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 and marginal models can both be utilized in power calculations for stepped wedge cluster randomized trials. Hussey and Hughes (2007) first provided the design and analysis for this particular type of cluster randomized trials based on the linear mixed effects model for continuous outcomes, along with an approximation to this approach for binary outcomes. Although not accurate for binary outcomes, it has been widely used. To improve the approximation for binary outcomes, Zhou, Liao, Kunz, Normand, Wang, and Spiegelman (2020) and Li, Turner, and Preisser (2018b) have recently published two new methods for stepped wedge designs (SWDs) of binary outcomes. However, these new methods have not been implemented in publicly available software. In this work, we have developed a SAS macro %swdpwr and an R package **swdpwr** to calculate power in SWDs in various settings for 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, improving administrative convenience and reducing 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 stepped wedge designs have been considered: 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 outcome measures at different time periods (Hemming and Girling 2013). A distinguishing feature of all 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 between-period correlation coefficients may be different and thus should be separately considered in designing SWDs (Martin *et al.* 2016a). An additional within-individual correlation should be included when it is a cohort SWD to account for the repeated measures within 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 responses 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 studying random effects with target population random effects, Li *et al.* (2018b) proposed methods for the design and analysis of SWDs using marginal models. Alternatively, the conditional model considers the causal effect of interventions on individual, under the assumption of no unmeasured confounders present or other source of bias.

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, estimated by weighted least squares for continuous outcomes, and provided an approximation to 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 over-

estimate 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 not yet been implemented in publicly available software, making it difficult for researchers to apply these new methods to design their studies rigorously. Additionally, existing software for SWDs focused on limited settings and did not have accurate power calculations for binary outcomes. Hemming and Girling (2014) developed a Stata menu-driven program steppedwedge based on Hussey and Hughes (2007), the **swCRTdesign** (Hughes, Hakhu, and Voldal 2019) package in R and an Excel spreadsheet (http://faculty.washington.edu/jphughes/pubs.html) also implemented Hussey and Hughes (2007), all of which utilized the linear mixed effects model for continuous outcomes and binary proportions considering only the traditional intra-cluster correlation coefficient for cross-sectional designs assuming fixed time effects. Hemming, Kasza, Hooper, Forbes, and Taljaard (2020) developed 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 included the linear mixed effects model approximation for binary outcomes. All these software did not implement accurate power calculations for SWDs with binary outcomes. There is another R package **SWSamp** (https://sites.google.com/a/statistica.it/gianluca/swsamp) based on Baio, Copas, Ambler, Hargreaves, Beard, and Omar (2015), which allowed simulation-based sample size and power calculations for several general scenarios including cross-sectional and cohort designs for continuous, binary and count outcomes. However, this package considered a less complex model and regarded time effects as fixed only. Hence, in an effort to make the new methods more accessible and incorporate as many variations as currently available, we have developed user-friendly and computationally efficient software based on the methods proposed by Zhou *et al.* (2020) and Li *et al.* (2018b) to implement power calculations of SWDs for binary and continuous outcomes. The software engine has been developed in Fortran and is incorporated into the SAS macro %swdpwr and the R package **swdpwr**.

2. Methods

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

$$H_0 : \beta = 0 \qquad \text{versus} \qquad H_A : \beta \neq 0, \tag{1}$$

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:

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

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 defined by either asymptotic theory for maximum likelihood estimation (MLE) in a conditional model, or by the theory of generalized 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 $1, \cdots, J$). In a cross-sectional SWD, the size of cluster i is given by $N_i = \sum_{j=1}^{J} K_{ij}$; in a cohort SWD, assuming that $K_{ij} = K_i$ over the active 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 surgery, getting the disease or not). For both outcome types, we first present the general models with three correlation parameters and then give more details about the cases implemented in this software.

2.1. Models for binary outcomes

Due to the correlated nature of the outcomes in SWDs, a generalized linear mixed effects model can be written:

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

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 outcome rate on the scale of the link function 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, the treatment effect. We assume that b_i is the between-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 assume that b_i, c_{ij} and π_{ik} are independent of each other. Let $p_{ijk} = E(Y_{ijk}|X_{ij}, b_i, c_{ij}, \pi_{ik}) = Pr(Y_{ijk} = 1|X_{ij}, b_i, c_{ij}, \pi_{ik})$ be the probability of the outcome for individual k conditioned on the random effects and design allocation, interpreted as the conditional mean response of the individual.

To account for the correlation of outcomes, 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 (corr($Y_{ijk}, Y_{ijk'}$) = α_0 for $k \neq k'$); (2) α_1, the between-period correlation, which measures the similarity between responses from different individuals within the same cluster in different time periods (corr($Y_{ijk}, Y_{ij'k'}$) = α_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 (corr($Y_{ijk}, Y_{ij'k}$) = α_2 for $j \neq j'$).

This framework can accommodate three common link functions $g(\cdot)$: identity, log and logit links. In a cross-sectional design, the correlation structure depends only on α_0 and α_1 since different sets of individuals are included at each time period, and α_2 is no longer relevant. However, in this circumstance, α_1 and α_2 are, by definition, the same, since actually different individuals are assessed at each time period, to obtain a two-level exchangeable correlation model to accommodate the cross-sectional design (Li *et al.* 2018b). Otherwise, a cohort design is implied.

Model (3) accommodates both cross-sectional and cohort SWDs. These designs can also be fitted by the marginal model with the same correlation structure but without random effects:

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

In our software, we implemented specific cases of designs with binary outcomes under conditional and marginal model assumptions.

Restrictions on the allowable parameter space For binary outcomes, the marginal and conditional mean responses limit the ranges of the correlation coefficients inside $[-1, 1]$ (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 working correlation matrix and that the correlations are between 0 and 1. To ensure valid probabilities between 0 and 1 under the identity and log link functions, we also need restrictions for parameters related to the treatment effect, baseline outcome rate on the scale of the link function and time effect parameters. For example, $\mu + \beta + \gamma_j$ should not exceed 0 under log link function. These restrictions apply to both the conditional and marginal models. When the input parameters for power calculations are out of range, the software will return an error message. More details will be discussed in Section 4.4.

Conditional model based on GLMM Zhou *et al.* (2020) considered a cross-sectional SWD modelled by the generalized linear mixed model (GLMM) as a special case of Model 3:

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

All notations are as defined previously. We assume a normal distribution for random effects, $b_i \sim N(0, \tau^2)$, although Zhou *et al.* (2020) showed that results are insensitive to the assumed form of the distribution of the random effects. This model only considers fixed time effects and does not include cluster-by-time interaction random effects. As α_2 is undefined for cross-sectional studies, the correlation structure for this case is $\alpha_0 = \alpha_1$ and Zhou *et al.* (2020) showed that they are equal to $\rho = \frac{Var(E(Y_{ij}|X_{ij}=0,b_i))}{Var(Y_{ij}|X_{ij}=0)}$, where ρ represents an intra-cluster correlation coefficient which measures the correlation between individuals in the same cluster. Under the identity link, $\alpha_0 = \alpha_1 = \frac{\tau^2}{\tau^2 + \mu(1-\mu)}$ (Hussey and Hughes 2007).

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

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

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

and is relevant when the time effects are expected to be minimum (Zhou, Liao, and Spiegelman 2017). The derivation of the likelihood formula and the calculation of the variance of $\hat{\beta}$ and power are similar to the procedures with time effects.

