is based on the log link function with Type I error of 0.05. The power obtained is then 0.871 for the alternative hypothesis $\beta_A = 0.18$.

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1,1),4),rep(c(0,0,1,1),4),rep(c(0,0,0,1),4)),
12, 4, byrow=TRUE)
R> swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 2, link = 2,
mu = -0.7, beta = 0.18, gammaJ = 0.33, alpha = 0.05, ICC0 = 0.03, ICC1 = 0.015,
ICC2 = 0.2)
$`design matrix dataset: row-cluster col-time`
      [,1] [,2] [,3] [,4]
 [1,]    0    1    1    1
 [2,]    0    1    1    1
 [3,]    0    1    1    1
 [4,]    0    1    1    1
 [5,]    0    0    1    1
 [6,]    0    0    1    1
 [7,]    0    0    1    1
 [8,]    0    0    1    1
 [9,]    0    0    0    1
[10,]    0    0    0    1
[11,]    0    0    0    1
[12,]    0    0    0    1
```

```
$Summary
                          [,1]
I                       12.000
J                        4.000
K                      100.000
total sample size     1200.000
response                 2.000
model                    2.000
link                     2.000
baseline effect mu      -0.700
treatment effect beta    0.180
time effect gamma_J      0.330
ICC0                     0.030
ICC1                     0.015
ICC2                     0.200
Type I error             0.050
Power                    0.871
```

The macro `swdpwr` is utilized for the calculation in SAS. Still, we first illustrate cohort design examples and a SAS data set `design2` is required:

```
data design2;
input numofclusters time1 time2 time3 time4;
cards;
4 0 1 1 1
4 0 0 1 1
4 0 0 0 1
;
run;
```

The macro called to do power calculations would be:

```
%swdpwr(I = 12, J = 4, K = 100, dataset = design2, response = 2, model = 2,
link = 1, mu = 0.5, beta = 0.15, gammaJ = 0, alpha = 0.05,
ICC0 = 0.1, ICC1 = 0.05, ICC2 = 0.2)
```

Based on this simulated stepped wedge cohort trial and conducting design using the marginal method with identity link function without time effects, we obtain the calculated power of 0.985 for the alternative hypothesis $\beta_A = 0.15$ in Table 4.

We also consider the cross-sectional settings under the marginal method, where ICC1 = ICC2 with different individuals at different time periods. The simulated example conducted by `swdpower` in R package **swdpwr**:

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1,1),4),rep(c(0,0,1,1),4),rep(c(0,0,0,1),4)),
```

Result
I = 12
J = 4
K = 100
Total sample size = 1200
response = 2
model = 2
link = 1
Baseline effect (mu): 0.5
Treatment effect (beta): 0.15
Time effect (gamma J): 0
ICC0: 0.1
ICC1: 0.05
ICC2: 0.2
Type I error = 0.05
Power = 0.985

Table 4: Output from macro `swdpwr`

```
12, 4, byrow=TRUE)
R> swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 2, link = 3,
mu = 0, beta = 0.4, gammaJ = 0.4, alpha = 0.05, ICC0 = 0.03, ICC1 = 0.015,
 ICC2 = 0.015)
$`design matrix dataset: row-cluster col-time`
      [,1] [,2] [,3] [,4]
 [1,]    0    1    1    1
 [2,]    0    1    1    1
 [3,]    0    1    1    1
 [4,]    0    1    1    1
 [5,]    0    0    1    1
 [6,]    0    0    1    1
 [7,]    0    0    1    1
 [8,]    0    0    1    1
 [9,]    0    0    0    1
[10,]    0    0    0    1
[11,]    0    0    0    1
[12,]    0    0    0    1

$Summary
                         [,1]
I                      12.000
J                       4.000
K                     100.000
total sample size    4800.000
response                2.000
model                   2.000
```

Result
I = 12
J = 4
K = 50
Total sample size = 2400
response = 2
model = 2
link = 1
Baseline effect (mu): 0.5
Treatment effect (beta): 0.15
Time effect (gamma J): 0.05
ICC0: 0.2
ICC1: 0.2
ICC2: 0.2
Type I error = 0.05
Power = 0.994

Table 5: Output from macro `swdpwr`

```
link                    3.000
baseline effect mu      0.000
treatment effect beta   0.400
time effect gamma_J     0.400
ICC0                    0.030
ICC1                    0.015
ICC2                    0.015
Type I error            0.050
Power                   0.691
```

Above is a cross-sectional SWD with 100 individuals at each time period in each cluster (a total sample size of 4800). The design is conducted on the marginal method under logit link function with time effects. In this design, the powe given by R is around 0.691 for the alternative hypothesis $\beta_A = 0.4$.

When illustrating the cross-sectional settings in SAS, we use a special example of cross-sectional cases as introduced in Zhou *et al.* (2020), where all three levels of correlation structure are equal and the identity link is utilized. Similarly, we use the data set `design2` generated previously as the study design.

