Fall 2019

ROBUST AND ADAPTIVE DESIGN APPROACHES FOR STEPPED WEDGE CLUSTER RANDOMIZED TRIALS

Jijia Wang
Southern Methodist University, jijiaw@smu.edu

ROBUST AND ADAPTIVE DESIGN APPROACHES FOR STEPPED WEDGE CLUSTER RANDOMIZED TRIALS

Approved by:

Dr. Song Zhang
Associate Professor in Biostatistics

Dr. Chul Ahn
Professor in Biostatistics

Dr. Daniel F. Heitjan
Professor in Biostatistics

Dr. Hon Keung Tony Ng
Professor in Statistics

ROBUST AND ADAPTIVE DESIGN APPROACHES FOR STEPPED WEDGE CLUSTER RANDOMIZED TRIALS

A Dissertation Presented to the Graduate Faculty of the

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Biostatistics

by

Jijia Wang

M.S., Statistical Science, Southern Methodist University

August 6, 2019

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my advisors Dr. Song Zhang and Dr. Chul Ahn. They were great advisors and great persons to work with. I learned a lot from them and very appreciate all their contributions of advices, time, endless encouragement, and continuous support of my study and research. Without them this work would not exist.

Second, I would like to thank Dr. Daniel Heitjan for his tremendous help and support over the past four years. I also thank Dr. Hon Keung Tony Ng for his guidance and advice on my competition project and my dissertaion. They were always supportive and patient. Without them this work would not exist.

Third, I would like to thank my family for being there for me. Although they lived far away from me, they were and will always be close to my heart. I very appreciate their endless support and love. Without them this work would not exist.

Finally, I would like to thank my colleagues and the staffs from Department of Statistical Science at SMU and Department of Population and Data Sciences at UTSW for facilitating a great work and study environment. I did learn from them and enjoy working with them. Moreover, I would like to thank my friends in Texas and everyone there for me, they are indeed true friends and I will miss them all.

Wang, Jijia M.S., Statistical Science, Southern Methodist University

Robust and Adaptive Design Approaches for Stepped Wedge Cluster Randomized Trials

Advisor: Dr. Song Zhang and Dr. Chul Ahn
Doctor of Philosophy degree conferred August 6, 2019
Dissertation completed July 11, 2019

The stepped wedge (SW) cluster randomized design has been increasingly employed by pragmatic trials in health services research. In this study, based on the GEE approach, we present a closed-form sample size that is applicable to both closed-cohort and cross-sectional SW trials with outcomes from the exponential family. Importantly, the proposed method is flexible to accommodate design issues routinely encountered in pragmatic trials, such as different within- and between-subject correlation structures, irregular crossover schedules for the switch to intervention, and missing data due to repeated measurements over prolonged follow-up. The closed-form formula also allows researchers to analytically assess the impact of different design factors on sample size requirement. We also recognize the potential issue of limited numbers of clusters in pragmatic SW trials and present an adjustment approach for underestimated variance of the intervention effect. We conduct extensive simulation to assess the performance of the proposed sample size method. An application example to a real clinical trial is presented.

Bayesian group sequential design is one of the popular adaptive designs and has been applied widely in clinical studies, especially in phase II and III studies. It is flexible and efficient in allowing early termination based on the accumulated data through Bayesian framework. However, so far there is no discussion on its application in SW cluster randomized trials, which become more pragmatic and popular in clinical and health care delivery studies. In this study, we proposed a Bayesian adaptive design for cross-sectional SW cluster randomized

trials. It is more adaptable than traditional designs because it allows early termination of the trial when interim data indicate that the intervention is sufficient efficacious or inefficacious. A decision to terminate or continue the trial will be made on the basis of the predictive probability. This probability is the chance of getting a conclusive result at the end of the study based on the interim data collected so far. We illustrate the early termination criteria and determine the design parameters subject to the constraints of power and type I error. The simulation studies show that the proposed method achieves good operating characteristics. We present an example using our proposed method for illustration purpose.

TABLE OF CONTENTS

LIST OF FIGURES ix

LIST OF TABLES x

CHAPTER

1. OVERVIEW OF STEPPED WEDGE CLUSTER RANDOMIZED TRIALS 1

2. SAMPLE SIZE DETERMINATION FOR SW-CRTS USING GEE 3

2.1. Introduction 3

2.2. Generalized estimating equations 4

2.3. Sample size calculation 6

2.4. Simulation studies 10

2.5. A real application example 16

2.6. Extension to outcomes from the exponential family 16

2.7. Discussion 26

3. A BAYESIAN ADAPTIVE DESIGN FOR SW-CRTS 27

3.1. Introduction 27

3.2. A Bayesian predictive probability approach 29

3.3. Simulation studies 33

3.4. A real application example 43

3.5. Discussion 45

APPENDIX

A. Derivation of equations 48

A.1. Derivation of Equation (2.3) 48

A.2. Derivation of Equations (2.4) and (2.5) 51

A.3. Sample size for cross-sectional SW-CRTs 53

A.4. Derivation of Equation (2.11) 54

BIBLIOGRAPHY 55

LIST OF FIGURES

Figure Page

1.1 An illustration of a SW-CRT with $T = 5$ time points and $S = 4$ sequences (Each cell represents a data collection point. Shaded and blank cells represent intervention and control, respectively.) 1

3.1 Empirical power (left) and expected number of subjects (right) using predictive probability method with different combination of design parameters. 41

3.2 Empirical power, probability of early stopping, and expected number of subjects using predictive probability method with different intervention effects 42

3.3 Empirical power, probability of early stopping, and expected number of subjects with varying π_U 45

LIST OF TABLES

Table Page

2.1 Required number of clusters (empirical power, empirical type I error) for closed-cohort SW-CRTs with between-subject correlation $\rho_2 = 0.03$ 13

2.2 Required number of clusters (empirical power, empirical type I error) for closed-cohort SW-CRTs with between-subject correlation $\rho_2 = 0.05$ 14

2.3 Required number of clusters (empirical power, empirical type I error) for cross-sectional SW-CRTs 15

2.4 Notations for different types of outcome 19

2.5 Required number of clusters (empirical power, empirical type I error) for binary outcomes with between-subject correlation $\rho_2 = 0.03$ 22

2.6 Required number of clusters (empirical power, empirical type I error) for binary outcomes with between-subject correlation $\rho_2 = 0.05$ 23

2.7 Required number of clusters (empirical power, empirical type I error) for cross-sectional SW-CRTs with binary outcomes 23

2.8 Required number of clusters (empirical power, empirical type I error) for count outcomes with between-subject correlation $\rho_2 = 0.03$ 24

2.9 Required number of clusters (empirical power, empirical type I error) for count outcomes with between-subject correlation $\rho_2 = 0.05$ 25

2.10 Required number of clusters (empirical power, empirical type I error) for cross-sectional SW-CRTs with count outcomes 25

3.1 Empirical power/type I error/expected number of subjects using predictive probability method with $n = 15$ 37

3.2 Empirical power/type I error/expected number of subjects using predictive probability method with $n = 20$ 38

3.3 Empirical power/type I error/expected number of subjects using predictive probability method with $n = 25$ 39

3.4 Empirical power/type I error/expected number of subjects for a SW-CRT with $n = 10$ and $\zeta = 0.3$ 44

This is dedicated to my family.

Chapter 1

OVERVIEW OF STEPPED WEDGE CLUSTER RANDOMIZED TRIALS

Stepped wedge cluster randomized trials (SW-CRTs) have been increasingly employed in health services research (Mhurchu et al., 2013; Bailet et al., 2009; Bacchieri et al., 2010; van Holland et al., 2012). They have a unique feature that allows stepwise introduction of interventions to a population, making it suitable to examine the effectiveness of interventions in real-life practice (Hussey and Hughes, 2007; Brown and Lilford, 2006). Figure 1.1 provides an illustration of the stepped wedge cluster randomized design. The horizontal axis indicates time and the vertical axis indicates the sequences. Each sequence includes one or more clusters of subjects, for example, formed by physicians and clinics. All clusters receive control treatment at the start and cross over to receive the intervention from a randomly assigned time onwards. Outcomes are measured at every time point (or step).

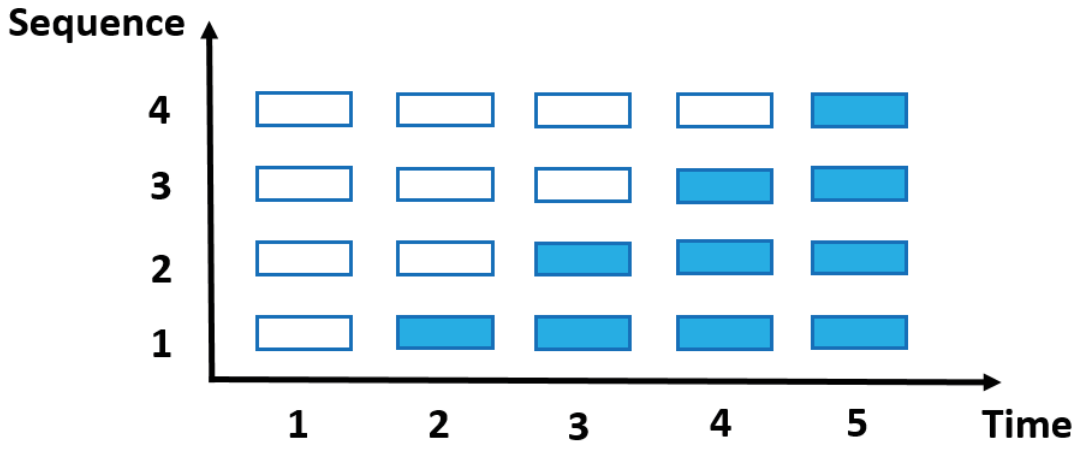

Figure 1.1. An illustration of a SW-CRT with $T = 5$ time points and $S = 4$ sequences (Each cell represents a data collection point. Shaded and blank cells represent intervention and control, respectively.)