Marginal model based on GEE Cohort SWDs 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), have been studied. The marginal model based on GEE is:

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

Here we denote $p_{ijk} = E(Y_{ijk}|X_{ij}) = P(Y_{ijk} = 1|X_{ij})$, interpreted as the marginal mean response of individuals, μ is the marginal outcome rate on the scale of the link function 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):

$$R_i = (1 - \alpha_0 + \alpha_1 - \alpha_2)I_{JK_i} + (\alpha_2 - \alpha_1)J_J \otimes I_{K_i} + (\alpha_0 - \alpha_1)I_J \otimes J_{K_i} + \alpha_1 J_{JK_i}. \tag{8}$$

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 define $\eta = (\mu, \gamma_2, \gamma_3, ..., \gamma_J, \beta)'$ 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 solved from $\sum_i D_i'V_i^{-1}(Y_i - p_i) = 0$, where $D_i = \partial p_i/\partial \eta'$, $V_i = A_i^{1/2}R_iA_i^{1/2}$, 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 $(\sum_i D_i'(\hat{\eta})V_i^{-1}(\alpha)D_i(\hat{\eta}))^{-1}$. Hence, we can calculate the variance of $\hat{\beta}$.

Cross-sectional SWDs can also be designed under the marginal model with appropriate specification of the correlation parameters α_0 and α_1.

2.2. Models for continuous outcomes

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

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

Notations are as defined previously. Here we assume that $\epsilon_{ijk} \sim N(0, \sigma_e^2)$, and as previously, b_i, c_{ij}, π_{ik} and ϵ_{ijk} are independent of each other, and the total variance of Y_{ijk} is $\sigma_t^2 = \sigma_b^2 + \sigma_c^2 + \sigma_\pi^2 + \sigma_e^2$. We note that, $\text{corr}(Y_{ijk}, Y_{ijk'}) = \alpha_0 = \frac{(\sigma_b^2 + \sigma_c^2)}{\sigma_t^2}$ for $k \neq k'$, $\text{corr}(Y_{ijk}, Y_{ij'k'}) = \alpha_1 = \frac{\sigma_b^2}{\sigma_t^2}$ for $j \neq j', k \neq k'$, and $\text{corr}(Y_{ijk}, Y_{ij'k}) = \alpha_2 = \frac{(\sigma_b^2 + \sigma_\pi^2)}{\sigma_t^2}$ for $j \neq j'$. The correlation structure R_i and the specification for cross-sectional and cohort designs are as in Section 2.1.

This general Model (9) agrees with the population-averaged marginal model used by Li *et al.* (2018b) for SWDs:

$$E(Y_{ijk}) = \mu_{ijk} = \mu + X_{ij}\beta + \gamma_j \tag{10}$$

where μ_{ijk} is the marginal expected mean response of Y_{ijk}. Our software allows for SWDs both with time effects (j in $1, \cdots, J$, and $\gamma_1 = 0$ for identifiability) and without time effects (all $\gamma_j = 0$).

Li *et al.* (2018b) and Hooper *et al.* (2016) both considered scenarios for continuous outcomes that accommodate three correlation parameters, under marginal and conditional model,

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 $JK_i * (J+1)$ 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.

Restrictions on the allowable parameter space The correlations are within $[-1, 1]$. When two or three correlation parameters are used for cross-sectional and cohort designs, respectively, a positive definite correlation matrix R_i should be ensured.

2.3. Remarks

This software includes procedures for power calculations in SWDs with binary outcomes under the 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. The treatment effect β is the same under the conditional and marginal model with the identity and log link functions (Ritz and Spiegelman 2004). In Appendix A, by proving the equivalence of the variance of the intervention effect in the conditional and marginal models with continuous outcomes, we showed that Li *et al.* (2018b) incorporated the cross-sectional designs in Hussey and Hughes (2007) and the cohort designs in Hooper *et al.* (2016), and obtained the same power formulae for the conditional and marginal models for each of these designs. Hence, we can directly implement Li *et al.* (2018b) in our software for all of these continuous outcome scenarios.

Regarding the cross-sectional settings where $\alpha_0 = \alpha_1$ with binary outcomes, we implement both the conditional (Zhou *et al.* 2020) and the marginal (Li *et al.* 2018b) models, for which the variance of these two estimates are different. As shown by Ritz and Spiegelman (2004), the treatment effect, β, is the same for these two models under the identity and log link functions given the same mean response rates. However, the estimates of the conditional and marginal model have different variances, and we found in numerical calculations that neither is globally greater than the other. Under the logit link, the parameter β has different interpretations for the two models given the same mean response rates, and again, neither has a greater power than the other although variance comparisons are not meaningful with parameters having different interpretations. Figure 1 provides some examples of this pattern, which is expected because GLMM is likelihood based, while the marginal models relies on unbiased estimating equations.

We emphasize that our software considers designs 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), time effects are assumed. However, Zhou *et al.* (2017) argued that SWDs are mostly used in the study of relatively short-term outcomes with short-term interventions, thus, it is reasonable to assume no time effects in the primary analysis. Zhou *et al.* (2020) developed their method considering cases with and without time effects. When developing this software, we include and generalize all the methods to cases with and without time effects.
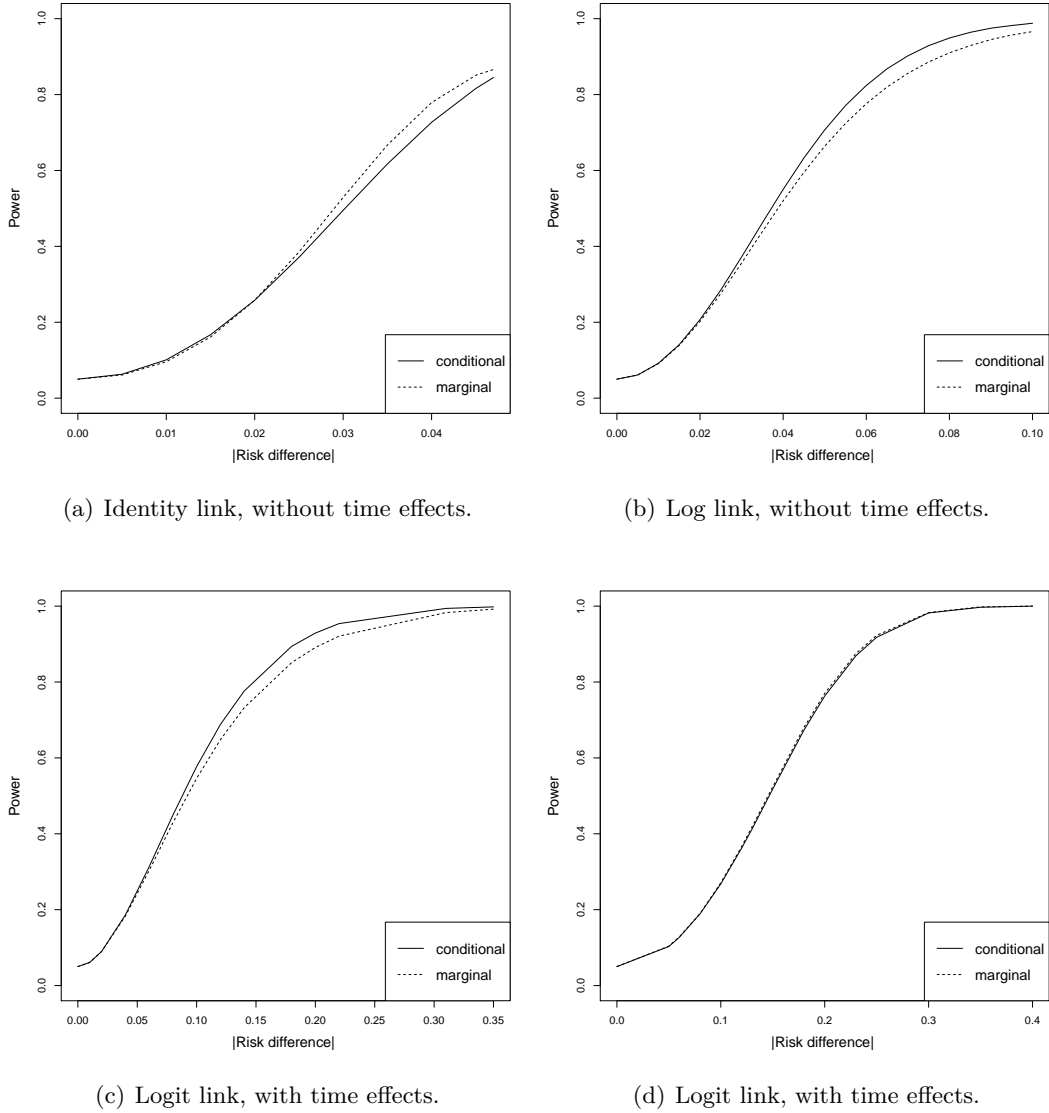
(a) Identity link, without time effects.