The macro called to do power calculations would be:

```
%swdpwr(I = 12, J = 4, K = 50, dataset = design2, response = 2, model = 2,
link = 1, mu = 0.5, beta = 0.15, gammaJ = 0.05, alpha = 0.05,
ICC0 = 0.2, ICC1 = 0.2, ICC2 = 0.2)
```

With a total sample size of 2400, the power under this scenario is 0.994 for the alternative hypothesis $\beta_A = 0.15$ as in Table 5 where a single correlation parameter is used to describe the working correlation.

4.3. Method with continuous outcomes under identity link

Linear mixed effects models (conditional method) have been widely used in the design and analysis of SWDs, for example, see: Hussey and Hughes (2007), Hooper *et al.* (2016), Hemming *et al.* (2015). Li *et al.* (2018b) proposed methods for analyzing SWDs with marginal model for continuous outcomes under the identity link function. As the equivalence between conditional models and marginal models when the regression is linear and in a mean 0 random effect with identity link (Ritz and Spiegelman 2004), along with the proof for equivalence of variance in Appendix A, it is straightforward to apply procedures in Li *et al.* (2018b) to power calculations with continuous outcomes, when three levels of correlation parameters are considered.

The design for continuous outcomes is conducted under the identity link function and is also estimated by GEE. Both cross-sectional and cohort designs with and without time effects can be accommodated. Here, the argument `sigma2` is a necessary parameter to be specified with continuous outcomes, which equals σ_t^2 in the marginal model setting. The three correlation parameters are still used for cross-sectional and cohort designs, which should ensure a positive definite correlation matrix R_i but without additional restrictions related to marginal means due to binary outcomes.

A simulated example from Li *et al.* (2018b) is used for illustrating the use of function `swdpower` in R:

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1),4),rep(c(0,0,1),4)),8,3, byrow=TRUE)
R> swdpower(I = 8, J = 3, K = 24, dataset, response = 1, model = 2, link = 1,
mu = 0.1, beta = 0.2, gammaJ = 0.1, sigma2 = 0.095, alpha = 0.05, ICC0 = 0.03,
ICC1 = 0.015, ICC2 = 0.2)
$`design matrix dataset: row-cluster col-time`
     [,1] [,2] [,3]
[1,]    0    1    1
[2,]    0    1    1
[3,]    0    1    1
[4,]    0    1    1
[5,]    0    0    1
[6,]    0    0    1
[7,]    0    0    1
[8,]    0    0    1

$Summary
                          [,1]
I                        8.000
J                        3.000
K                       24.000
total sample size      192.000
response                 1.000
model                    2.000
link                     1.000
baseline effect mu       0.100
```

```
treatment effect beta    0.200
time effect gamma_J      0.100
marginal variance        0.095
ICC0                     0.030
ICC1                     0.015
ICC2                     0.200
Type I error             0.050
Power                    0.965
```

In this simulated trial, the outcome is continuous with total marginal variance of 0.095, conducted in 8 clusters and 3 time periods (a total sample size of 192). Marginal method under identity link function with time effects is utilized for the calculation. This trial is in a cohort design with three different levels of correlation parameter. The power obtained from the software is around 0.965 for the alternative hypothesis $\beta_A = 0.2$.

Also, we illustrate a cross-sectional example in macro `swdpwr` using SAS, and prepare a SAS data set `design3`:

```
data design3;
input numofclusters time1 time2 time3;
cards;
4 0 1 1
4 0 0 1
;
run;
```

The macro called to do power calculations would be:

```
%swdpwr(I = 8, J = 3, K = 50, dataset = design3, response = 1,
 model = 2, link = 1, mu = 0.1, beta = 0.2, gammaJ = 0, sigma2 = 0.095,
 alpha = 0.05, ICC0 = 0.3, ICC1 = 0.2, ICC2 = 0.2)
```

With a total sample size of 1200, we get the power 0.994 for the alternative hypothesis $\beta_A = 0.2$ from SAS output in Table 6.

It should be noticed that for continuous outcomes only identity link function is established in the software, scenarios with or without time effects can both be considered.

4.4. Error message occurs

When the input parameters such as correlation parameters, Type I error, and mean response for binary outcomes are out of range, an error message will return in the software. Here we give examples including all scenarios that error occurs, and users could check the input parameters according to the error message and the suggestions about revision of them.

In the first example, the requirement that R_i is positive-definite is violated as correlation parameters exceed plausible ranges. This error could occur for both types of outcomes in all models.

Result
I = 8
J = 3
K = 50
Total sample size = 1200
response = 1
model = 2
link = 1
Baseline effect (mu): 0.1
Treatment effect (beta): 0.2
Time effect (gamma J): 0
Marginal variance: 0.095
ICC0: 0.3
ICC1: 0.2
ICC2: 0.2
Type I error = 0.05
Power = 0.994

Table 6: Output from macro `swdpwr`

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1,1),4),rep(c(0,0,1,1),4),rep(c(0,0,0,1),4)),
12, 4, byrow=TRUE)
R> swdpower(I = 12, J = 4, K = 100, dataset, response = 1, model = 2, link = 3,
mu = 0, beta = 0.2, gammaJ = 0.1, sigma2 = 0.095, alpha = 0.05, ICC0 = 0.01,
ICC1 = 0.2, ICC2 = 0.2)
Error in swdpower(I = 12, J = 4, K = 100, dataset, response = 1, model = 2, :
Correlation matrix R is not positive definite. Please check whether the inter-
period correlation is unrealistically larger than the within-period correlation
or the within-individual correlation.
```