SW-CRTs have several advantages. First, by switching all clusters to intervention by the end of the study, it mitigates ethical dilemma of withholding an effective intervention when the intervention is believed to be superior to the control (Edwards, 2013). Second, it is more efficient because the evidence of effectiveness is provided by comparisons from two directions: vertical comparisons of clusters between intervention and control at each time point, and horizontal comparisons of observations before and after switching to intervention within each cluster (Hargreaves et al., 2015). Third, outcomes are measured longitudinally, offering insight into the temporal trend of intervention effect. On the other hand, SW-CRTs present various challenges to researchers, which include longer study duration, higher risk of attrition, and difficulty in blinding. For example, the correlation structure is much more complicated than that in parallel cluster randomized designs. There are intracluster correlation among subjects from the same cluster as well as longitudinal correlation among measurements obtained over time from the same subjects. Moreover, the proportion of missing data is usually higher due to repeated measurements over a longer duration, which cannot be ignored in the design.

There are three major types of SW-CRTs (Copas et al., 2015; Beard et al., 2015; Martin et al., 2016). The first is called closed-cohort, where all subjects are followed through the study hence longitudinal measurements are collected from the same individuals. The second is called cross-sectional, where subjects selected for measurements at a specific time point will not be selected again in the future. The third is called open-cohort, where some subjects are followed through the study, while others can leave or join in mid-study.

Chapter 2

SAMPLE SIZE DETERMINATION FOR SW-CRTS USING GEE

2.1. Introduction

Many researchers have investigated sample size calculation methods for cross-sectional SW-CRTs (Hussey and Hughes, 2007; Hemming et al., 2015; Woertman et al., 2013; Moulton et al., 2007), where the correlation structure is simpler. There have been relatively fewer publications on sample size methods for closed-cohort SW-CRTs, some of which were purely simulation-based (Baio et al., 2015), some based on the linear mixed-effect model approach (Hooper et al., 2016; Girling and Hemming, 2016), and some based on the Bayesian method (Cunanan et al., 2016). Li et al. (2018) proposed a sample size determination procedure for closed-cohort SW-CRTs using the generalized estimating equations (GEE) approach (Liang and Zeger, 1986), which considered a block exchangeable correlation structure defined by three types of correlations: within-period, inter-period, and within-individual correlations. It is noteworthy that all existing methods only consider ideal scenarios such as no missing data, which are unlikely to hold in pragmatic trials conducted in the settings of real world clinical practice.

In this study we investigate sample size calculation for SW-CRTs within the context of pragmatic trials. Based on the GEE approach, we derived closed-form sample size formulas for SW-CRTs with outcomes from exponential family, which can be applied to either the closed-cohort or cross-sectional SW-CRTs by modifying the correlation structures. The proposed sample size method is flexible to address pragmatic issues such as missing data due to various mechanisms, complicated correlation structures, or arbitrary crossover schemes. Specifically, (1) we introduce the GEE approach for SW-CRTs with continuous outcomes; (2) we derive closed-form sample size formulas that accommodates various pragmatic design

issues, such as unbalanced randomization of clusters, correlation structures, and missing data; (3) we conduct simulation studies to assess the performance of the proposed method under various designing configurations; (4) we illustrate the proposed method with a real example; (5) we extend the proposed method to SW-CRTs with outcomes from exponential family; and (6) we conclude with a discussion of proposed method.

2.2. Generalized estimating equations

We consider a closed-cohort SW-CRT with continuous outcomes. Suppose there are n clusters and T steps. Each cluster is randomly assigned to one of the S sequences ($S = T-1$). Let Y_{ijt} be the continuous response obtained at time t ($t = 1, \ldots, T$) from subject j ($j = 1, \ldots, J$) of the ith cluster ($i = 1, \ldots, n$), where J is the number of subjects in each cluster. Thus $N = nJ$ is the total number of subjects. A linear model is assumed for $\boldsymbol{Y_{ij}} = (Y_{ij1}, \ldots, Y_{ijT})'$:

$$E\left(\boldsymbol{Y}_{ij}\right) = \boldsymbol{\lambda} + \boldsymbol{u}_i\zeta = \boldsymbol{X}_i\boldsymbol{\beta}, \tag{2.1}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ contains time-specific intercepts that implies a flexible model for the temporal trend under control; $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iT})'$ represents the treatments received by cluster i over time, with $u_{it} = 0/1$ denoting control/intervention; and ζ is a scalar parameter representing the intervention effect, which is assumed to be constant over time. We define $\boldsymbol{X}_i = (\boldsymbol{I}_T, \boldsymbol{u}_i)$ to be the subject-specific design matrix, where $\boldsymbol{I}_T$ denotes the $T \times T$ identity matrix. Hence $\boldsymbol{\beta} = \left(\boldsymbol{\lambda}', \zeta\right)'$ is the vector of regression coefficients. Depending on randomization, we have $P(\boldsymbol{u}_i = \boldsymbol{v}_s) = p_s$ ($s = 1, \ldots, S$), where

$$\begin{aligned}
\boldsymbol{v_1} &= (0, 1, 1, \ldots, 1, 1)', \\
\boldsymbol{v_2} &= (0, 0, 1, \ldots, 1, 1)', \\
&\vdots \\
\boldsymbol{v_S} &= (0, 0, 0, \ldots, 0, 1)',
\end{aligned}$$

and $\sum_{s=1}^{S} p_s = 1$. Under the above specification, the proportion of subjects receiving intervention at step t is $\bar{u}_t = \sum_{s=1}^{S} p_s v_{st}$. Here v_{st} is the tth element of $\boldsymbol{v}_s$. Accordingly, we define the vector $\overline{\boldsymbol{u}} = (\overline{u}_1, \ldots, \overline{u}_T)$.

We assume $\text{Var}\,(Y_{ijt}) = \sigma^2$. For each cluster, we use $\text{Corr}\,(\boldsymbol{Y_{ij}}) = \boldsymbol{\Omega}$ to denote longitudinal (or within-subject) correlation and $\text{Corr}\,(\boldsymbol{Y_{ij}}, \boldsymbol{Y_{ij'}}) = \boldsymbol{\Phi}$ for $j \neq j'$ to denote between-subject correlation within the same cluster. We assume the subjects to be independent across clusters. Hence, $\boldsymbol{\Omega}$ and $\boldsymbol{\Phi}$ fully specify the correlation structure of $\boldsymbol{Y_i} = \left(\boldsymbol{Y'_{i1}}, \ldots, \boldsymbol{Y'_{iJ}}\right)'$. Specifically,

$$\text{Corr}\,(\boldsymbol{Y_i}) = \boldsymbol{I}_J \otimes (\boldsymbol{\Omega} - \boldsymbol{\Phi}) + (\mathbf{1}_J \mathbf{1}'_J) \otimes \boldsymbol{\Phi}. \tag{2.2}$$

Here $\otimes$ indicates the operation of Kronecker product, $\boldsymbol{I}_J$ denotes a $J \times J$ identity matrix, and $\mathbf{1}_J$ is a vector of length J with all elements being 1.

The primary interest is to test the null hypothesis $H_0 : \zeta = 0$. We estimate parameters based on the GEE approach using an independent working correlation structure, which serves as an instrument to facilitate the estimating process. The estimates remain consistent but the computation is much simpler and more stable in terms of convergence (Crowder, 1995). It has also been demonstrated that in many situations the resulting estimates are highly efficient (Liang and Zeger, 1986; McDonald, 1993) and successfully employed in sample size calculation (Ahn et al., 2014; Zhang et al., 2018; Lou et al., 2017; Zhu et al., 2017). The GEE estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} J \boldsymbol{X}_i' \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \sum_{j=1}^{J} \boldsymbol{X}_i' \boldsymbol{Y}_{ij} \right).$$

Under mild regularity conditions (Liang and Zeger, 1986), as $n \to \infty$, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N\,(\mathbf{0},\, \boldsymbol{\Sigma}_n)$ in distribution. The covariance matrix $\boldsymbol{\Sigma}_n$ can be consistently estimated by $\boldsymbol{A}_n^{-1} \boldsymbol{E}_n \boldsymbol{A}_n^{-1}$, where

$$\boldsymbol{A}_n = n^{-1} \sum_{i=1}^{n} J \boldsymbol{X}_i' \boldsymbol{X}_i$$

and

$$\boldsymbol{E}_n = n^{-1} \sum_{i=1}^{n} \left(\sum_{j=1}^{J} \boldsymbol{X}_i' \hat{\boldsymbol{e}}_{ij} \right)^{\otimes 2}.$$

Here $\hat{\boldsymbol{e}}_{ij} = \boldsymbol{Y}_{ij} - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}$ is the residual and $\boldsymbol{c}^{\otimes 2} = \boldsymbol{c}\boldsymbol{c}'$ for a vector $\boldsymbol{c}$. The null hypothesis $H_0 : \zeta = 0$ is rejected if $\sqrt{n}|\hat{\zeta}|/\hat{\sigma}_\zeta > z_{1-\alpha/2}$, where $\hat{\zeta}$ is the $(T+1)$th element of $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_\zeta^2$ is the $(T+1, T+1)$-component of $\boldsymbol{\Sigma}_n$, and $z_{1-\alpha/2}$ is the $100\,(1-\alpha/2)$th percentile of the standard normal distribution.

2.3. Sample size calculation

Let $\boldsymbol{A}$ and $\boldsymbol{E}$ be the limits of $\boldsymbol{A}_n$ and $\boldsymbol{E}_n$ as $n \to \infty$, then $\boldsymbol{\Sigma}_n$ converges to $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{E}\boldsymbol{A}^{-1}$. Specifically, we have

$$\boldsymbol{A} = J \sum_{s=1}^{S} p_s \boldsymbol{W}_s' \boldsymbol{W}_s$$

and

$$\boldsymbol{E} = \sigma^2 J \sum_{s=1}^{S} p_s \boldsymbol{W}_s' \left[\boldsymbol{\Omega} + (J-1)\, \boldsymbol{\Phi} \right] \boldsymbol{W}_s,$$

where $\boldsymbol{W}_s = (\boldsymbol{I}_T, \boldsymbol{v}_s)$. It can be shown that the $(T+1, T+1)$-component of $\boldsymbol{\Sigma}$ is

$$\sigma_\zeta^2 = \frac{\sigma^2 \sum_{s=1}^{S} p_s \left(\boldsymbol{v}_s - \bar{\boldsymbol{u}} \right)' \left[\boldsymbol{\Omega} + (J-1)\, \boldsymbol{\Phi} \right] \left(\boldsymbol{v}_s - \bar{\boldsymbol{u}} \right)}{J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t \right) \right]^2}.$$

Under true intervention effect ζ_0, to reject the null hypothesis $H_0 : \zeta = 0$ with power $1-\gamma$ at two-sided significance level α, the required number of clusters can be calculated by

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\gamma} \right)^2 \sigma^2 \sum_{s=1}^{S} p_s \left(\boldsymbol{v}_s - \bar{\boldsymbol{u}} \right)' \left[\boldsymbol{\Omega} + (J-1)\, \boldsymbol{\Phi} \right] \left(\boldsymbol{v}_s - \bar{\boldsymbol{u}} \right)}{\zeta_0^2 J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t \right) \right]^2}, \tag{2.3}$$

Details of derivation are presented in Appendix.