(b) Log link, without time effects.

(c) Logit link, with time effects.

(d) Logit link, with time effects.

Figure 1: Power vs. the absolute value of risk difference between the control and intervention group at the end of the cross-sectional study, for I = 6, J = 3, K = 40, and $\alpha_0 = \alpha_1 = 0.02$. There are no time effects in the first row. Figure 1(a) has the identity link and the baseline mean response rate in the control group of 0.05. The conditional model has a slightly larger power when the absolute value of risk difference is smaller, the opposite conclusion is shown when the absolute value of risk difference is larger. Figure 1(b) has the log link and the baseline mean response rate in the control group of 0.045. The conditional and marginal models have almost the same power when risk difference is smaller, and conditional model has a larger power when risk difference increases. The second row considers time effects and the logit link. In Figure 1(c), the baseline mean response rate in the control group is 0.03 and the mean response rate in the control group at the end of the study is 0.04. The conditional model has a larger power as risk difference increases. In Figure 1(d), the baseline mean response rate in the control group is 0.3 and the mean response rate in the control group at the end of the study is 0.4. The marginal model has a larger power as risk difference increases.

		Conditional model	Marginal model
Binary outcomes (identity, log, logit links)	Cross-sectional design	Zhou *et al.* (2020)	Li *et al.* (2018b)
	Cohort design	N/A	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: Methods implemented in **swdpwr**.

3. Software Description

Table 1 displays all the scenarios that are implemented in the software, accommodating cases and methods with and without time effects. The input parameters are the same for R and SAS. Different input parameters values, including for the anticipated mean response rate in the control group at the start of the study, the anticipated mean response rate in the control group at the end of the study, the anticipated mean response rate in the intervention group at the end of the study, the study design, and the intraclass correlations can be specified based on the preliminary data. The arguments of software **swdpwr** are shown in Table 2.

In this section, we give a detailed description for the implementation in R and SAS. Examples under different scenarios will be discussed in Section 4. We assume balanced cluster-period sizes so that $K_{ij} = K$ over all i and j.

3.1. Supply of regression model parameters

To account for time effects, we assume that the difference of time effects between neighbouring time periods is constant, which implies that the absolute value of γ_j increases linearly in j. Hence, parameter for time effects γ_J is a single parameter. This software considers two complementary ways for users to supply the regression model parameters μ, β and γ_J. One approach is to specify `meanresponse_start`, `meanresponse_end0` and `meanresponse_end1`, which will help identify the baseline outcome rate on the scale of the link function μ, the treatment effect β and the time effect γ_J on the link function scale. The alternative approach is to specify `meanresponse_start`, `meanresponse_end0` and `effectsize_beta`. This approach provides the treatment effect β directly, while μ and γ_J can also be identified from the input parameters.

It was remarked in Li *et al.* (2018b) that the power calculations for continuous outcomes do not depend on μ and γ_J, hence users are allowed to get the results by just providing the accurate parameter `effectsize_beta` if it is continuous outcome and assumes presence of time effects. However, users will additionally need to specify different values for `meanresponse_start` and `meanresponse_end0` to indicate the presence of time effects, although the values do not require to be accurate. If assuming no time effects and continuous outcomes, Equation (26) in Appendix A.2 also shows the independence of power on μ, and users can only supply

`effectsize_beta` to conduct the calculation, with the default of no time effects.

In case of any contradiction of parameter specification, either one of the approaches for parameter specification is allowed in the software only. Users are not allowed to supply `meanresponse_start`, `meanresponse_end0`, `meanresponse_end1` and `effectsize_beta` at the same time.

3.2. Description of the R package

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

```
swdpower(K, design, family = "binomial", model = "conditional",
link = "identity", type = "cross-sectional", meanresponse_start = NA,
meanresponse_end0 = meanresponse_start, meanresponse_end1 =  NA,
effectsize_beta = NA, sigma2 = 0, typeIerror = 0.05, alpha0 = 0.1,
alpha1 = alpha0/2, alpha2 = NA)
```

with details provided in Table 2 along with defaults if any (those above with an equal sign have default values). The input argument `design` is a matrix generated in R with elements of 0 (control) or 1 (intervention) for users to define the detail of the SWD, in which each row represents a cluster and each column represents a time period. The argument `sigma2` is only allowed for continuous outcomes. The argument `alpha2` should not be an input under cross-sectional designs although it is numerically identical to `alpha1` in this scenario by definition. The objects returned by `swdpower` function are the design matrix, summary features of the design and the power for the scenario specified.

3.3. 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

Argument	Description	Default
K	number of individuals at each time period in a cluster	
design	I*J dimensional data set that describes the study design	
family	specify family="gaussian" for continuous outcome, family="binomial" for binary outcome	"binomial"
model	specify model="conditional" for conditional model, model="marginal" for marginal model	"conditional"
link	choose link function from link="identity", link="log" and link="logit"	"identity"
type	specify type="cohort" for closed cohort study, type="cross-sectional" for cross-sectional study	"cross-sectional"
meanresponse_start	the anticipated mean response rate in the control group at the start of the study	NA
meanresponse_end0	the anticipated mean response rate in the control group at the end of the study	meanresponse_start
meanresponse_end1	the anticipated mean response rate in the intervention group at the end of the study	NA
effectsize_beta	the anticipated effect size	NA
sigma2	marginal variance of the outcome (only allowed if is continuous outcome)	0
typeIerror	Type I error	0.05
alpha0	within-period correlation α_0	0.1
alpha1	between-period correlation α_1	alpha0/2
alpha2	within-individual correlation α_2 (only allowed if type="cohort")	NA

Table 2: Arguments of software **swdpwr**.

```
% macro swdpwr(K, design, family = "binomial", model = "conditional",
link = "identity", type = "cross-sectional", meanresponse_start = NA,
meanresponse_end0 = meanresponse_start, meanresponse_end1 =  NA,
effectsize_beta = NA, sigma2 = 0, typeIerror = 0.05, alpha0 = 0.1,
alpha1 = alpha0/2, alpha2 = NA)
```

with details given in Table 2 together with defaults if any. 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. Each section will contain examples under both platforms.