The second example is with binary outcomes, which is also related to the range of correlations as they are additionally restricted by marginal means due to binary outcomes (Qaqish 2003; Ridout *et al.* 1999).

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1,1),4),rep(c(0,0,1,1),4),rep(c(0,0,0,1),4)),
12, 4, byrow=TRUE)
R> swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 2, link = 1,
mu = 0.1, beta = 0.7, gammaJ = 0, alpha = 0.05, ICC0 = 0.1, ICC1 = 0.05,
ICC2 = 0.2)
Error in swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 2,  :
Correlation paramters do not satisfy the restrictions of Qaqish (2003).
Please check whether it is possible to reduce the large difference between
baseline mean response and mean response at the end of the trial, or make
adjustments to the ICCs.
```

The third example also has binary outcomes, where the mean responses can be interpreted

as the probability of outcomes. Sometimes, the input parameters related to mean responses (mu, beta, gammaJ) may exceed the range for valid probability. For continuous outcomes, there is no concern about these parameters.

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1,1),4),rep(c(0,0,1,1),4),rep(c(0,0,0,1),4)),
12, 4, byrow=TRUE)
R> swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 1, link = 1,
mu = 1.1, beta = -0.3, gammaJ = -0.1, alpha = 0.05, ICC0 = 0.1, ICC1 = 0.05,
ICC2 = 0.2)
Error in swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 1, :
Violate theory of probability under identity link: max(mu+gammaJ,mu+beta+gammaJ,
mu,mu+beta)>1. Please check whether any of these four values are out of range.
```

Besides, ICCs and Type I error should be between 0 and 1 as well, for instance:

```
R> library(swdpwr)
R> dataset = matrix(c(rep(c(0,1,1,1),4),rep(c(0,0,1,1),4),rep(c(0,0,0,1),4)),
12, 4, byrow=TRUE)
R> swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 2, link = 1,
mu = 0.6, beta = 0.2, gammaJ = 0.1, alpha = 0.05, ICC0 = 1.1, ICC1 = 0.05,
ICC2 = 0.2)
Error in swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 2,  :
Violate range of ICC: max(ICC0,ICC1,ICC2)>1. Please correct the values of
correlation parameters (between 0 and 1).

R> swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 2, link = 1,
mu = 0.6, beta = 0.2, gammaJ = 0.1, alpha = 1.05, ICC0 = 0.1, ICC1 = 0.05,
ICC2 = 0.2)
Error in swdpower(I = 12, J = 4, K = 100, dataset, response = 2, model = 2, :
Type I error is larger than 1, it should be between 0 and 1.
```

5. Application

Illustrations in the previous section are based on simulated examples, here we give two real world applications for the marginal method and the conditional method, respectively: the Washington Expedited Partner Therapy (EPT) trial and the Tanzania postpartum intrauterine device (PPIUD) study.

The EPT trial was a community-level trial and employed cluster randomized SWD for promoting the intervention of EPT. The outcome was Chlamydia positive or not, which is binary. 24 local health jurisdictions (LHJs) were eligible in this trial and each represented a cluster. The number of time periods was 5 and the intervention was initiated at four time periods, with 6 clusters entering the intervention group at each time period. We can describe this design in a SAS data set `ept`:

```
data ept;
input numofclusters time1 time2 time3 time4 time5;
```

```
cards;
6 0 1 1 1 1
6 0 0 1 1 1
6 0 0 0 1 1
6 0 0 0 0 1
;
run;
```

The design of the study's outcomes used a generalized linear mixed model under log link with covariates for intervention status and time period (Golden, Kerani, Stenger, Hughes, Aubin, Malinski, and Holmes 2015). This cross-sectional design assumes 162 individuals in each cluster at each time period (a total sample size of 19440). Based on preliminary data: the baseline prevalence of Chlamydia is about 0.05; the coefficient of variation is 0.3; the Type I error is set to be 0.05; and a prevalence ratio of 0.7 is to be tested. Coefficient of variance is defined for explaining cluster effects on the variance (Hayes and Bennett 1999) and is closely related to regular intraclass correlation. Calculated following Hussey and Hughes (2007), the intraclass correlation is approximately 0.0054 and in this scenario $\alpha_0 = \alpha_1 = \alpha_2 = 0.0054$. We estimate from preliminary data that $\mu = -3$ and $\beta = -0.36$. We also assume a time effect of $\gamma_J = -0.1$. We conduct this design using the marginal method in SAS. The macro called to do power calculations with the marginal method would be:

```
%swdpwr(I = 24, J = 5, K = 162, dataset = ept, response = 2, model = 2,
link = 2, mu = -3, beta = -0.36, gammaJ= -0.1, alpha = 0.05, ICC0 = 0.0054,
ICC1 = 0.0054, ICC2 = 0.0054)
```