We have a few observations about sample size (2.3). First, by introducing notations of $\boldsymbol{\Omega}$ and $\boldsymbol{\Phi}$, as well as p_s and $\boldsymbol{v}_s$, the proposed sample size provides great flexibility for researchers to account for complicated correlation structures and arbitrary crossover schedules encountered in real world SW-CRTs. Furthermore, it is noteworthy that time-specific intercepts ($\boldsymbol{\lambda}$) are not involved in Equation (2.3). That is, the temporal trend under control

approximately has negligible impact on sample size requirement for SW-CRTs with continuous outcomes. Because all clusters receive control at Step 1 and experimental intervention at Step T, it is obvious that the first elements of $\boldsymbol{v_s}$ $(s = 1, \cdots, S)$ and $\bar{\boldsymbol{u}}$ all take value 0, and the last elements all take value 1. In turn, the first and last elements of $(\boldsymbol{v_s} - \bar{\boldsymbol{u}})$ equal 0 for $s = 1, \cdots, S$. From this observation we have the following fact:

Fact: *Define* $\boldsymbol{Y}_{ij}^* = (Y_{ij2}, \cdots, Y_{ij(T-1)})'$, $\boldsymbol{\Omega}^* = Corr(\boldsymbol{Y}_{ij}^*)$, *and* $\boldsymbol{\Phi}^* = Corr(\boldsymbol{Y}_{ij}^*, \boldsymbol{Y}_{ij'}^*)$, *where* $j \neq j'$. *The correlation matrices* $\boldsymbol{\Omega}$ *and* $\boldsymbol{\Phi}$ *affect sample size (2.3) only through submatrices* $\boldsymbol{\Omega}^*$ *and* $\boldsymbol{\Phi}^*$.

This fact is particularly meaningful for special scenarios where $T = 3$ or $T = 4$. Under $T = 3$, between-subject correlation $\boldsymbol{\Phi}$ affects sample size only through $\text{Corr}(Y_{ij2}, Y_{ij'2})$, while within-subject correlation $\boldsymbol{\Omega}$ has no effect. Under $T = 4$, within-subject correlation $\boldsymbol{\Omega}$ affects sample size only through $\text{Corr}(Y_{ij2}, Y_{ij3})$. Hence whether $\boldsymbol{\Omega}$ has a CS or AR(1) correlation structure will lead to the same sample size as long as $\text{Corr}(Y_{ij2}, Y_{ij3})$ remains the same. The expressions of sample size (2.3) under $T = 3$ and $T = 4$ are presented in Appendix.

Traditionally researchers have assumed a balanced randomization scheme. That is, $p_s = \frac{1}{S}$ for $s = 1, \ldots, S$. The sample size (2.3) under balanced randomization becomes

$$n = \frac{36\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 S\sigma^2 \sum_{s=1}^{S} \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[\boldsymbol{\Omega} + (J-1)\boldsymbol{\Phi}\right] \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)}{\zeta_0^2 J \left(S^2 - 1\right)^2}. \tag{2.4}$$

The correlation matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Phi}$ can be specified directly based on preliminary knowledge. Alternatively, their specification can be motivated by linear mixed-effect models. For example, when the model includes cluster and patient random effects, the implied within-subject (longitudinal) correlation matrix $\boldsymbol{\Omega}$ is compound symmetric (CS) with off-diagonal elements being ρ_1 and the between-subject correlation matrix can be written as $\boldsymbol{\Phi} = \mathbf{1}\mathbf{1}'\rho_2$, then the sample size becomes

$$n = \frac{3\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 S\sigma^2 \left[(S-2)\rho_1 + S(J-1)\rho_2 + 2\right]}{\zeta_0^2 J \left(S^2 - 1\right)}. \tag{2.5}$$

The derivation of (2.4) and (2.5) is presented in Appendix. It is noteworthy that sample size (2.5) is a linear function of ρ_1 and ρ_2. Equation (2.5) also shows that the coefficient of ρ_1

is $S-2$, which is much smaller than the coefficient of ρ_2, $S(J-1)$, under moderate cluster sizes. Hence the between-subject correlation ρ_2 has greater impact on sample size than within-subject correlation ρ_1. Finally, as cluster size $J \to \infty$, sample size (2.5) approaches

$$\frac{3\left(z_{1-\alpha/2}+z_{1-\gamma}\right)^2 S^2\sigma^2\rho_2}{\zeta_0^2\left(S^2-1\right)}.$$

This lower limit for the number of clusters as $J \to \infty$ suggests that, in practice, increasing enrollment within existing clusters (increasing J) cannot always compensate for the lack of participation of unique clusters.

The above sample size derivation is presented in the context of closed-cohort SW-CRTs. It can easily accommodate cross-sectional SW-CRTs through the specification of $\boldsymbol{\Omega}$ and $\boldsymbol{\Phi}$. Under the cross-sectional stepped wedge design, the observations obtained at each step are contributed by different panels of subjects from the same clusters. The resulting correlation structure is characterized by $\boldsymbol{\Omega} = \mathbf{1}\mathbf{1}'\rho + (1-\rho)\boldsymbol{I}$ and $\boldsymbol{\Phi} = \mathbf{1}\mathbf{1}'\rho$, as considered by Hussey and Hughes (2007). Here ρ is the intracluster correlation coefficient (ICC). The required number of clusters is

$$n = \frac{\left(z_{1-\alpha/2}+z_{1-\gamma}\right)^2\sigma^2\sum_{s=1}^{S}p_s\left\{J\rho\left[\sum_{t=1}^{T}(v_{st}-\overline{u}_t)\right]^2+(1-\rho)\sum_{t=1}^{T}(v_{st}-\overline{u}_t)^2\right\}}{\zeta_0^2 J\left[\sum_{t=1}^{T}\overline{u}_t\left(1-\overline{u}_t\right)\right]^2}. \quad (2.6)$$

Under the special case of $p_s = \frac{1}{S}$ for $s = 1, \ldots, S$, it can be simplified as

$$n = \frac{3\left(z_{1-\alpha/2}+z_{1-\gamma}\right)^2 S\sigma^2\left[(JS-2)\rho+2\right]}{\zeta_0^2 J\left(S^2-1\right)}. \quad (2.7)$$

Details of derivation are presented in Appendix.

The problem of missing data is frequently encountered in clinical trials. For closed-cohort SW-CRTs, this problem is particularly pronounced due to the prolonged follow-up. Here we present an extension to sample size (2.3) that accommodates missing data and still has a closed form. First we introduce missing data indicator Δ_{ijt}, which takes value 0 or 1 for missed or observed measurement from the jth subject within the ith cluster at time t. Under the assumption of missing completely at random (MCAR), $\boldsymbol{A}_n$ and $\boldsymbol{\Sigma}_n$ can be rewritten as

$$\boldsymbol{A}_n^* = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\boldsymbol{X}_{\boldsymbol{i}}^{'}\mathrm{diag}\left(\boldsymbol{\Delta_{ij}}\right)\boldsymbol{X_i}$$

and

$$\boldsymbol{E}_n^* = n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\boldsymbol{X}_{\boldsymbol{i}}^{'}(\boldsymbol{\Delta_{ij}}\circ\hat{\boldsymbol{e}}_{\boldsymbol{ij}})\right)^{\otimes 2},$$

where $\boldsymbol{\Delta_{ij}} = \left(\Delta_{ij1},\ldots,\Delta_{ijT}\right)^{'}$, $\mathrm{diag}\left(\boldsymbol{c}\right)$ is a diagonal matrix with $\boldsymbol{c}$ being the diagonal elements, and $\circ$ indicates the operation of Hadamard product. Specifically, for two matrices M_1 and M_2 of the same dimension, the (i,j)th-component of $M_1 \circ M_2$ is the product of (i,j)th-component of M_1 and (i,j)th-component of M_2. We assume that $P\left(\Delta_{ijt}=1\right)=\delta_t$. We further define the joint probability of observing measurements at both time t and t' as $P\left(\Delta_{ijt}\Delta_{ijt'}=1\right)=\delta_{tt'}$. Usually the occurrence of missing data increases over time, hence $\delta_1 \geq \cdots \geq \delta_T$. Given the same set of marginal missing probabilities, there can be different missing data patterns. For example, missed visits cause missing values to occur independently over time, which is called the independent missing (IM) pattern with $\delta_{tt'} = \delta_t\delta_{t'}$ for $t \neq t'$, and $\delta_{tt} = \delta_t$. On the other hand, patient dropout leads to a pattern such that once a patient misses a measurement, all his/her subsequent measurements are missing. This is called the monotone missing (MM) pattern with $\delta_{tt'} = \delta_{t'}$ for $t \leq t'$, and $\delta_{tt} = \delta_t$.