4.1. Conditional model with binary outcome and cross-sectional design

Zhou *et al.* (2020) proposed this maximum likelihood method under cross-sectional settings, which discussed scenarios with time effects and without time effects. With binary outcomes, the variance argument `sigma2` is undefined and should not be specified. Due to this specific cross-sectional setting, we specify $\alpha_0 = \alpha_1$ in the function and α_2 is also undefined as discussed in Section 2.1.

R package   The study design should be supplied by an I*J matrix, where I is the number of clusters, J is the number of time periods. We will utilize the `rep()` function to specify the elements of the study design matrix by providing different allocations and repeated times of each allocation. More details can be found in the following examples. Fitting the model with all required inputs 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(K = 50, design = dataset, family = "binomial",
+ model = "conditional", link = "identity", type = "cross-sectional",
+ meanresponse_start = 0.2, meanresponse_end0 = 0.25,
+ meanresponse_end1 = 0.38, typeIerror = 0.05, alpha0 = 0.01,  alpha1 = 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"
J                  "3"
K                  "50"
total sample size  "1800"
```

```
study type              "cross-sectional"
family of outcomes      "binomial"
model                   "conditional"
link                    "identity"
baseline  mu            "0.2"
treatment effect beta   "0.13"
time effect gamma_J     "0.05"
alpha0                  "0.01"
alpha1                  "0.01"
alpha2                  "0.01"
Type I error            "0.05"
Power                   "0.899"
```

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 for a total sample size of 1800. The calculation was conducted utilizing the conditional model with time effects and the identity link function. The intraclass correlations, here interpreted as the correlation between different individuals within the same cluster, were all set as 0.01 and the Type I error was 0.05. The power obtained from `swdpower` for this scenario is 0.899 for the alternative hypothesis $\beta_A = 0.13$.

The model can also be considered under other link functions such as log and logit link functions. For instance:

```
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(K = 50, design = dataset, family = "binomial",
+ model = "conditional",  link = "logit", type = "cross-sectional",
+ meanresponse_start = 0.2,  meanresponse_end0 = 0.25,
+ meanresponse_end1 = 0.38, typeIerror = 0.05, alpha0 = 0.01,  alpha1 = 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"
J                      "3"
K                      "50"
total sample size      "1800"
study type             "cross-sectional"
family of outcomes     "binomial"
model                  "conditional"
link                   "logit"
baseline  mu           "-1.405"
treatment effect beta  "0.616"
time effect gamma_J     "0.291"
alpha0                 "0.01"
alpha1                 "0.01"
alpha2                 "0.01"
Type I error           "0.05"
Power                  "0.838"
```

This example assumes a conditional model with time effects and the logit link function, with other inputs the same as the above example. R package gives the power of 0.838 for the alternative hypothesis $\beta_A = 0.616$.

SAS macro   The maro %swdpwr is used in SAS. The same examples are illustrated in SAS as above. As in R, %swdpwr in SAS requires that the input of a design matrix and a SAS data set exmple as follows is a pre-requisite:

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

The call to %swdpwr is:

```
%swdpwr(K = 50, design = example, family = "binomial", model = "conditional",
link = "identity", type = "cross-sectional", meanresponse_start = 0.2,
meanresponse_end0 = 0.25,  meanresponse_end1 = 0.38, typeIerror = 0.05,
alpha0 = 0.01, alpha1 = 0.01)

%swdpwr(K = 50, design = example, family = "binomial", model = "conditional",
link = "logit", type = "cross-sectional", meanresponse_start = 0.2,
meanresponse_end0 = 0.25,  meanresponse_end1 = 0.38, typeIerror = 0.05,
alpha0 = 0.01, alpha1 = 0.01)
```

The data set example should be generated as described in Section 3. The inputs of the SAS are the same as the R examples. Calculations were based on the conditional model with time

effects. The same results are obtained as in the R examples above. SAS gives the power of 0.899 under this scenario with the identity link function for the alternative hypothesis $\beta_A = 0.13$, and the power of 0.838 under this scenario with the logit link function for the alternative hypothesis $\beta_A = 0.616$.

If considering scenarios without time effects, users can supply the same value for `meanresponse_start` and `meanresponse_end0`.

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

A method for the marginal model based on GEE was proposed in Li *et al.* (2018b), employing a block exchangeable correlation structure applicable to both cohort and cross-sectional designs. This method accommodates scenarios with and without time effects as well as three link functions. When the outcome is binary, because the variance is a strict function of the mean, the marginal variance argument `sigma2` is undefined.

R package with cohort design   We first utilize a cohort design example to illustrate the use of the software in R for function `swdpower` in package **swdpwr**. A cohort design requires the specification of α_0, α_1 and α_2.

```
R> library("swdpwr")
R> dataset = matrix(c(rep(c(0, 1, 1, 1), 6), rep(c(0, 0, 1, 1), 6)), 12, 4,
+ byrow = TRUE)
R> swdpower(K = 100, design = dataset, family = "binomial",
+ model = "marginal", link = "log", type = "cohort",
+ meanresponse_start = 0.156, meanresponse_end0 = 0.1765,
+ effectsize_beta = 0.75, typeIerror = 0.05, alpha0 = 0.03, alpha1 = 0.015,
+ alpha2 = 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    1    1    1
 [6,]    0    1    1    1
 [7,]    0    0    1    1
 [8,]    0    0    1    1
 [9,]    0    0    1    1
[10,]    0    0    1    1
[11,]    0    0    1    1
[12,]    0    0    1    1


$Summary
                       [,1]
I                      "12"
```

```
J                      "4"
K                      "100"
total sample size      "1200"
study type             "cohort"
family of outcomes     "binomial"
model                  "marginal"
link                   "log"
baseline  mu           "-1.858"
treatment effect beta  "0.75"
time effect gamma_J     "0.123"
alpha0                 "0.03"
alpha1                 "0.015"
alpha2                 "0.2"
Type I error           "0.05"
Power                  "0.983"
```

This example illustrates a cluster randomized cohort SWD including 100 individuals in each cluster, with 12 clusters and 4 time periods for a total sample size of 1200. We conducted the calculation using the marginal model under the log link function with time effects. The three correlation parameters were 0.03, 0.015 and 0.2. The power calculated given by R for this scenario is around 0.983 for the alternative hypothesis $\beta_A = 0.75$.

Different link functions can be considered for the calculation. For instance, with the same regression model parameters and other inputs but under the logit link function, the power obtained is then 0.843 for the alternative hypothesis $\beta_A = 0.75$.