We get the power 0.834 for the alternative hypothesis $\beta_A = -0.36$ from the marginal model in Table 7, which corresponds to the anticipated power of 80% from similar scenarios considered in Golden *et al.* (2015) and Hussey and Hughes (2007).

The Tanzania PPIUD study utilized a SWD to assess the causal effect of PPIUD intervention on subsequent pregnancy (Canning, Shah, Pearson, Pradhan, Karra, Senderowicz, Bärnighausen, Spiegelman, and Langer 2016). The binary outcome whether the participant is currently pregnant or has had a pregnancy that was terminated will be obtained at 18 months postpartum. 6 hospitals were selected into the trial and the study lasted for 18 months with 4 time periods. We can describe this design in a SAS data set `PPIUD`:

```
data PPIUD;
input numofclusters time1 time2 time3 time4;
cards;
3 0 1 1 1
3 0 0 0 1
;
run;
```

A generalized linear mixed model under identity link without time effects is employed for design. For illustrative purposes, we consider a small cluster size and this cross-sectional design assumes 120 individuals in each cluster at each time period (a total sample size of 2880). As in Canning *et al.* (2016), the baseline proportion of pregnancy is about 0.24; the

Result
I = 24
J = 5
K = 162
Total sample size = 19440
response = 2
model = 2
link = 2
Baseline effect (mu): -3
Treatment effect (beta): -0.36
Time effect (gamma J): -0.1
ICC0: 0.0054
ICC1: 0.0054
ICC2: 0.0054
Type I error = 0.05
Power = 0.834

Table 7: Output from macro `swdpwr`

ICC is 0.15; the Type I error is set to be 0.05; and a prevalence ratio of around 0.8 is to be tested. We estimate from preliminary data that $\mu = 0.24$ and $\beta = -0.046$. We also assume no time effects. We conduct this design using the conditional method in SAS. The macro called to do power calculations with the conditional method would be:

```
%swdpwr(I = 6, J = 4, K = 120, dataset = PPIUD, response = 2, model = 1,
 link = 1, mu = 0.24, beta = -0.046, gammaJ = 0, alpha = 0.05, ICC0 = 0.15,
 ICC1 = 0.15, ICC2 = 0.15)
```

We get the power 0.846 for the alternative hypothesis $\beta_A = -0.046$ from the conditional model in Table 8. This result is consistent with the conclusion of 80% power or more in Canning *et al.* (2016).

6. Discussion

This article has described the use of R package **swdpwr** and SAS macro `swdpwr` for power calculations in SWDs. The software is designed under two platforms for users to specify different input parameters in different scenarios, which accommodates cross-sectional and cohort settings, binary and continuous outcomes, marginal method and conditional method, different link functions, with and without time effects, etc. The development of this software addresses the implementation gap between newly proposed methodology and applying them to obtain more accurate power calculation for binary outcomes in SWDs, instead of relying on the approximation method in Hussey and Hughes (2007). The core of the software is developed in Fortran to ensure that the computation is efficient, and is linked to SAS and R by the foreign function interface. As the number of time periods and individuals in each cluster increases, the calculation may become time-consuming, thus we have implemented some tricks to save the computational cost. For example, we derived a simplified close-form

Result
I = 6
J = 4
K = 120
Total sample size = 2880
response = 2
model = 1
link = 1
Baseline effect (mu): 0.24
Treatment effect (beta): -0.046
Time effect (gamma J): 0
ICC0: 0.15
ICC1: 0.15
ICC2: 0.15
Type I error = 0.05
Power = 0.846

Table 8: Output from macro `swdpwr`

expression of matrix inversion for estimation under the marginal method, as in Li *et al.* (2018b). Hence, the current version of the software has high efficiency in calculation for the marginal method. Package **swdpwr** is available via the Comprehensive R Archive Network (CRAN) as a contributed package at: . The package is also available as a **shiny** (Chang, Cheng, Allaire, Xie, and McPherson 2019) app, available online ( ) such that users without programming skills can also use this software easily.

In the future we plan to accommodate different numbers of individuals in each cluster or even at each time period to make the software more flexible. Furthermore, we are developing a partition method (Zhou *et al.* 2020) and corresponding algorithms for the conditional method to reduce much more computational cost of this method. In addition, current software only accommodates nested exchangeable and block exchangeable correlation structures, which can be extended for exponential decay (Kasza, Hemming, Hooper, Matthews, and Forbes 2019) and proportional decay correlation structures (Li 2020) further.