As $n \to \infty$, $\boldsymbol{A}_n^* \to \boldsymbol{A}^*$ and $\boldsymbol{E}_n^* \to \boldsymbol{E}^*$, where

$$\boldsymbol{A}^* = J\sum_{s=1}^{S} p_s \boldsymbol{W}_{\boldsymbol{s}}^{'}\mathrm{diag}\left(\boldsymbol{\delta}\right)\boldsymbol{W_s}$$

and

$$\boldsymbol{E}^* = \sigma^2 J\sum_{s=1}^{S} p_s \boldsymbol{W}_{\boldsymbol{s}}^{'}\left[\widetilde{\boldsymbol{\delta}}\circ\boldsymbol{\Omega}+(J-1)\,\mathrm{diag}\left(\boldsymbol{\delta}\right)\boldsymbol{\Phi}\mathrm{diag}\left(\boldsymbol{\delta}\right)\right]\boldsymbol{W_s}.$$

Here $\boldsymbol{\delta} = \left(\delta_1,\ldots,\delta_T\right)^{'}$, and $\widetilde{\boldsymbol{\delta}}$ is a matrix with the (t,t)th element being δ_t and the $(t,t^{'})$th $(t \neq t')$ element being $\delta_{tt'}$. Based on the derivation similar to that for sample size (2.3), it can be shown that the general sample size which accommodates missing data still has a closed form:

$$n = \frac{\left(z_{1-\alpha/2}+z_{1-\gamma}\right)^2\sigma^2\sum_{s=1}^{S}p_s\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)^{'}\left[\widetilde{\boldsymbol{\delta}}\circ\boldsymbol{\Omega}+(J-1)\,\mathrm{diag}\left(\boldsymbol{\delta}\right)\boldsymbol{\Phi}\mathrm{diag}\left(\boldsymbol{\delta}\right)\right]\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)}{\zeta_0^2 J\left[\sum_{t=1}^{T}\delta_t\overline{u}_t\left(1-\overline{u}_t\right)\right]^2}. \tag{2.8}$$

Correspondingly, given n, the power can be calculated by

$$P\left(Z < \frac{\sqrt{nJ}|\zeta_0|\sum_{t=1}^{T}\delta_t\overline{u}_t\left(1-\overline{u}_t\right)}{\sigma\sqrt{\sum_{s=1}^{S}p_s\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)'\left[\widetilde{\boldsymbol{\delta}}\circ\boldsymbol{\Omega}+(J-1)\,\mathrm{diag}\left(\boldsymbol{\delta}\right)\boldsymbol{\Phi}\mathrm{diag}\left(\boldsymbol{\delta}\right)\right]\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)}} - z_{1-\alpha/2}\right), \tag{2.9}$$

where Z is a random variable following the standard normal distribution. For a closed-cohort SW-CRT, the total number of subjects is jointly determined by the number of clusters (n) and the number of subjects per cluster (J). Given n, we can calculate the cluster size by

$$J = \frac{\sum_{s=1}^{S}p_s\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)'\widetilde{\boldsymbol{\delta}}\circ\boldsymbol{\Omega}\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)-C}{n\zeta_0^2\left[\sum_{t=1}^{T}\overline{u}_t\left(1-\overline{u}_t\right)\right]^2\left[\left(z_{1-\alpha/2}+z_{1-\gamma}\right)\sigma\right]^{-2}-C}, \tag{2.10}$$

where $C=\sum_{s=1}^{S}p_s\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)'\left[\mathrm{diag}\left(\boldsymbol{\delta}\right)\boldsymbol{\Phi}\mathrm{diag}\left(\boldsymbol{\delta}\right)\right]\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)$.

2.4. Simulation studies

To investigate the performance of the proposed method, we conducted simulation studies for SW-CRTs with continuous outcomes. Two cluster sizes were explored: $J = 20$ and 40. We assumed the subjects to transition to intervention over $T = 5$ time periods. We assumed $\sigma^2 = 1$ and intervention effect $\zeta_0 = 0.2$. The time-specific intercepts were specified as $\lambda_t = 0.1\,(t-1)$ for $t = 1,\ldots,T$. For closed-cohort SW-CRTs, we explored two structures for the within-subject correlation matrix $\boldsymbol{\Omega}$: CS and AR(1). Specifically, under CS the off-diagonal elements are $\rho_{tt'} = \rho_1$ $(t \neq t')$ and under AR(1) we specify $\rho_{tt'} = \rho_1^{|t-t'|/S}$ $(t \neq t')$. The between-subject correlation matrix was specified as $\boldsymbol{\Phi} = \mathbf{11'}\rho_2$. We explored $(\rho_1, \rho_2) = \{(0.15, 0.03), (0.15, 0.05), (0.30, 0.03), (0.30, 0.05)\}$ to assess the impact of correlations on sample size. Moreover, we investigated different missing patterns, which include independent missing (IM) and monotone missing (MM). For marginal observed probabilities

$\boldsymbol{\delta} = (\delta_1, \ldots, \delta_T)'$, we consider four sets of observational probabilities:

$$\begin{aligned}
\boldsymbol{\delta}_1 &= (1.000, 1.000, 1.000, 1.000, 1.000)\,, \\
\boldsymbol{\delta}_2 &= (1.000, 0.790, 0.760, 0.730, 0.700)\,, \\
\boldsymbol{\delta}_3 &= (1.000, 0.925, 0.850, 0.775, 0.700)\,, \\
\boldsymbol{\delta}_4 &= (1.000, 1.000, 1.000, 0.800, 0.700)\,.
\end{aligned}$$

Here $\boldsymbol{\delta}_1$ represents the scenario of no missing data, while $\boldsymbol{\delta}_2$, $\boldsymbol{\delta}_3$, $\boldsymbol{\delta}_4$ represent scenarios where an increasing number of subjects miss visits over time with a dropout rate of 0.3 at the end of study. However, their change patterns are different. Under $\boldsymbol{\delta}_2$ the observational probability drops quickly initially but plateaus afterwards. Under $\boldsymbol{\delta}_3$ the observational probability drops following a linear trend. And under $\boldsymbol{\delta}_4$ the observational probability remains high initially but drop quickly afterwards. For cross-sectional SW-CRTs, we specified $\boldsymbol{\Omega} = \mathbf{1}\mathbf{1}'\rho + (1-\rho)\boldsymbol{I}$ and $\boldsymbol{\Phi} = \mathbf{1}\mathbf{1}'\rho$, as discussed previously. We explored two values for the ICC: $\rho = 0.03$ and 0.05. To test the null hypothesis $H_0 : \zeta = 0$, we set the two-sided type I error rate $\alpha = 0.05$ and power $1 - \gamma = 0.8$, respectively, and assumed balanced randomization $p_1 = \cdots = p_S = 1/4$.

For each combination of design parameters, the simulation scheme is presented as follows:

1. Calculate the required number of clusters (n).

2. Generate a set of observations ($\boldsymbol{Y_i}$) based on simulation parameters under the null hypothesis $H_0 : \zeta = 0$ or alternative hypothesis $H_1 : \zeta = \zeta_0$. $\boldsymbol{Y_i}$ is generated from a multivariate normal random distribution with mean and covariance matrix implied by the design parameters.

3. Generate missing indicators under different missing patterns and probabilities $\boldsymbol{\delta}$.

4. Calculate $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_n$. The test statistic is $Z = \sqrt{n}|\hat{\zeta}|/\hat{\sigma}_\zeta$. If $Z > z_{1-\alpha/2}$, then set rejection indicator $D = 1$; otherwise $D = 0$.

5. Repeat Steps 2-4 for $L = 10000$ times to obtain D_l for $l = 1, \ldots, L$. The empirical type I error and empirical power are estimated by $\sum_{l=1}^{L} D_l/L$ under the null and alternative hypothesis, respectively.

Tables 2.1 and 2.2 present the required number of clusters under different parameter settings for closed-cohort SW-CRTs with between-subject correlation $\rho_2 = 0.03$ and 0.05, respectively. We have several observations. First, the comparison between Table 1 and Table 2 shows a drastic change in sample size when ρ_2 changes from 0.03 to 0.05. On the other hand, within each table, sample sizes under $\rho_1 = 0.15$ and $\rho_1 = 0.30$ don't change as much. This observation corroborates the theoretical conclusion that between-subject correlation (ρ_2) has a greater impact on sample size than within-subject correlation (ρ_1). Second, under the same between-subject correlation ρ_2 (see Table 2.1 or 2.2), the required number of clusters is influenced by longitudinal correlation structures. For example the first cell in Table 1, under complete data ($\boldsymbol{\delta_1}$), the required number of clusters is 28 for CS structure, but it is 31 for AR(1) structure. Third, the trials with larger cluster size (J) require less number of clusters. Fourth, the trials with subjects dropping quickly initially ($\boldsymbol{\delta_2}$) need larger sample sizes to compensate the information loss compared to the other missing scenarios ($\boldsymbol{\delta_3}$ and $\boldsymbol{\delta_4}$), despite the same dropout rate at the end of study. We also observe that the existence of various types of correlations mitigates the impact of missing data in SW-CRTs. Specifically, with a dropout rate of 0.3 at the end of study, the increase in sample size that is needed to compensate for missing data is at most 19% over the missing patterns considered. Table 2.3 presents the required number of clusters for cross-sectional SW-CRTs. Here we assume no missing data because each patient only contributes one measurement. Increased ICC (ρ) leads to increase in required number of clusters, which has been shown in Equation (2.7). Moreover, larger cluster size leads to less required number of clusters.

Table 2.1. Required number of clusters (empirical power, empirical type I error) for closed-cohort SW-CRTs with between-subject correlation $\rho_2 = 0.03$