```
R> library("swdpwr")
R> dataset = matrix(c(rep(c(0, 1, 1, 1),  6), rep(c(0, 0, 1, 1), 6)), 12, 4,
+ byrow = TRUE)
R> swdpower(K = 100, design = dataset, family = "binomial",
+ model = "marginal", link = "logit", type = "cohort",
+ meanresponse_start = 0.1349,  meanresponse_end0 = 0.1499,
+ effectsize_beta = 0.75, typeIerror = 0.05, alpha0 = 0.03, alpha1 = 0.015,
+ alpha2 = 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    1    1    1
 [6,]    0    1    1    1
 [7,]    0    0    1    1
 [8,]    0    0    1    1
 [9,]    0    0    1    1
[10,]    0    0    1    1
[11,]    0    0    1    1
```

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

```
$Summary
                        [,1]
I                       "12"
J                       "4"
K                       "100"
total sample size       "1200"
study type              "cohort"
family of outcomes      "binomial"
model                   "marginal"
link                    "logit"
baseline  mu            "-1.858"
treatment effect beta   "0.75"
time effect gamma_J      "0.123"
alpha0                  "0.03"
alpha1                  "0.015"
alpha2                  "0.2"
Type I error            "0.05"
Power                   "0.843"
```

SAS macro with cohort design Next, we illustrate the use of SAS macro `%swdpwr` for cohort design examples and a SAS data set `design2` is required:

```
data design2;
input numofclusters time1 time2 time3 time4;
cards;
6 0 1 1 1
6 0 0 1 1
;
run;
```

To implement the same examples in R as above, the macro called to do power calculations would be:

```
% swdpwr(K = 100, design = design2, family = "binomial", model = "marginal",
link = "log", type = "cohort", meanresponse_start = 0.156,
meanresponse_end0 = 0.1765, effectsize_beta = 0.75, typeIerror = 0.05,
alpha0 = 0.03, alpha1 = 0.015, alpha2 = 0.2)
```

```
% swdpwr(K = 100, design = design2, family = "binomial", model = "marginal",
link = "logit", type = "cohort", meanresponse_start = 0.1349,
meanresponse_end0 = 0.1499, effectsize_beta = 0.75, typeIerror = 0.05,
alpha0 = 0.03, alpha1 = 0.015, alpha2 = 0.2)
```

Based on the sample stepped wedge cohort trial and conducting design using the marginal model with time effects, we obtain the same calculated power as above: power of 0.983 under

the log link function and power of 0.843 under the logit link function for the alternative hypothesis $\beta_A = 0.75$.

R package with cross-sectional design   We next consider the cross-sectional design under the marginal model, where α_0 and α_1 should be specified. The example conducted by swdpower in R package **swdpwr**:

```
R> library("swdpwr")
R> dataset = matrix(c(rep(c(0, 1, 1, 1), 6), rep(c(0, 0, 1, 1), 6)), 12, 4,
+ byrow = TRUE)
R> swdpower(K = 100, design = dataset, family = "binomial",
+ model = "marginal", link = "identity", type = "cross-sectional",
+ meanresponse_start = 0.15, meanresponse_end0 = 0.15,
+ meanresponse_end1  = 0.2, typeIerror = 0.05, alpha0 = 0.02, alpha1 = 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    1    1    1
 [6,]    0    1    1    1
 [7,]    0    0    1    1
 [8,]    0    0    1    1
 [9,]    0    0    1    1
[10,]    0    0    1    1
[11,]    0    0    1    1
[12,]    0    0    1    1

$Summary
                        [,1]
I                       "12"
J                       "4"
K                       "100"
total sample size       "4800"
study type              "cross-sectional"
family of outcomes      "binomial"
model                   "marginal"
link                    "identity"
baseline  mu            "0.15"
treatment effect beta   "0.05"
time effect gamma_J     "0"
alpha0                  "0.02"
alpha1                  "0.015"
alpha2                  "0.015"
```

```
Type I error          "0.05"
Power                 "0.946"
```

Above is a cross-sectional SWD with 100 individuals at each time period in each cluster for a total sample size of 4800. The design was conducted on the marginal model under identity link function without time effects. In this example, the powe given by R is around 0.946 for the alternative hypothesis $\beta_A = 0.05$.

SAS macro with cross-sectional design   To illustrate the cross-sectional settings in SAS, we use the same example in R as above. Similarly, we use the data set `design2` generated previously as the study design.

The macro called to do power calculations would be:

```
%swdpwr(K = 100, design = design2, family = "binomial", model = "marginal",
link = "identity", type = "cross-sectional", meanresponse_start = 0.15,
meanresponse_end0 = 0.15,  meanresponse_end1  = 0.2, typeIerror = 0.05,
alpha0 = 0.02, alpha1 = 0.015)
```

With a total sample size of 2400, the power under this scenario is 0.946 for the alternative hypothesis $\beta_A = 0.05$, which accords with the results in R.

4.3. Method with continuous outcomes under identity link

Linear mixed effects models 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 for continuous outcomes under the identity link function. Because it has been shown that conditional and marginal models are equivalent when the regression is linear with mean 0 random effect and the identity or log link functions (Ritz and Spiegelman 2004), with the addition of a new proof for equivalence of variance of treatment effect in Appendix A, the procedures in Li *et al.* (2018b) are used for power calculations with continuous outcomes, with three levels of correlation parameters are considered.

The designs for continuous outcomes are conducted under the identity link function. Both cross-sectional and cohort designs with and without time effects can be accommodated in our software. Here, the argument `sigma2` must be specified, which equals σ_t^2 as in Section 2.2.

R package   An 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(K = 24, design = dataset, family = "gaussian",
+ model = "marginal", link = "identity", type = "cohort",
+ meanresponse_start = 0.1, meanresponse_end0 = 0.2,
+ effectsize_beta = 0.2, sigma2 = 0.095,  typeIerror = 0.05, alpha0 = 0.03,
```

```
+ alpha1 = 0.015, alpha2 = 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"
J                        "3"
K                        "24"
total sample size        "192"
study type               "cohort"
family of outcomes       "gaussian"
model                    "marginal"
link                     "identity"
treatment effect beta    "0.2"
time effect gamma_J       "0.1"
marginal variance        "0.095"
alpha0                   "0.03"
alpha1                   "0.015"
alpha2                   "0.2"
Type I error             "0.05"
Power                    "0.965"
```

In this example trial, the outcome is continuous with total marginal variance of 0.095, conducted in 8 clusters and 3 time periods for a total sample size of 192. Marginal model 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 $\beta_A = 0.2$.

We also consider the same example without time effects:

```
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(K = 24, design = dataset,  family= "gaussian",
+ model = "marginal",  link = "identity", type = "cohort",
+ effectsize_beta = 0.2, sigma2 = 0.095,  typeIerror = 0.05, alpha0 = 0.03,
+ alpha1 = 0.015, alpha2 = 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"
J                       "3"
K                       "24"
total sample size       "192"
study type              "cohort"
family of outcomes      "gaussian"
model                   "marginal"
link                    "identity"
treatment effect beta   "0.2"
time effect gamma_J     "0"
marginal variance       "0.095"
alpha0                  "0.03"
alpha1                  "0.015"
alpha2                  "0.2"
Type I error            "0.05"
Power                   "1"
```

In this example, only `effectsize_beta` is specified and the trial was conducted without time effects by default. The calculated power is around 1 for the alternative hypothesis $\beta_A = 0.2$.