Acknowledgments

References

Breslow NE, Clayton DG (1993). "Approximate inference in generalized linear mixed models." *Journal of the American statistical Association*, **88**(421), 9–25. doi:10.1080/01621459.1993.10594284.

Canning D, Shah IH, Pearson E, Pradhan E, Karra M, Senderowicz L, Bärnighausen T, Spiegelman D, Langer A (2016). "Institutionalizing postpartum intrauterine device (IUD) services in Sri Lanka, Tanzania, and Nepal: study protocol for a cluster-

randomized stepped-wedge trial." *BMC pregnancy and childbirth*, **16**(1), 362. doi:10.1186/s12884-016-1160-0.

Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2019). *shiny: Web Application Framework for R*. URL https://CRAN.R-project.org/package=shiny.

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**(1), 352. doi:10.1186/s13063-015-0842-7.

Golden MR, Kerani RP, Stenger M, Hughes JP, Aubin M, Malinski C, Holmes KK (2015). "Uptake and population-level impact of expedited partner therapy (EPT) on Chlamydia trachomatis and Neisseria gonorrhoeae: the Washington State community-level randomized trial of EPT." *PLOS medicine*, **12**(1). doi:10.1371/journal.pmed.1001777.

Hayes R, Bennett S (1999). "Simple sample size calculation for cluster-randomized trials." *International journal of epidemiology*, **28**(2), 319–326. doi:10.1093/ije/28.2.319.

Hemming K, Girling A (2013). "The efficiency of stepped wedge vs. cluster randomized trials: stepped wedge studies do not always require a smaller sample size." *Journal of clinical epidemiology*, **66**(12), 1427. doi:10.1016/j.jclinepi.2013.07.007.

Hemming K, Girling A (2014). "A menu-driven facility for power and detectable-difference calculations in stepped-wedge cluster-randomized trials." *The Stata Journal*, **14**(2), 363–380. doi:10.1177/1536867X1401400208.

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*. 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), 181–196. doi:10.1002/sim.6325.

Hemming K, Taljaard M (2016). "Sample size calculations for stepped wedge and cluster randomised trials: a unified approach." *Journal of clinical epidemiology*, **69**, 137–146. doi:10.1016/j.jclinepi.2015.08.015.

Hooper R, Teerenstra S, de Hoop E, Eldridge S (2016). "Sample size calculation for stepped wedge and other longitudinal cluster randomised trials." *Statistics in medicine*, **35**(26), 4718–4728. doi:10.1002/sim.7028.

Hughes J, Hakhu NR, Voldal E (2019). *swCRTdesign: Stepped Wedge Cluster Randomized Trial (SW CRT) Design*. URL https://CRAN.R-project.org/package=swCRTdesign.

Hughes JP, Granston TS, Heagerty PJ (2015). "Current issues in the design and analysis of stepped wedge trials." *Contemporary clinical trials*, **45**, 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, Hemming K, Hooper R, Matthews J, Forbes A (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.

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, Turner EL, Preisser JS (2018a). "Optimal allocation of clusters in cohort stepped wedge designs." *Statistics & Probability Letters*, **137**, 257–263. doi:10.1016/j.spl.2018.02.002.

Li F, Turner EL, Preisser JS (2018b). "Sample size determination for GEE analyses of stepped wedge cluster randomized trials." *Biometrics*, **74**(4), 1450–1458. doi:10.1111/biom.12918.

Liang KY, Zeger SL (1986). "Longitudinal data analysis using generalized linear models." *Biometrika*, **73**(1), 13–22. doi:10.1093/biomet/73.1.13.

Liao X, Zhou X, Spiegelman D (2015). "A note on "Design and analysis of stepped wedge cluster randomized trials"." *Contemporary clinical trials*, **45**(Pt B), 338. doi:10.1016/j.cct.2015.09.011.

Martin J, Girling A, Nirantharakumar K, Ryan R, Marshall T, Hemming K (2016a). "Intra-cluster and inter-period correlation coefficients for cross-sectional cluster randomised controlled trials for type-2 diabetes in UK primary care." *Trials*, **17**(1), 402. doi:10.1186/s13063-016-1532-9.

Martin J, Taljaard M, Girling A, Hemming K (2016b). "Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials." *BMJ open*, **6**(2), e010166. doi:10.1136/bmjopen-2015-010166.

McCulloch CE (1997). "Maximum likelihood algorithms for generalized linear mixed models." *Journal of the American statistical Association*, **92**(437), 162–170. doi:10.1080/01621459.1997.10473613.

Murray DM, *et al.* (1998). *Design and analysis of group-randomized trials*, volume 29. Oxford University Press, USA.

Pinheiro J, Bates D (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

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.