		Missing		GEE		Adjusted GEE	
	J	Pattern	$\boldsymbol{\delta}$	$\rho_1 = 0.15$	$\rho_1 = 0.30$	$\rho_1 = 0.15$	$\rho_1 = 0.30$
CS	40	IM	$\boldsymbol{\delta}_1$	28 (0.8136, 0.0764)	29 (0.8146, 0.0789)	30 (0.8024, 0.0532)	31 (0.7930, 0.0539)
			$\boldsymbol{\delta}_2$	30 (0.8039, 0.0763)	32 (0.8176, 0.0746)	32 (0.7882, 0.0509)	34 (0.7992, 0.0492)
			$\boldsymbol{\delta}_3$	29 (0.8092, 0.0812)	30 (0.8099, 0.0716)	31 (0.7944, 0.0499)	32 (0.7951, 0.0508)
			$\boldsymbol{\delta}_4$	28 (0.8043, 0.0778)	30 (0.8159, 0.0754)	30 (0.7922, 0.0551)	32 (0.8006, 0.0497)
		MM	$\boldsymbol{\delta}_2$	31 (0.8201, 0.0738)	32 (0.8123, 0.0732)	33 (0.8079, 0.0516)	34 (0.7932, 0.0528)
			$\boldsymbol{\delta}_3$	29 (0.8077, 0.0770)	31 (0.8136, 0.0708)	31 (0.7926, 0.0576)	33 (0.8033, 0.0542)
			$\boldsymbol{\delta}_4$	28 (0.8120, 0.0710)	30 (0.8275, 0.0788)	30 (0.7918, 0.0556)	32 (0.7975, 0.0530)
	20	IM	$\boldsymbol{\delta}_1$	36 (0.8107, 0.0642)	39 (0.8110, 0.0664)	38 (0.7929, 0.0512)	41 (0.7977, 0.0512)
			$\boldsymbol{\delta}_2$	41 (0.8043, 0.0668)	44 (0.8064, 0.0634)	43 (0.7887, 0.0516)	46 (0.8062, 0.0492)
			$\boldsymbol{\delta}_3$	39 (0.8096, 0.0707)	42 (0.8178, 0.0687)	41 (0.7930, 0.0507)	44 (0.7956, 0.0485)
			$\boldsymbol{\delta}_4$	37 (0.8017, 0.0730)	40 (0.8078, 0.0605)	39 (0.7971, 0.0503)	42 (0.7996, 0.0517)
		MM	$\boldsymbol{\delta}_2$	42 (0.8080, 0.0637)	45 (0.8141, 0.0666)	44 (0.7922, 0.0485)	47 (0.7963, 0.0548)
			$\boldsymbol{\delta}_3$	40 (0.8160, 0.0688)	42 (0.8111, 0.0664)	42 (0.8079, 0.0533)	44 (0.8038, 0.0544)
			$\boldsymbol{\delta}_4$	37 (0.8117, 0.0688)	40 (0.8117, 0.0690)	39 (0.7855, 0.0546)	42 (0.7966, 0.0511)
AR(1)	40	IM	$\boldsymbol{\delta}_1$	31 (0.8097, 0.0737)	32 (0.8114, 0.0716)	33 (0.8074, 0.0541)	34 (0.8009, 0.0541)
			$\boldsymbol{\delta}_2$	34 (0.8189, 0.0716)	35 (0.8143, 0.0733)	36 (0.8008, 0.0521)	37 (0.8042, 0.0546)
			$\boldsymbol{\delta}_3$	33 (0.8247, 0.0702)	34 (0.8190, 0.0727)	35 (0.8058, 0.0564)	36 (0.8001, 0.0546)
			$\boldsymbol{\delta}_4$	32 (0.8147, 0.0749)	33 (0.8135, 0.0790)	34 (0.8054, 0.0549)	35 (0.8059, 0.0560)
		MM	$\boldsymbol{\delta}_2$	35 (0.8174, 0.0703)	36 (0.8189, 0.0761)	37 (0.7941, 0.0516)	38 (0.8005, 0.0524)
			$\boldsymbol{\delta}_3$	33 (0.8113, 0.0725)	34 (0.8152, 0.0758)	35 (0.7950, 0.0527)	36 (0.7972, 0.0589)
			$\boldsymbol{\delta}_4$	32 (0.8176, 0.0760)	33 (0.8142, 0.0743)	34 (0.8081, 0.0519)	35 (0.8004, 0.0548)
	20	IM	$\boldsymbol{\delta}_1$	43 (0.8099, 0.0660)	45 (0.8100, 0.0677)	45 (0.8006, 0.0531)	47 (0.7995, 0.0584)
			$\boldsymbol{\delta}_2$	48 (0.8065, 0.0659)	50 (0.8099, 0.0620)	50 (0.8004, 0.0523)	52 (0.8035, 0.0532)
			$\boldsymbol{\delta}_3$	46 (0.8082, 0.0676)	48 (0.8064, 0.0599)	48 (0.7980, 0.0557)	50 (0.7984, 0.0529)
			$\boldsymbol{\delta}_4$	44 (0.8126, 0.0677)	46 (0.8127, 0.0643)	46 (0.7951, 0.0558)	48 (0.7987, 0.0541)
		MM	$\boldsymbol{\delta}_2$	51 (0.8050, 0.0638)	53 (0.8080, 0.0645)	53 (0.8058, 0.0526)	55 (0.8066, 0.0516)
			$\boldsymbol{\delta}_3$	47 (0.8060, 0.0621)	49 (0.8117, 0.0604)	49 (0.8112, 0.0522)	51 (0.8005, 0.0506)
			$\boldsymbol{\delta}_4$	44 (0.8023, 0.0658)	46 (0.8095, 0.0689)	46 (0.8014, 0.0580)	48 (0.8076, 0.0552)

Table 2.2. Required number of clusters (empirical power, empirical type I error) for closed-cohort SW-CRTs with between-subject correlation $\rho_2 = 0.05$

		Missing		GEE		Adjusted GEE	
	J	Pattern	$\boldsymbol{\delta}$	$\rho_1 = 0.15$	$\rho_1 = 0.30$	$\rho_1 = 0.15$	$\rho_1 = 0.30$
CS	40	IM	$\boldsymbol{\delta}_1$	40 (0.8112, 0.0724)	41 (0.8069, 0.0659)	42 (0.7939, 0.0532)	43 (0.7879, 0.0553)
			$\boldsymbol{\delta}_2$	43 (0.8162, 0.0684)	44 (0.8141, 0.0663)	45 (0.8024, 0.0509)	46 (0.7983, 0.0497)
			$\boldsymbol{\delta}_3$	42 (0.8163, 0.0680)	43 (0.8136, 0.0666)	44 (0.8060, 0.0534)	45 (0.8004, 0.0561)
			$\boldsymbol{\delta}_4$	41 (0.8197, 0.0695)	42 (0.8120, 0.0682)	43 (0.7956, 0.0484)	44 (0.8077, 0.0490)
		MM	$\boldsymbol{\delta}_2$	43 (0.8064, 0.0678)	44 (0.8075, 0.0607)	45 (0.7997, 0.0564)	46 (0.7871, 0.0505)
			$\boldsymbol{\delta}_3$	42 (0.8105, 0.0678)	43 (0.8097, 0.0680)	44 (0.8023, 0.0514)	45 (0.8051, 0.0549)
			$\boldsymbol{\delta}_4$	41 (0.8164, 0.0716)	42 (0.8042, 0.0699)	43 (0.7999, 0.0530)	44 (0.7964, 0.0578)
	20	IM	$\boldsymbol{\delta}_1$	48 (0.8112, 0.0658)	51 (0.8113, 0.0660)	50 (0.7965, 0.0521)	53 (0.7992, 0.0521)
			$\boldsymbol{\delta}_2$	53 (0.8085, 0.0623)	56 (0.8062, 0.0653)	55 (0.8015, 0.0506)	58 (0.7978, 0.0536)
			$\boldsymbol{\delta}_3$	51 (0.8040, 0.0657)	54 (0.8060, 0.0647)	53 (0.8016, 0.0530)	56 (0.8112, 0.0511)
			$\boldsymbol{\delta}_4$	49 (0.8057, 0.0661)	52 (0.8092, 0.0676)	51 (0.7919, 0.0516)	54 (0.8019, 0.0572)
		MM	$\boldsymbol{\delta}_2$	54 (0.8071, 0.0666)	57 (0.8037, 0.0611)	56 (0.7970, 0.0570)	59 (0.7946, 0.0509)
			$\boldsymbol{\delta}_3$	51 (0.8097, 0.0668)	54 (0.8109, 0.0608)	53 (0.7926, 0.0525)	56 (0.8013, 0.0525)
			$\boldsymbol{\delta}_4$	49 (0.8078, 0.0647)	52 (0.8050, 0.0596)	51 (0.7935, 0.0511)	54 (0.8033, 0.0553)
AR(1)	40	IM	$\boldsymbol{\delta}_1$	43 (0.7994, 0.0671)	44 (0.8086, 0.0683)	45 (0.8007, 0.0574)	46 (0.7973, 0.0579)
			$\boldsymbol{\delta}_2$	46 (0.8070, 0.0649)	47 (0.8139, 0.0681)	48 (0.7979, 0.0533)	49 (0.7982, 0.0553)
			$\boldsymbol{\delta}_3$	45 (0.8062, 0.0658)	46 (0.8099, 0.0662)	47 (0.7983, 0.0527)	48 (0.8053, 0.0525)
			$\boldsymbol{\delta}_4$	44 (0.8086, 0.0612)	45 (0.8125, 0.0622)	46 (0.8041, 0.0511)	47 (0.8127, 0.0580)
		MM	$\boldsymbol{\delta}_2$	47 (0.8056, 0.0668)	49 (0.8110, 0.0674)	49 (0.8015, 0.0556)	51 (0.8063, 0.0566)
			$\boldsymbol{\delta}_3$	45 (0.8024, 0.0668)	47 (0.8107, 0.0684)	47 (0.8006, 0.0546)	49 (0.8079, 0.0564)
			$\boldsymbol{\delta}_4$	44 (0.8078, 0.0663)	45 (0.8105, 0.0673)	46 (0.8052, 0.0528)	47 (0.8043, 0.0545)
	20	IM	$\boldsymbol{\delta}_1$	55 (0.8159, 0.0631)	57 (0.8154, 0.0668)	57 (0.8056, 0.0552)	59 (0.7952, 0.0567)
			$\boldsymbol{\delta}_2$	60 (0.8063, 0.0620)	62 (0.8067, 0.0647)	62 (0.8064, 0.0520)	64 (0.8014, 0.0525)
			$\boldsymbol{\delta}_3$	58 (0.8057, 0.0659)	60 (0.8064, 0.0651)	60 (0.8018, 0.0516)	62 (0.7981, 0.0543)
			$\boldsymbol{\delta}_4$	56 (0.8057, 0.0631)	58 (0.8080, 0.0611)	58 (0.7978, 0.0507)	60 (0.8003, 0.0513)
		MM	$\boldsymbol{\delta}_2$	63 (0.8192, 0.0570)	65 (0.8061, 0.0617)	65 (0.8011, 0.0531)	67 (0.8025, 0.0530)
			$\boldsymbol{\delta}_3$	59 (0.8074, 0.0625)	61 (0.8039, 0.0602)	61 (0.8051, 0.0550)	63 (0.7951, 0.0529)
			$\boldsymbol{\delta}_4$	56 (0.8001, 0.0640)	58 (0.8061, 0.0599)	58 (0.8003, 0.0522)	60 (0.7997, 0.0495)

Table 2.3. Required number of clusters (empirical power, empirical type I error) for cross-sectional SW-CRTs