SAS macro Next, we illustrate the R examples above 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(K = 24, design = design3, family = "gaussian", model = "marginal",
link = "identity", type = "cohort", meanresponse_start = 0.1,
```

```
meanresponse_end0 = 0.2, effectsize_beta = 0.2, sigma2 = 0.095,
typeIerror = 0.05, alpha0 = 0.03, alpha1 = 0.015, alpha2 = 0.2)

%swdpwr(K = 24, design = design3, family = "gaussian", model = "marginal",
link = "identity", type = "cohort", effectsize_beta = 0.2, sigma2 = 0.095,
typeIerror = 0.05, alpha0 = 0.03, alpha1 = 0.015, alpha2 = 0.2)
```

The same results were obtained as in R, with the power of around 0.965 considering time effects and 1 considering no time effects, respectively, for the alternative hypothesis $\beta_A = 0.2$.

4.4. Warnings and error messages generation

When the input arguments are inappropriate, warnings and error messages may occur in the software. This section will give an illustration of these scenarios and advise users how to correct for the errors.

The program will output reliable results although warnings are returned due to some invalid input arguments, including mainly five scenarios. First, conditional model with binary outcomes only allows for cross-sectional settings, so if `"cohort"` is specified for the argument `type`, a warning message will occur and `type="cross-sectional"` will be forced to resume the power calculation. Second, $\alpha_0 = \alpha_1$ is to be ensured for conditional model with binary outcome and cross-sectional designs, if it is violated, the value of α_1 is set to the value of α_0 and a warning will be returned. Third, for cross-sectional designs, if α_2 is specified, it will be ignored and a warning message will remind users that α_2 is undefined and should not be an input. Fourth, when `sigma2` is supplied for binary outcomes, it will be ignored and the software will return a warning that marginal variance should not be specified for binary outcomes. Last, if it is continuous outcome supplied by link functions of logit or log, a warning message will explain that the program will be conducted with the link function forced to be identity.

There are also cases where error messages occur and the program will be stopped, then users will need to revise the input arguments. The most common errors will be missing of input arguments. However, users will need to notice that although the software includes two complementary ways to supply regression model parameters as in Section 3.1, an error will be generated if users specify `meanresponse_start`, `meanresponse_end0`, `meanresponse_end1` and `effectsize_beta` simultaneously, in case that contradictions occur for the value of parameters. Besides, the specification of input arguments cannot be identified when typos occur in arguments such as `family`, `model`, `link` and `type`, then errors will be reported.

When input parameters, including the correlation parameters, Type I error rate, or response for binary outcomes, are out of range, an error message will be returned. Before giving the detailed examples, we discuss the specification of correlation parameters for both cross-sectional and cohort designs. As described in Section 2.1, these correlations represent the within-period, between-period, and within-individual correlations, respectively, with the natural restriction of within $[-1, 1]$. For the correlation structure of SWDs, Li *et al.* (2018a) identified four linear eigenvalue constraints to ensure a positive definite correlation matrix. These constraints are enforced in our software for all models and both outcome types. In a cross-sectional design, because there are no repeated measures within subjects, α_2 is not required and the block exchangeable correlation structure reduces to the nested exchangeable structure as in Teerenstra, Lu, Preisser, Van Achterberg, and Borm (2010). Thus, in cross-

sectional designs, a single intra-cluster correlation coefficient which measures the correlation between individuals in the same cluster describes the correlation structure in Zhou *et al.* (2020) such that the working correlation reduces to $\alpha_0 = \alpha_1$.

Additional restrictions for α_0, α_1 and α_2 are needed as the marginal and conditional means limit the ranges of correlation due to binary outcomes (Qaqish 2003), where $i \neq j$:

$$max(0, \mu_i + \mu_j - 1) \leq E[Y_i Y_j] \leq min(\mu_i, \mu_j) \tag{11}$$

This can be written to require that, the following three criteria must be satisfied over $i = 1, 2, ..., I; j, l = 1, 2, ..., J; k, m = 1, 2, ..., K$:

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

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

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

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, and $j \neq l$ plus $k \neq m$. These restrictions apply to both the conditional and marginal model with binary outcomes.

Now, we give examples for scenarios where representative errors occur in R, so that users can check the input parameter according to the error message and obtain suggestions about how to revise them. The same errors will occur if conducted in SAS as well.

In the first example, the requirement that the 2- or 3-way working correlation matrix R_i is positive definite is violated as correlation parameters exceed plausible ranges although they are still within $[0, 1]$. This error could occur for both outcome types and in all models.

```
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(K = 100, design = dataset, family = "gaussian",
+ model = "marginal", link = "identity", type = "cohort",
+ effectsize_beta = 0.05, sigma2 = 0.095, typeIerror = 0.05, alpha0 = 0.015,
+ alpha1 = 0.2, alpha2 = 0.1)
Error in swdpower(K = 100, design = dataset, family = "gaussian",
model = "marginal", : Correlation matrix R is not positive definite. Please
check whether the between-period correlation is unrealistically larger
than the within-period correlation or the within-individual correlation.
```

The fix to this error can be revising `alpha1` to be much smaller, then the program will be processed without errors:

```
R> swdpower(K = 100, design = dataset, family = "gaussian",
+ model = "marginal", link = "identity", type = "cohort",
+ effectsize_beta = 0.05, sigma2 = 0.095, typeIerror = 0.05, alpha0 = 0.015,
+ alpha1 = 0.01, alpha2 = 0.1)
```

```
$`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"
J                        "4"
K                        "100"
total sample size        "1200"
study type               "cohort"
family of outcomes       "gaussian"
model                    "marginal"
link                     "identity"
treatment effect beta    "0.05"
time effect gamma_J      "0"
marginal variance        "0.095"
alpha0                   "0.015"
alpha1                   "0.01"
alpha2                   "0.1"
Type I error             "0.05"
Power                    "0.994"
```

The second example occurred with a binary outcome. This error is also related to the correlations values given because they are additionally restricted by means due to binary outcomes (Qaqish 2003; Ridout *et al.* 1999) as discussed before.

```
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(K = 100, design = dataset, family = "binomial",
+ model = "marginal", link = "identity", type = "cohort",
+ meanresponse_start = 0.1, meanresponse_end0 = 0.2, effectsize_beta = 0.7,
+ typeIerror = 0.05, alpha0 = 0.1, alpha1 = 0.05, alpha2 = 0.2)
Error in swdpower(K = 100, design = dataset, family = "binomial",
```

```
model = "marginal", : Correlation parameters do not satisfy the
restrictions of Qaqish (2003). Please check whether it is possible to reduce
the effect size, or make adjustments to the intraclass correlations.
```

This error can be fixed by revising the intraclass correlations, for example:

```
R> swdpower(K = 100, design = dataset, family = "binomial",
+ model = "marginal", link = "identity", type = "cohort",
+ meanresponse_start = 0.1, meanresponse_end0 = 0.2, effectsize_beta = 0.7,
+ typeIerror = 0.05, alpha0 = 0.05, alpha1 = 0.05, alpha2 = 0.1)