Ridout MS, Demetrio CG, Firth D (1999). "Estimating intraclass correlation for binary data." *Biometrics*, **55**(1), 137–148. doi:10.1111/j.0006-341X.1999.00137.x.

Ritz J, Spiegelman D (2004). "Equivalence of conditional and marginal regression models for clustered and longitudinal data." *Statistical Methods in Medical Research*, **13**(4), 309–323. doi:10.1191/0962280204sm368ra.

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), 1230–1237. doi:10.1111/j.1541-0420.2009.01374.x.

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.

Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S (2013). "Stepped wedge designs could reduce the required sample size in cluster randomized trials." *Journal of clinical epidemiology*, **66**(7), 752–758. doi:10.1016/j.jclinepi.2013.01.009.

Zhou X, Liao X, Kunz LM, Normand SLT, Wang M, Spiegelman D (2020). "A maximum likelihood approach to power calculations for stepped wedge designs of binary outcomes." *Biostatistics*, **21**(1), 102–121. doi:10.1093/biostatistics/kxy031.

Zhou X, Liao X, Spiegelman D (2017). ""Cross-sectional" stepped wedge designs always reduce the required sample size when there is no time effect." *Journal of clinical epidemiology*, **83**, 108–109. doi:10.1016/j.jclinepi.2016.12.011.

A. Equivalence of three models with continuous outcomes

As mentioned in the main part of the paper, we include procedures for power calculation of SWDs with continuous outcomes in our software based on Li *et al.* (2018b), which employed a block exchangeable correlation structure with three correlation parameters that accommodate both cohort designs and cross-sectional designs under the GEE framework. Hussey and Hughes (2007) utilized the linear mixed effects model with cluster-level means as the response for a cross-sectional setting including only a random effect for clusters. Hooper *et al.* (2016) generalized Hussey and Hughes (2007)'s method to closed cohort CRTs with random effects of individual within cluster, time within cluster and cluster effect, which is the same as in Li *et al.* (2018b). However, Hooper *et al.* (2016) was modelled under a linear mixed effects model with individual level responses, and Li *et al.* (2018b) used the GEE with marginal mean responses.

In this note, we would like to show the equivalence of power obtained in these three models (Hussey and Hughes 2007; Hooper *et al.* 2016; Li *et al.* 2018b) with continuous response under correpsonding scenarios through the equivalence of variance for the intervention effect. Hence, power calculation for continuous outcomes under different scenarios can all be incorporated by Li *et al.* (2018b) directly.

A.1. Cases with time effects

With I clusters and J time periods, Hussey and Hughes (2007) defined:

$$Y_{ijk} = \mu + X_{ij}\beta + \gamma_j + b_i + \epsilon_{ijk} \tag{15}$$

where Y_{ijk} is the individual continuous response in cluster i at time period j of individual k, X_{ij} is a binary treatment assignment (1=intervention; 0=standard of care) in cluster i at time period j, β is the treatment effect, μ is the baseline effect in control groups, γ_j is the fixed time effect corresponding to time period j (with $\gamma_1 = 0$), b_i is the random effect for cluster i with $b_i \sim N(0, \tau^2)$, and $\epsilon_{ijk} \sim N(0, \sigma_e^2)$. We also assume that ϵ_{ijk} is independent of b_i, and K individuals at each time period in each cluster.

The cluster means model obtained by summing over individuals within a cluster is:

$$Y_{ij} = \mu + X_{ij}\beta + \gamma_j + b_i + \epsilon_{ij} \tag{16}$$

where $\epsilon_{ij} = \sum_k \epsilon_{ijk}/K \sim N(0, \sigma^2)$ and $\sigma^2 = \sigma_e^2/K$.

The estimate of the fixed parameter vector $\eta = (\mu, \gamma_2, \gamma_3, ..., \gamma_J, \beta)$ is obtained by weighted least squares (WLS). We assume that $\mathbf{Z}$ is the $IJ * (J+1)$ design matrix corresponds to the parameter vector η under the SWD framework. In order to obtain the power, we are most interested in the covariance matrix of η, which is $(\mathbf{Z'V^{-1}Z})^{-1}$, where $\mathbf{V}$ is an $IJ * IJ$ block diagonal matrix that measures the covariance of mean response between different time periods in all clusters. When assuming K individuals at each time period per cluster, we can show that (Hussey and Hughes 2007):

$$Var(\hat{\beta}) = \frac{I\sigma^2(\sigma^2 + J\tau^2)}{(IU - W)\sigma^2 + (U^2 + IJU - JW - IV)\tau^2} \tag{17}$$

where $U = \sum_{ij} X_{ij}$, $W = \sum_j (\sum_i X_{ij})^2$ and $V = \sum_i (\sum_j X_{ij})^2$.