	GEE		Adjusted GEE	
J	$\rho = 0.03$	$\rho = 0.05$	$\rho = 0.03$	$\rho = 0.05$
40	27 (0.8132, 0.0796)	39 (0.8006, 0.0745)	29 (0.7979, 0.0522)	41 (0.7911, 0.0526)
20	35 (0.8158, 0.0703)	47 (0.8075, 0.0634)	37 (0.8000, 0.0495)	49 (0.8026, 0.0491)

The proposed method was developed based on the GEE approach, which requires the large sample theory to ensure the validity of statistical inference. When the number of clusters is small, even if the total number of unique patients is large, the variance of treatment effect can be underestimated, which leads to inflated type I error rate. To adjust for the estimation bias under small numbers of clusters, researchers have proposed many methods, such as making correction on the GEE variance estimator (Mancl and DeRouen, 2001; Kauermann and Carroll, 2001; Ziegler and Vens, 2010; Morel et al., 2003), making inferences based on t-distribution or F-distribution, or combination of different adjustment methods (Fay and Graubard, 2001; Pan and Wall, 2002; McCaffrey and Bell, 2006; Fan et al., 2013). The first two columns of Tables 2.1-2.3 show results based on GEE approach without correction. We observe that the empirical powers and type I errors are both inflated, a typical symptom of underestimated variance. We explored various adjustment methods and found that the combination of Morel et al. (2003) (MBN) and Donner and Klar (2000) offers a good balance between easy implementation and maintaining the power and type I error at nominal levels. Specifically, the MBN method adds a correction term to the GEE covariance estimator. This correction term converges to zero as the number of clusters increases. Donner and Klar (2000) suggested adding one cluster per intervention when the sample size is determined under 0.05 type I error. Simulation results based on the adjusted GEE are presented in the last two columns of Tables 2.1-2.3, which show that after correction, the empirical powers and type I errors are close to the nominal levels of 0.8 and 0.05, respectively.

2.5. A real application example

We illustrate the proposed method in the determination of sample size for a SW-CRT evaluating the effect of a physician training program (in interpersonal and communication skills) on female patients' satisfaction with doctor-women relationship in labor and delivery rooms (Bashour et al., 2013). The primary outcome is defined as the total score over a series of questions on a Likert scale modified from the Medical Interview Satisfaction Scale. The data will be collected $T = 5$ times evenly over a 10-month follow-up period. Preliminary data indicates that the mean of the satisfaction score is 3.23 (SD 0.72) in the control group and 3.42 (SD 0.73) in the intervention group. A balanced randomization will be employed, i.e., $p_s = \frac{1}{4}$ for $s = 1, \ldots, 4$. We would like to detect a difference of 0.2 in mean satisfaction score with 80% power at a two-sided significance level of 0.05. For a closed-cohort SW-CRT where the within-subject correlation is CS with $\rho_1 = 0.15$ and between-subject correlation is $\boldsymbol{\Phi} = \mathbf{1}\mathbf{1}'\rho_2$ with $\rho_2 = 0.03$, the required numbers of clusters are 36 and 40 for complete data and missing data with $\boldsymbol{\delta} = (1.00, 0.85, 0.80, 0.75, 0.70)$ under IM missing pattern, respectively when cluster size is 20. If cluster size is 50, the required numbers of clusters are 26 and 28, respectively. For a cross-sectional SW-CRT with ICC $\rho = 0.03$, the required numbers of clusters are 35 and 25 when cluster size is 20 and 50, respectively.

2.6. Extension to outcomes from the exponential family

Besides continuous outcomes, binary and count patient-centered outcomes are often encountered in clinical trials. For those types of outcomes, binomial and Poisson models can be employed. A similar sample size derivation approach based on GEE can be implemented because they all belong to the exponential family. The procedure for outcomes from the exponential family is developed as follows. Suppose Y_{ijt} follows a distribution from an exponential family with mean $\boldsymbol{\mu_{ij}} = E(\boldsymbol{Y_{ij}})$ and a link function $g(\boldsymbol{\mu_{ij}})$, we have

$$g(\boldsymbol{\mu_{ij}}) = \boldsymbol{\lambda} + \boldsymbol{u}_i\zeta = \boldsymbol{X}_i\boldsymbol{\beta}.$$

The correlation structure of $\boldsymbol{Y_i} = \left(\boldsymbol{Y_{i1}'}, \ldots, \boldsymbol{Y_{iJ}'}\right)'$ is

$$Corr\left(\boldsymbol{Y_i}\right) = \boldsymbol{I_J} \otimes \left(\boldsymbol{\Omega} - \boldsymbol{\Phi}\right) + \left(\boldsymbol{1_J 1_J'}\right) \otimes \boldsymbol{\Phi}.$$

The variance of $\boldsymbol{Y_{ij}}$ is shown in Table 2.4.

Assuming an independent working correlation structure, the GEE estimator of $\boldsymbol{\beta}$ is obtained by solving the score functions $\boldsymbol{U}_n\left(\boldsymbol{\beta}\right) = 0$, where

$$\boldsymbol{U}_n\left(\boldsymbol{\beta}\right) = \left\{ \begin{array}{c} \sum_{i=1}^{n}\sum_{j=1}^{J}\left(Y_{ij1} - \mu_{ij1}\left(\boldsymbol{\beta}\right)\right) \\ \vdots \\ \sum_{i=1}^{n}\sum_{j=1}^{J}\left(Y_{ijT} - \mu_{ijT}\left(\boldsymbol{\beta}\right)\right) \\ \sum_{i=1}^{n}\sum_{j=1}^{J}\left[\boldsymbol{u_i'}\left(\boldsymbol{Y_{ij}} - \boldsymbol{\mu_{ij}}\left(\boldsymbol{\beta}\right)\right)\right] \end{array} \right\}$$

with $\boldsymbol{\mu_{ij}}\left(\boldsymbol{\beta}\right) = g^{-1}\left(\boldsymbol{X_i\beta}\right)$.

Under mild regularity conditions, as $n \to \infty$, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_n\right)$ in distribution. The covariance matrix $\boldsymbol{\Sigma}_n$ can be consistently estimated by $\boldsymbol{A}_n^{-1}\boldsymbol{E}_n\boldsymbol{A}_n^{-1}$, where

$$\boldsymbol{A}_n = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\boldsymbol{X_i'B_{ij}X_i}$$

with $\boldsymbol{B_{ij}}$ being a diaginal matix as shown in Table 2.4.

$$\boldsymbol{E}_n = n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\boldsymbol{X_i'}\hat{\boldsymbol{e}}_{\boldsymbol{ij}}\right)^{\otimes 2}$$

$\hat{\boldsymbol{e}}_{\boldsymbol{ij}} = \boldsymbol{y_{ij}} - \boldsymbol{\mu_{ij}}\left(\hat{\boldsymbol{\beta}}\right)$ is the residual and $\boldsymbol{c}^{\otimes 2} = \boldsymbol{cc'}$ for a vector $\boldsymbol{c}$.

Let $\boldsymbol{A}$ and $\boldsymbol{E}$ be the limits of $\boldsymbol{A}_n$ and $\boldsymbol{E}_n$ as $n \to \infty$, then $\boldsymbol{\Sigma}_n$ converges to $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{E}\boldsymbol{A}^{-1}$, where

$$\boldsymbol{A} = J\sum_{s=1}^{S}p_s\boldsymbol{W_s'B_sW_s}$$

and

$$\boldsymbol{E} = J\sum_{s=1}^{S}p_s\boldsymbol{W_s'G_s}\left[\boldsymbol{\Omega} + \left(J-1\right)\boldsymbol{\Phi}\right]\boldsymbol{G_sW_s}, \tag{2.11}$$

where $\boldsymbol{W_s} = (\boldsymbol{I_T}, \boldsymbol{v_s})$. $\boldsymbol{B_s}$ and $\boldsymbol{G_s}$ are shown in Table 2.4. The derivation is shown in Appendix. To reject the null hypothesis $H_0 : \zeta = 0$ with a power $1-\gamma$ at a significance level of α, the required number of clusters is

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma_\zeta^2}{\zeta_0^2}$$

where ζ_0 is the true intervention effect and σ_ζ^2 is the $(T+1, T+1)$-component of $\boldsymbol{\Sigma}$.

If missing data is considered with missing indicator Δ_{ijt}, Under the assumption of missing complete at random (MCAR), $\boldsymbol{A}_n$ and $\boldsymbol{\Sigma}_n$ can be rewritten as

$$\boldsymbol{A}_n^* = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J} \boldsymbol{X_i'} diag\left(\boldsymbol{\Delta_{ij}}\right) \boldsymbol{B_{ij}} \boldsymbol{X_i}$$

and

$$\boldsymbol{E}_n^* = n^{-1} \sum_{i=1}^{n} \left(\sum_{j=1}^{J} \boldsymbol{X_i'} (\boldsymbol{\Delta_{ij}} \circ \hat{\boldsymbol{e}}_{\boldsymbol{ij}}) \right)^{\otimes 2}$$

As $n \to \infty$, $\boldsymbol{A}_n^* \to \boldsymbol{A}^*$ and $\boldsymbol{E}_n^* \to \boldsymbol{E}^*$, where

$$\boldsymbol{A}^* = J \sum_{s=1}^{S} p_s \boldsymbol{W_s'} diag\left(\boldsymbol{\delta}\right) \boldsymbol{B_s} \boldsymbol{W_s}$$

and

$$\boldsymbol{E}^* = J \sum_{s=1}^{S} p_s \boldsymbol{W_s'} \left[\widetilde{\boldsymbol{\delta}} \circ \boldsymbol{G_s} \boldsymbol{\Omega} \boldsymbol{G_s} + (J-1)\, diag\left(\boldsymbol{\delta}\right) \boldsymbol{G_s} \boldsymbol{\Phi} \boldsymbol{G_s} diag\left(\boldsymbol{\delta}\right) \right] \boldsymbol{W_s}$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_T)^{'}$, $\widetilde{\boldsymbol{\delta}}$ is a matrix with diagonal (t,t)th element being δ_t and off-diagonal $(t,t^{'})$th element being $\delta_{tt'}$. Then, the required number of clusters can be calculated using

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma_\zeta^2}{\zeta_0^2},$$

where σ_ζ^2 is the $(T+1, T+1)$-component of $\boldsymbol{A}^{*-1} \boldsymbol{E}^* \boldsymbol{A}^{*-1}$.