$`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"
J                      "4"
K                      "100"
total sample size      "1200"
study type             "cohort"
family of outcomes      "binomial"
model                  "marginal"
link                   "identity"
baseline mu            "0.1"
treatment effect beta "0.7"
time effect gamma_J    "0.1"
alpha0                 "0.05"
alpha1                 "0.05"
alpha2                 "0.1"
Type I error           "0.05"
Power                  "1"
```

The third example also pertains to a binary outcome. Sometimes, the input parameters related to mean responses may exceed the range for valid probability between $[0, 1]$. This error is easy to handle with and is not relevant to continuous outcomes. 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(K = 100, design = dataset, family = "binomial",
+ model = "marginal", link = "identity", type = "cohort",
+ meanresponse_start = 0.1, meanresponse_end0 = 0.2, effectsize_beta = 0.9,
+ typeIerror = 0.05, alpha0 = 0.05, alpha1 = 0.05, alpha2 = 0.1)
Error in swdpower(K = 100, design = dataset, family = "binomial",
model = "marginal", : Violation of valid probability, given input parameters:
max(meanresponse_start, meanresponse_end0, meanresponse_end1)>1. Please
check whether any of these values are out of range and revise one or more of
them.
```

The fourth example occurs only in scenarios of binary outcomes under conditional models with time effects. It is discussed in Zhou *et al.* (2020) that as the sample size K for each cluster at each time period increases, the estimated running time increases prohibitively to beyond what is possible even for high performance computing under these scenarios. Hence, we set an upper limit of 150 for K in the current software. However, even with values of K within this allowable range, computing time may still be low. To address this, users can consider switching to the marginal model specification or dropping time effects. An example for this error is:

```
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(K = 160, design = dataset, family = "binomial",
+ model = "conditional", link = "logit", type = "cross-sectional",
+ meanresponse_start = 0.1, meanresponse_end0 = 0.2, effectsize_beta = 0.6,
+ typeIerror = 0.05, alpha0 = 0.05, alpha1 = 0.05)
Error in swdpower(K = 100, design = dataset, family = "binomial",
model = "conditional", : K should be at least smaller than 150 for this
scenario as the running time is too long with this K for the power
calculation of binary outcomes under conditional model with time effects.
Please reduce K or use the model without time effects or use
marginal models.
```

In addition, the intraclass correlations and Type I error must be between 0 and 1 as well for all models, 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(K = 100, design = dataset, family = "binomial",
+ model = "marginal", link = "identity", type = "cohort",
+ meanresponse_start = 0.1, meanresponse_end0 = 0.2, effectsize_beta = 0.5,
+ typeIerror = 0.05, alpha0 = 1.1, alpha1 = 0.05, alpha2 = 0.1)
Error in swdpower(K = 100, design = dataset, family = "binomial",
```

```
model = "marginal",  : Violate range of intraclass correlations:
max(alpha0,alpha1,alpha2)>1. Please correct the values of correlation
parameters, they must be between 0 and 1.

R> swdpower(K = 100, design = dataset, family = "binomial",
+ model = "marginal", link = "identity", type="cohort",
+ meanresponse_start = 0.1, meanresponse_end0 = 0.2, effectsize_beta = 0.5,
+ typeIerror = 1.05, alpha0 = 0.1, alpha1 = 0.05, alpha2 = 0.1)
Error in swdpower(K = 100, design = dataset, family = "binomial",
model = "marginal",  : Type I error provided is larger than 1, it must be
between 0 and 1, and is usually 0.05.
```

5. Application

The examples in the previous section are based on hypothetical examples. Now, we provide two real world applications, one using the marginal model and the other using conditional model.

The Washington Expedited Partner Therapy (EPT) trial was a community-based trial employing a cluster randomized SWD for promoting EPT. The outcome was Chlamydia status, a binary variable. 24 local health jurisdictions (LHJs) were included in this trial and each represented a cluster. There were 5 time periods and the intervention was initiated at four of them, with 6 clusters entering the intervention group at each time period. We can describe this design in the 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 used a generalized linear mixed model under the log link function 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 for a total sample size of 19440. Based on preliminary data, the baseline prevalence of Chlamydia was about 0.05, the cluster effect coefficient of variation was 0.3, the Type I error was set of 0.05, and a prevalence ratio of 0.7 was to be tested. The coefficient of variation is defined to characterize the cluster effects on the variance of responses (Hayes and Bennett 1999) and is closely related to regular intraclass correlation. Following Hussey and Hughes (2007) that the coefficient of variation was τ/μ', where μ' is denoted as the baseline prevalence of Chlamydia 0.05, the intraclass correlation $\frac{\tau^2}{\tau^2+\mu'(1-\mu')}$ was approximately 0.0047 and hence in this study $\alpha_0 = \alpha_1 = 0.0047$. We also assumed the presence of time effects.

Result
I = 24
J = 5
K = 162
Total sample size = 19440
Family = binomial
Model = marginal
Link = log
Type = cross-sectional
Baseline (mu): -2.996
Treatment effect (beta): -0.336
Time effect (gamma J): -0.020
alpha0: 0.005
alpha1: 0.005
alpha2: 0.005
Type I error = 0.050
Power = 0.812

Table 3: Output from macro `%swdpwr` for EPT trial

The input parameters were calculated and summarized based on these information. Hussey and Hughes (2007) used the conditional model and the approximation method for binary outcomes, here, we conducted this design using the marginal model, as follows:

```
%swdpwr(K = 162, design = ept, family = "binomial", model = "marginal",
link = "log", type = "cross-sectional", meanresponse_start = 0.05,
meanresponse_end0 = 0.049, meanresponse_end1 = 0.035, typeIerror = 0.05,
alpha0 = 0.0047, alpha1 = 0.0047)
```

We obtained power of 0.812 for the alternative hypothesis $\beta_A = -0.336$ from the marginal model in Table 3, which corresponds to the anticipated power of 80% from similar scenarios with the condiitonal model considered in Golden *et al.* (2015) and Hussey and Hughes (2007).

The Tanzania PPIUD study utilized a SWD to assess the causal effect of a PPIUD intervention on subsequent pregnancy (Canning, Shah, Pearson, Pradhan, Karra, Senderowicz, Bärnighausen, Spiegelman, and Langer 2016). The binary outcome was defined as a current or terminated pregnancy 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;
```

Result
I = 6
J = 4
K = 120
Total sample size = 2880
Family = binomial
Model = conditional
Link = identity
Type = cross-sectional
Baseline (mu): 0.240
Treatment effect (beta): -0.046
Time effect (gamma J): 0.000
alpha0: 0.150
alpha1: 0.150
alpha2: 0.150
Type I error = 0.050
Power = 0.846

Table 4: Output from macro `%swdpwr` for PPIUD study

A generalized linear mixed model under identity link without time effects was 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 for a total sample size of 2880. As in Canning *et al.* (2016), the baseline proportion of pregnancy was assumed to be 0.24, the intraclass correlation was 0.15, the Type I error was set of 0.05, and a prevalence ratio of around 0.8 was to be tested. We estimated from preliminary data that the effect size was -0.046, assuming no time effects. We conducted this design using the conditional model in SAS. The macro called for this power calculation would be:

```
%swdpwr(K = 120, design = PPIUD, family = "binomial", model = "conditional",
link = "identity", type = "cross-sectional", meanresponse_start = 0.24,
meanresponse_end0 = 0.24, effectsize_beta = -0.046, typeIerror = 0.05,
alpha0 = 0.15, alpha1 = 0.15)
```

We get the power 0.846 for the alternative hypothesis $\beta_A = -0.046$ from the conditional model in Table 4. 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 the R package **swdpwr** and the SAS macro `%swdpwr` for power calculations in SWDs. The software is designed under two computer platforms where users specify input parameters for different scenarios of interest, accommodating cross-sectional and cohort designs, binary and continuous outcomes, marginal and conditional models, three link functions, with and without time effects. This software addresses the implementation gap between newly proposed methodology and their application to obtain

more accurate power calculation for binary outcomes in SWDs, rather than relying on the approximation method in Hussey and Hughes (2007). The core of the software is developed in Fortran to ensure efficient computations, 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 calculations may become time-consuming, thus we have implemented some methods to decrease the computational cost. For example, we applied a simplified closed-form expression for matrix inversion under the marginal model, as derived in Li *et al.* (2018b). Package **swdpwr** is available via the Comprehensive R Archive Network (CRAN) as a contributed package at: `https://CRAN.R-project.org/package=swdpwr`. The package is also available as a **shiny** (Chang, Cheng, Allaire, Xie, and McPherson 2019) app, available online at: `https://jiachenchen322.shinyapps.io/swdpwr_shinyapp/` to enable users without SAS or R programming skills to also use this software easily.