To generalize the cross-sectional settings in Hussey and Hughes (2007) to cohort studies, Hooper *et al.* (2016) assumed a linear mixed effects model:

$$Y_{ijk} = \mu + X_{ij}\beta + \gamma_j + b_i + c_{ij} + \pi_{ik} + \epsilon_{ijk} \tag{18}$$

where b_i is the cluster random effect distributed by $N(0, \sigma_b^2)$, c_{ij} is the cluster-by-time interaction random effect distributed by $N(0, \sigma_c^2)$, π_{ik} is the random effect for repeated measures of one individual distributed by $N(0, \sigma_\pi^2)$, and $\epsilon_{ijk} \sim N(0, \sigma_e^2)$. We also assume that b_i, c_{ij}, π_{ik} and ϵ_{ijk} are independent of each other, and the total variance of Y_{ijk} is $\sigma_t^2 = \sigma_b^2 + \sigma_c^2 + \sigma_\pi^2 + \sigma_e^2$. The three-level correlation structure R_i is defined the same with that in Section 2.2 and can be specified for cross-sectional or cohort designs. Model 18 is described at individual-level responses based on mixed effects model, which agrees with the population-averaged marginal model in Li *et al.* (2018b):

$$\mu_{ijk} = \mu + X_{ij}\beta + \gamma_j \tag{19}$$

where μ_{ijk} is the marginal mean response of Y_{ijk}. Li *et al.* (2018b) and Hooper *et al.* (2016) are proposed in the same scenarios that accommodate three correlation parameters, under marginal method and conditional method, respectively. In both models, the covariance is estimated by the model-based estimator $(\sum_i \mathbf{Z_i'} \mathbf{V_i^{-1}} \mathbf{Z_i})^{-1}$, where $\mathbf{Z_i}$ is the $J * (J+1)$ design matrix corresponds to the parameter vector η under the SWD framework within cluster i and $\mathbf{V_i} = \sigma_t^2 R_i$ is the working covariance matrix within cluster i based on a block exchangeable correlation structure. Hence model (18) and (19) are equivalent and obtain the same power.

As shown in Li *et al.* (2018b), the variance of intervention effect estimator is:

$$Var(\hat{\beta}) = \frac{(\phi/K)IJ\lambda_3\lambda_4}{(U^2 + IJU - JW - IV)\lambda_4 - (U^2 - IV)\lambda_3} \tag{20}$$

where $U = \sum_{ij} X_{ij}$, $W = \sum_j (\sum_i X_{ij})^2$, $V = \sum_i (\sum_j X_{ij})^2$, $\lambda_3 = 1 + (K-1)(\alpha_0 - \alpha_1) - \alpha_2$, $\lambda_4 = 1 + (K-1)\alpha_0 + (J-1)(K-1)\alpha_1 + (J-1)\alpha_2$. Here ϕ is the total variance σ_t^2 of Y_{ijk}.

The equivalence of model (16) and (19) is proved under the cross-sectional setting in Hussey and Hughes (2007) with $\alpha_0 = \alpha_1 = \alpha_2$. By fitting into the variance of the intervention effect of (20) with $\alpha_0 = \alpha_1 = \alpha_2 = \frac{\tau^2}{\sigma_e^2 + \tau^2}$ and $\phi = \sigma_e^2 + \tau^2$, we get exactly (17) with $\sigma^2 = \frac{\sigma_e^2}{K}$. Thus we can conclude the equivalence of these three models in cases with time effects.

A.2. Cases without time effects

Here we also consider cases without fixed time effects and derive these three models accordingly. Other notations are the same as cases with time effects.

For Hussey and Hughes (2007) the model without time effects is denoted as:

$$Y_{ijk} = \mu + X_{ij}\beta + b_i + \epsilon_{ijk} \tag{21}$$

where Y_{ijk} is the individual continuous response in cluster i at time period j of individual k, X_{ij} is a binary treatment assignment (1=intervention; 0=standard of care) in cluster i at time period j, β is the treatment effect, μ is the baseline effect in control groups, b_i is the random effect for cluster i with $b_i \sim N(0, \tau^2)$, and $\epsilon_{ijk} \sim N(0, \sigma_e^2)$. We also assume that ϵ_{ijk} is independent of b_i.

The cluster means model obtained by summing over individuals within a cluster is:

$$Y_{ij} = \mu + X_{ij}\beta + b_i + \epsilon_{ij} \tag{22}$$

where $\epsilon_{ij} = \sum_k \epsilon_{ijk}/K \sim N(0, \sigma^2)$ and $\sigma^2 = \sigma_e^2/K$.