Table 2.4. Notations for different types of outcome

	Continuous	Binary	Count
Distribution	Normal	Bernoulli	Poisson
$g(\boldsymbol{\mu_{ij}})$	$\boldsymbol{\mu_{ij}}$	$ln\left(\frac{\boldsymbol{\mu_{ij}}}{1-\boldsymbol{\mu_{ij}}}\right)$	$ln(\boldsymbol{\mu_{ij}})$
$Var(\mathbf{Y_{ij}})$	$\begin{bmatrix} \sigma^2 \\ \vdots \\ \sigma^2 \end{bmatrix}$	$\begin{bmatrix} \mu_{ij1}(1-\mu_{ij1}) \\ \vdots \\ \mu_{ijT}(1-\mu_{ijT}) \end{bmatrix}$	$\begin{bmatrix} \mu_{ij1} \\ \vdots \\ \mu_{ijT} \end{bmatrix}$
$\boldsymbol{B_{ij}}$	$\boldsymbol{I_T}$	$\begin{bmatrix} \mu_{ij1}(1-\mu_{ij1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_{ijT}(1-\mu_{ijT}) \end{bmatrix}$	$\begin{bmatrix} \mu_{ij1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_{ijT} \end{bmatrix}$
$\boldsymbol{B_s}$	$\boldsymbol{I_T}$	$\begin{bmatrix} \mu_{s1}(1-\mu_{s1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_{sT}(1-\mu_{sT}) \end{bmatrix}$	$\begin{bmatrix} \mu_{s1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_{sT} \end{bmatrix}$
$\boldsymbol{G_s}$	$\sigma\boldsymbol{B_s^{1/2}}$	$\boldsymbol{B_s^{1/2}}$	$\boldsymbol{B_s^{1/2}}$

μ_{st} is the mean of outcomes from sth sequence at tth time point.

We conducted simulation studies for closed-cohort and cross-sectional SW-CRTs with binary and count outcomes. Suppose subjects were transited into intervention over $T = 4$ time points and cluster size was $J = 15$. For binary outcomes, we set time-specific intercepts $\lambda_t = 0.01\,(t-1)$ for $t = 1, \ldots, T$ and $\zeta = \{log(1.5), log(1.8)\}$ corresponding to odds ratio 1.5 and 1.8, respectively. For count outcomes, we set time-specific intercepts $\lambda_t = 1 + 0.3\,(t-1)$ for $t = 1, \ldots, T$ and $\zeta = \{0.10, 0.13\}$. For closed-cohort SW-CRTs, we explored CS and AR(1) structures for within-subject correlation matrix $\boldsymbol{\Omega}$ and $\boldsymbol{\Phi} = \mathbf{1}\mathbf{1}'\rho_2$ for between-subject correlation matrix. We also explored $(\rho_1, \rho_2) = \{(0.2, 0.03), (0.2, 0.05), (0.4, 0.03), (0.4, 0.05)\}$ to assess the impact of correlations on sample size. Moreover, we investigated different missing patterns with independent missing (IM) and monotone missing (MM), as well as four sets of observational probabilities:

$$\begin{aligned}
\boldsymbol{\delta}_1 &= (1.00,\ 1.00,\ 1.00,\ 1.00)\,, \\
\boldsymbol{\delta}_2 &= (1.00,\ 0.80,\ 0.75,\ 0.70)\,, \\
\boldsymbol{\delta}_3 &= (1.00,\ 0.90,\ 0.80,\ 0.70)\,, \\
\boldsymbol{\delta}_4 &= (1.00,\ 1.00,\ 0.85,\ 0.70)\,.
\end{aligned}$$

For cross-sectional SW-CRTs, we specified $\boldsymbol{\Omega} = \mathbf{1}\mathbf{1}'\rho + (1-\rho)\boldsymbol{I}$ and $\boldsymbol{\Phi} = \mathbf{1}\mathbf{1}'\rho$, as was discussed in SW-CRTs with continuous outcomes. We explored two values for the ICC: $\rho = 0.03$ and 0.05. To test the null hypothesis $H_0 : \zeta = 0$, we set the two-sided type I error rate $\alpha = 0.05$ and power $1 - \gamma = 0.8$, respectively, and assumed balanced randomization $p_1 = \cdots = p_S = 1/3$.

Tables 2.5-2.6 and Tables 2.8-2.9 present the required number of clusters with empirical power and type I error for closed-cohort SW-CRTs with binary and count outcomes, respectively. We have several observations. First, more clusters are required when between-subject correlation and longitudinal correlation get larger. For example, the first row in Table 2.5, the required number of clusters changes from 51 to 54 when longitudinal correlation (ρ_1) increases from 0.2 to 0.4, respectively. On the other hand, the first cells in Tables 2.5-2.6, the required number of clusters increases from 51 to 63 when between-subject correlation

(ρ_2) increases from 0.03 to 0.05. Second, the longitudinal correlation structures affect the required number of clusters, which can be shown by comparing the first and second panels in each table for closed-cohort SW-CRTs. Third, larger ζ leads to smaller sample size, which can be easily proofed theoretically. Fourth, different missing patterns and observational probabilities affect the required number of clusters. The missing scenario with subjects dropping quickly initially ($\boldsymbol{\delta}_2$) requires more additional clusters to achieve the desired power. On the other hand, the scenario with subjects dropping quickly near the end of the trial ($\boldsymbol{\delta}_4$) requires less additional clusters to compensate the loss due to missing compared to the other missing scenarios with the same dropout rate at the end of the study. Tables 2.7 and 2.10 present the required number of clusters with empirical power and type I error for cross-sectional SW-CRTs with binary and count outcomes, respectively. Similar to observations from closed-cohort SW-CRTs, the larger ζ and smaller correlations lead to smaller required number of clusters.

The last two columns in Tables 2.5-2.10 show the empirical powers and type I errors after adjustment using the combination of MBN and Donner and Klar's method as described previously in SW-CRTs with continuous outcomes. The results show that after adjustment, the empirical powers and type I errors are very close to their nominal values, which are 0.8 and 0.05, respectively. For example, in Table 2.5 when required number of clusters is less than 30, the inflated type I error is larger than 0.07. After adjustment, the type I error drops to around 0.05. Therefore, this proposed combination method for adjusting inflated type I error works very well in SW-CRTs not only for continuous outcomes, but also for binary and count outcomes.

Table 2.5. Required number of clusters (empirical power, empirical type I error) for binary outcomes with between-subject correlation $\rho_2 = 0.03$

		Missing		GEE		Adjusted GEE	
	ζ	Pattern	$\boldsymbol{\delta}$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.2$	$\rho_1 = 0.4$
CS	log(1.5)	IM	$\boldsymbol{\delta}_1$	51 (0.7998, 0.0587)	54 (0.8063, 0.0576)	53 (0.7920, 0.0518)	56 (0.8007, 0.0479)
			$\boldsymbol{\delta}_2$	60 (0.8030, 0.0566)	63 (0.8140, 0.0581)	62 (0.8050, 0.0499)	65 (0.8061, 0.0483)
			$\boldsymbol{\delta}_3$	56 (0.8014, 0.0615)	59 (0.8046, 0.0664)	58 (0.8019, 0.0494)	61 (0.7959, 0.0489)
			$\boldsymbol{\delta}_4$	54 (0.8124, 0.0596)	56 (0.8003, 0.0636)	56 (0.8004, 0.0504)	58 (0.7978, 0.0534)
		MM	$\boldsymbol{\delta}_2$	60 (0.8110, 0.0643)	64 (0.7993, 0.0601)	62 (0.8013, 0.0549)	66 (0.7993, 0.0496)
			$\boldsymbol{\delta}_3$	57 (0.8073, 0.0616)	60 (0.8074, 0.0595)	59 (0.8047, 0.0497)	62 (0.8026, 0.0515)
			$\boldsymbol{\delta}_4$	54 (0.8167, 0.0596)	56 (0.8075, 0.0642)	56 (0.8072, 0.0494)	58 (0.7961, 0.0520)
	log(1.8)	IM	$\boldsymbol{\delta}_1$	25 (0.8165, 0.0737)	27 (0.8233, 0.0745)	27 (0.7979, 0.0498)	29 (0.8052, 0.0525)
			$\boldsymbol{\delta}_2$	29 (0.8108, 0.0757)	31 (0.8127, 0.0734)	31 (0.7928, 0.0497)	33 (0.8042, 0.0546)
			$\boldsymbol{\delta}_3$	28 (0.8189, 0.0732)	29 (0.8095, 0.0703)	30 (0.8020, 0.0480)	31 (0.7952, 0.0496)
			$\boldsymbol{\delta}_4$	26 (0.8124, 0.0758)	28 (0.8144, 0.0755)	28 (0.7971, 0.0474)	30 (0.8074, 0.0525)
		MM	$\boldsymbol{\delta}_2$	30 (0.8244, 0.0716)	31 (0.8137, 0.0704)	32 (0.8107, 0.0499)	33 (0.7991, 0.0491)
			$\boldsymbol{\delta}_3$	28 (0.8180, 0.0737)	29 (0.8180, 0.0713)	30 (0.7976, 0.0491)	31 (0.7983, 0.0515)
			$\boldsymbol{\delta}_4$	26 (0.8062, 0.0748)	28 (0.8189, 0.0740)	28 (0.7925, 0.0514)	30 (0.8063, 0.0511)
AR(1)	log(1.5)	IM	$\boldsymbol{\delta}_1$	57 (0.8104, 0.0567)	59 (0.8019, 0.0611)	59 (0.7993, 0.0517)	61 (0.8102, 0.0505)
			$\boldsymbol{\delta}_2$	65 (0.8055, 0.0610)	68 (0.8038, 0.0602)	67 (0.8008, 0.0506)	70 (0.8028, 0.0555)
			$\boldsymbol{\delta}_3$	62 (0.8128, 0.0606)	64 (0.7968, 0.0604)	64 (0.8014, 0.0550)	66 (0.8017, 0.0483)
			$\boldsymbol{\delta}_4$	59 (0.8023, 0.0629)	61 (0.8092, 0.0608)	61 (0.7988, 0.0518)	63 (0.7945, 0.0520)
		MM	$\boldsymbol{\delta}_2$	67 (0.8032, 0.0587)	70 (0.8093, 0.0606)	69 (0.7985, 0.0494)	72 (0.7996, 0.0467)
			$\boldsymbol{\delta}_3$	63 (0.8147, 0.0619)	65 (0.8027, 0.0614)	65 (0.8029, 0.0544)	67 (0.7977, 0.0513)
			$\boldsymbol{\delta}_4$	59 (0.8015, 0.0606)	61 (0.8007, 0.0637)	61 (0.7983, 0.0494)	63 (0.7901, 0.0528)
	log(1.8)	IM	$\boldsymbol{\delta}_1$	28 (0.8145, 0.0720)	29 (0.8129, 0.0723)	30 (0.8061, 0.0507)	31 (0.8038, 0.0524)
			$\boldsymbol{\delta}_2$	32 (0.8227, 0.0720)	33 (0.8185, 0.0704)	34 (0.8015, 0.0508)	35 (0.8043, 0.0513)
			$\boldsymbol{\delta}_3$	30 (0.8104, 0.0732)	32 (0.8247, 0.0714)	32 (0.7948, 0.0495)	34 (0.8107, 0.0490)
			$\boldsymbol{\delta}_4$	29 (0.8086, 0.0779)	30 (0.8111, 0.0661)	31 (0.7925, 0.0495)	32 (0.8018, 0.0541)
		MM	$\boldsymbol{\delta}_2$	33 (0.8173, 0.0692)	34 (0.8077, 0.0715)	35 (0.7963, 0.0486)	36 (0.8042, 0.0534)
			$\boldsymbol{\delta}_3$	31 (0.8157, 0.0731)	32 (0.8108, 0.0772)	33 (0.8090, 0.0521)	34 (0.8013, 0.0567)
			$\boldsymbol{\delta}_4$	29 (0.8182, 0.0762)	30 (0.8067, 0.0765)	31 (0.7985, 0.0540)	32 (0.8033, 0.0562)