In the future, we plan to accommodate various numbers of individuals per cluster and even at each time period to make the software even more flexible. Furthermore, we are working to implement the partition method developed in Zhou *et al.* (2020) for the conditional model with binary outcomes to permit power and sample size calculations with cluster sizes, K, greater than 150 as occurs in practice. In addition, current software only accommodates nested exchangeable and block exchangeable correlation structures, which could be further extended to allow for damped exponential correlation structures (Munoz, Carey, Schouten, Segal, and Rosner 1992) and proportional decay correlation structures (Li 2020).

Acknowledgments

References

Baio G, Copas A, Ambler G, Hargreaves J, Beard E, Omar RZ (2015). "Sample Size Calculation for a Stepped Wedge Trial." *Trials*, **16**(1), 354. `doi:10.1186/s13063-015-0840-9`.

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*. R package version 1.5.0, 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*. R package version 3.1, 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`.

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`.

Munoz A, Carey V, Schouten JP, Segal M, Rosner B (1992). "A Parametric Family of Correlation Structures for the Analysis of Longitudinal Data." *Biometrics*, pp. 733–742. `doi:10.2307/2532340`.

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 the variance for continuous outcomes

In this paper, we include procedures for power calculations in SWDs with continuous outcomes based on Li *et al.* (2018b), which employed a block exchangeable correlation structure with three correlation parameters accommodating both cohort and cross-sectional designs under the GEE framework. Hussey and Hughes (2007) developed methods based on a linear mixed effects model with a continuous outcome for a cross-sectional design with a single random cluster effect. Hooper *et al.* (2016) generalized Hussey and Hughes (2007)'s method to closed cohort CRTs with random effects of individuals within cluster, between- and within-period effect, as in Li *et al.* (2018b).

Here, we show the equivalence of power obtained by Hussey and Hughes (2007); Hooper *et al.* (2016); Li *et al.* (2018b) for a continuous outcome. Then, power calculation for continuous outcomes under all scenarios covered in this software can all be incorporated by Li *et al.* (2018b) directly.

A.1. 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 outcome rate on the scale of the link function 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.

It can be shown that, 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 corresponding to the parameter vector η under SWD. To obtain power, we are interested in the variance of $\hat{\beta}$, which is the $(J+1) \times (J+1)$ element of $(\mathbf{Z'V^{-1}Z})^{-1}$, where $\mathbf{V}$ is an $IJ * IJ$ block diagonal matrix that measures the covariance of mean response between different time periods in all clusters. With K individuals at each time period per cluster, Hussey and Hughes (2007) gave:

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

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 cross-sectional SWDs considered by Hussey and Hughes (2007) to cohort SWDs, Hooper *et al.* (2016) also assumed a linear mixed effects model:

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

The notations are the same as Section 2.2. 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 working correlation matrix is the same as the three-level correlation structure R_i given in Section 2.2 for cross-sectional and cohort designs. Model (18) corresponds to individual-level responses based on mixed effects model, which agrees with the population-averaged marginal model in Li *et al.* (2018b):

$$E(Y_{ijk}) = \mu_{ijk} = \mu + X_{ij}\beta + \gamma_j \tag{19}$$

where μ_{ijk} is the mean response of Y_{ijk} under both the marginal and conditional model. Li *et al.* (2018b) and Hooper *et al.* (2016) are developed to accommodate the same three correlation parameters, under marginal and conditional models, 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 $JK_i * (J+1)$ design matrix corresponding to the parameter vector η under SWD 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, the estimators of β under Model (18) and (19) are equivalent and obtain the same power.

As shown in Li *et al.* (2018b), this variance for the estimated intervention effect is:

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

where $U = \sum_{ij} X_{ij}$, $W = \sum_j (\sum_i X_{ij})^2$, $V = \sum_i (\sum_j X_{ij})^2$, $\lambda_3 = 1 + (K-1)(\alpha_0 - \alpha_1) - \alpha_2$, $\lambda_4 = 1 + (K-1)\alpha_0 + (J-1)(K-1)\alpha_1 + (J-1)\alpha_2$.

The equivalence of Model (16) and (19) is proved under the cross-sectional setting in Hussey and Hughes (2007) with $\alpha_0 = \alpha_1 = \alpha_2 = \frac{\tau^2}{\sigma_e^2 + \tau^2}$, where Equation (17) is equal to Equation (20). The treatment effect β is the same under the conditional and marginal model with the identity link function (Ritz and Spiegelman 2004). Thus we can conclude the equivalence of these three models in cases with time effects.

A.2. Without time effects

Here, we consider cases without fixed time effects and derive the three models by Hussey and Hughes (2007); Hooper *et al.* (2016); Li *et al.* (2018b) accordingly.

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

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

The notations are the same as previously.

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 corresponding to the parameter vector η under SWD. To obtain power, we are interested in the variance of $\hat{\beta}$, which is the 2×2

element of $(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}$, 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. With K individuals at each time period per cluster, Liao, Zhou, and Spiegelman (2015) showed that the variance of the estimated 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:

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

The notations are defined in the same way as 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):

$$E(Y_{ijk}) = \mu_{ijk} = \mu + X_{ij}\beta \tag{25}$$

where μ_{ijk} is the mean response of Y_{ijk} under both the marginal and conditional model. Li *et al.* (2018b) and Hooper *et al.* (2016) are developed to accommodate the three three correlation parameters, under marginal and conditional models, 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 $JK_i * 2$ design matrix corresponding to the parameter vector η under SWD 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 the estimated intervention effect under cases without time effects is:

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

where $U = \sum_{ij} X_{ij}$ and $V = \sum_i (\sum_j X_{ij})^2$, $\lambda_3 = 1 + (K-1)(\alpha_0 - \alpha_1) - \alpha_2$, $\lambda_4 = 1 + (K-1)\alpha_0 + (J-1)(K-1)\alpha_1 + (J-1)\alpha_2$.

The equivalence of Model (22) and (25) is proved under the cross-sectional setting in Hussey and Hughes (2007) with $\alpha_0 = \alpha_1 = \alpha_2 = \frac{\tau^2}{\sigma_e^2 + \tau^2}$, where Equation (23) equals to Equation (26). The treatment effect β is the same under the conditional and marginal model with the identity link function (Ritz and Spiegelman 2004). 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/