The estimate of the fixed parameter vector $\eta = (\mu, \beta)$ is obtained by weighted least squares (WLS). We assume that $\mathbf{Z}$ is the $IJ * 2$ design matrix corresponds to the parameter vector η under the SWD framework. In order to obtain the power, we are most interested in the covariance matrix of η, which is $(\mathbf{Z'V^{-1}Z})^{-1}$, where $\mathbf{V} = \sigma_t^2 R_i$ is an $IJ * IJ$ block diagonal matrix that measures the covariance of mean response between different time periods in all clusters. When assuming K individuals at each time period per cluster, Liao, Zhou, and Spiegelman (2015) showed that the variance of intervention effect is:

$$Var(\hat{\beta}) = \frac{IJ(\sigma^2 + J\tau^2)\sigma^2}{(IJU - U^2)\sigma^2 + IJ(JU - V)\tau^2} \tag{23}$$

where $U = \sum_{ij} X_{ij}$ and $V = \sum_i (\sum_j X_{ij})^2$.

The model that accounts for closed cohort designs without time effects under Hooper *et al.* (2016) is:

$$Y_{ijk} = \mu + X_{ij}\beta + b_i + c_{ij} + \pi_{ik} + \epsilon_{ijk} \tag{24}$$

where b_i is the cluster random effect distributed by $N(0, \sigma_b^2)$, c_{ij} is the cluster-by-time interaction random effect distributed by $N(0, \sigma_c^2)$, π_{ik} is the random effect for repeated measures of one individual distributed by $N(0, \sigma_\pi^2)$, and $\epsilon_{ijk} \sim N(0, \sigma_e^2)$. Three correlation parameters are defined in the same way as that in the previous section. This model is described at individual-level responses based on mixed effects model, which still agrees with the population-averaged marginal model in Li *et al.* (2018b):

$$\mu_{ijk} = \mu + X_{ij}\beta \tag{25}$$

where μ_{ijk} is the marginal mean response of Y_{ijk}. Li *et al.* (2018b) and Hooper *et al.* (2016) are proposed under the same scenarios that accommodate three correlation parameters, under marginal method and conditional method, respectively. In both models, the covariance is estimated by the model-based estimator $(\sum_i \mathbf{Z_i'V_i^{-1}Z_i})^{-1}$, where $\mathbf{Z_i}$ is the $J*2$ design matrix corresponds to the parameter vector η under the SWD framework within cluster i and $\mathbf{V_i}$ is the working covariance matrix within cluster i based on a block exchangeable correlation structure. Hence model (24) and (25) are equivalent, and obtain the same power.

According to our derivation following Li *et al.* (2018b), the variance of intervention effect estimator under cases without time effects is:

$$Var(\hat{\beta}) = \frac{(\phi/K)IJ\lambda_3\lambda_4}{(IJU - IV)\lambda_4 - (U^2 - IV)\lambda_3} \tag{26}$$

where $U = \sum_{ij} X_{ij}$ and $V = \sum_i (\sum_j X_{ij})^2$, $\lambda_3 = 1 + (K-1)(\alpha_0 - \alpha_1) - \alpha_2$, $\lambda_4 = 1 + (K-1)\alpha_0 + (J-1)(K-1)\alpha_1 + (J-1)\alpha_2$. Here ϕ is the total variance σ_t^2 of Y_{ijk}.

The equivalence of model (22) and (25) is proved under the cross-sectional setting in Hussey and Hughes (2007) with $\alpha_0 = \alpha_1 = \alpha_2$. By fitting into the variance of the intervention effect of (26) with $\alpha_0 = \alpha_1 = \alpha_2 = \frac{\tau^2}{\sigma_e^2 + \tau^2}$ and $\phi = \sigma_e^2 + \tau^2$, we get exactly (23) with $\sigma^2 = \frac{\sigma_e^2}{K}$. Thus we can conclude the equivalence of these three models in cases without time effects.

Affiliation:

Jiachen Chen
Department of Biostatistics
Center for Methods in Implementation and Prevention Science (CMIPS)
Yale School of Public Health, New Haven, CT
E-mail: jiachen.chen@yale.edu
URL: https://medicine.yale.edu/profile/jiachen_chen/

Xin Zhou
Department of Biostatistics
Center for Methods in Implementation and Prevention Science (CMIPS)
Yale School of Public Health, New Haven, CT
E-mail: xin.zhou@yale.edu
URL: https://publichealth.yale.edu/profile/xin_zhou/

Fan Li
Department of Biostatistics
Center for Methods in Implementation and Prevention Science (CMIPS)
Yale School of Public Health, New Haven, CT
E-mail: fan.f.li@yale.edu
URL: https://publichealth.yale.edu/profile/fan_f_li/

Donna Spiegelman *(corresponding author)*
Department of Biostatistics
Center for Methods in Implementation and Prevention Science (CMIPS)
Yale School of Public Health, New Haven, CT
and
Channing Division of Network Medicine
Department of Medicine
Harvard T.H. Chan School of Public Health, Boston, MA
E-mail: donna.spiegelman@yale.edu
URL: https://publichealth.yale.edu/cmips/profile/donna_spiegelman/

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
MMMMMM YYYY, Volume VV, Issue II *Submitted:* yyyy-mm-dd
doi:10.18637/jss.v000.i00 *Accepted:* yyyy-mm-dd