Table 2.6. Required number of clusters (empirical power, empirical type I error) for binary outcomes with between-subject correlation $\rho_2 = 0.05$

		Missing		GEE		Adjusted GEE	
	ζ	Pattern	$\boldsymbol{\delta}$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.2$	$\rho_1 = 0.4$
CS	log(1.5)	IM	$\boldsymbol{\delta}_1$	63 (0.8003, 0.0638)	66 (0.8033, 0.0576)	65 (0.7973, 0.0513)	68 (0.7923, 0.0502)
			$\boldsymbol{\delta}_2$	72 (0.8039, 0.0571)	75 (0.8082, 0.0598)	74 (0.8043, 0.0486)	77 (0.8079, 0.0486)
			$\boldsymbol{\delta}_3$	69 (0.8086, 0.0582)	72 (0.8051, 0.0578)	71 (0.8039, 0.0527)	74 (0.8078, 0.0523)
			$\boldsymbol{\delta}_4$	66 (0.8004, 0.0625)	69 (0.8064, 0.0563)	68 (0.7984, 0.0499)	71 (0.7979, 0.0490)
		MM	$\boldsymbol{\delta}_2$	73 (0.8109, 0.0613)	76 (0.8083, 0.0583)	75 (0.8074, 0.0565)	78 (0.8001, 0.0500)
			$\boldsymbol{\delta}_3$	69 (0.8048, 0.0590)	72 (0.8038, 0.0609)	71 (0.7981, 0.0505)	74 (0.7958, 0.0462)
			$\boldsymbol{\delta}_4$	66 (0.8045, 0.0634)	69 (0.8079, 0.0631)	68 (0.7950, 0.0536)	71 (0.8035, 0.0547)
	log(1.8)	IM	$\boldsymbol{\delta}_1$	31 (0.8105, 0.0680)	33 (0.8210, 0.0708)	33 (0.7973, 0.0534)	35 (0.8059, 0.0504)
			$\boldsymbol{\delta}_2$	35 (0.8173, 0.0695)	37 (0.8164, 0.0665)	37 (0.7972, 0.0462)	39 (0.8092, 0.0521)
			$\boldsymbol{\delta}_3$	34 (0.8097, 0.0667)	35 (0.8144, 0.0711)	36 (0.8081, 0.0496)	37 (0.7919, 0.0517)
			$\boldsymbol{\delta}_4$	32 (0.8140, 0.0717)	34 (0.8145, 0.0693)	34 (0.8019, 0.0542)	36 (0.8008, 0.0549)
		MM	$\boldsymbol{\delta}_2$	36 (0.8233, 0.0687)	37 (0.8055, 0.0688)	38 (0.8064, 0.0515)	39 (0.7922, 0.0504)
			$\boldsymbol{\delta}_3$	34 (0.8105, 0.0753)	35 (0.8061, 0.0661)	36 (0.8023, 0.0543)	37 (0.7991, 0.0527)
			$\boldsymbol{\delta}_4$	32 (0.8075, 0.0703)	34 (0.8110, 0.0710)	34 (0.7917, 0.0516)	36 (0.8011, 0.0533)
AR(1)	log(1.5)	IM	$\boldsymbol{\delta}_1$	69 (0.8043, 0.0599)	71 (0.8022, 0.0569)	71 (0.7966, 0.0502)	73 (0.7955, 0.0484)
			$\boldsymbol{\delta}_2$	78 (0.8087, 0.0572)	80 (0.8042, 0.0536)	80 (0.8002, 0.0465)	82 (0.7966, 0.0502)
			$\boldsymbol{\delta}_3$	74 (0.8056, 0.0552)	77 (0.8091, 0.0549)	76 (0.8028, 0.0485)	79 (0.8049, 0.0502)
			$\boldsymbol{\delta}_4$	72 (0.8010, 0.0591)	74 (0.8047, 0.0575)	74 (0.8056, 0.0535)	76 (0.8058, 0.0492)
		MM	$\boldsymbol{\delta}_2$	80 (0.8107, 0.0596)	83 (0.8071, 0.0551)	82 (0.8008, 0.0523)	85 (0.8080, 0.0515)
			$\boldsymbol{\delta}_3$	75 (0.8090, 0.0621)	78 (0.8104, 0.0548)	77 (0.8020, 0.0508)	80 (0.7993, 0.0507)
			$\boldsymbol{\delta}_4$	72 (0.7978, 0.0604)	74 (0.8116, 0.0543)	74 (0.7998, 0.0499)	76 (0.8046, 0.0506)
	log(1.8)	IM	$\boldsymbol{\delta}_1$	34 (0.8130, 0.0697)	35 (0.8111, 0.0732)	36 (0.8052, 0.0536)	37 (0.7972, 0.0556)
			$\boldsymbol{\delta}_2$	38 (0.8156, 0.0622)	39 (0.8186, 0.0687)	40 (0.8014, 0.0511)	41 (0.8047, 0.0490)
			$\boldsymbol{\delta}_3$	36 (0.8094, 0.0699)	38 (0.8201, 0.0713)	38 (0.7907, 0.0520)	40 (0.8030, 0.0564)
			$\boldsymbol{\delta}_4$	35 (0.8130, 0.0711)	36 (0.8066, 0.0686)	37 (0.8021, 0.0528)	38 (0.8022, 0.0512)
		MM	$\boldsymbol{\delta}_2$	39 (0.8137, 0.0661)	40 (0.8072, 0.0641)	41 (0.8005, 0.0536)	42 (0.8047, 0.0531)
			$\boldsymbol{\delta}_3$	37 (0.8182, 0.0670)	38 (0.8077, 0.0677)	39 (0.7950, 0.0540)	40 (0.8041, 0.0570)
			$\boldsymbol{\delta}_4$	35 (0.8100, 0.0678)	36 (0.8205, 0.0681)	37 (0.7941, 0.0542)	38 (0.7980, 0.0522)

Table 2.7. Required number of clusters (empirical power, empirical type I error) for cross-sectional SW-CRTs with binary outcomes

	GEE		Adjusted GEE	
ζ	$\rho = 0.03$	$\rho = 0.05$	$\rho = 0.03$	$\rho = 0.05$
log(1.5)	49 (0.8077, 0.0600)	61 (0.8116, 0.0616)	51 (0.7935, 0.0495)	63 (0.7919, 0.0543)
log(1.8)	24 (0.8091, 0.0850)	30 (0.8136, 0.0781)	26 (0.7948, 0.0492)	32 (0.8001, 0.0511)

Table 2.8. Required number of clusters (empirical power, empirical type I error) for count outcomes with between-subject correlation $\rho_2 = 0.03$

		Missing		GEE		Adjusted GEE	
	ζ	Pattern	$\boldsymbol{\delta}$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.2$	$\rho_1 = 0.4$
CS	0.10	IM	$\boldsymbol{\delta}_1$	46 (0.8191, 0.0668)	48 (0.8176, 0.0662)	48 (0.8149, 0.0525)	50 (0.8037, 0.0541)
			$\boldsymbol{\delta}_2$	53 (0.8105, 0.0661)	56 (0.8150, 0.0640)	55 (0.8049, 0.0541)	58 (0.8033, 0.0534)
			$\boldsymbol{\delta}_3$	51 (0.8169, 0.0570)	53 (0.8079, 0.0625)	53 (0.8049, 0.0537)	55 (0.7958, 0.0546)
			$\boldsymbol{\delta}_4$	48 (0.8082, 0.0644)	51 (0.8136, 0.0640)	50 (0.8044, 0.0526)	53 (0.8087, 0.0544)
		MM	$\boldsymbol{\delta}_2$	54 (0.8104, 0.0650)	57 (0.8059, 0.0621)	56 (0.8023, 0.0498)	59 (0.8005, 0.0493)
			$\boldsymbol{\delta}_3$	51 (0.8209, 0.0607)	54 (0.8149, 0.0623)	53 (0.8101, 0.0523)	56 (0.8001, 0.0506)
			$\boldsymbol{\delta}_4$	48 (0.8037, 0.0658)	51 (0.8120, 0.0577)	50 (0.7980, 0.0521)	53 (0.8103, 0.0524)
	0.13	IM	$\boldsymbol{\delta}_1$	27 (0.8215, 0.0792)	28 (0.8145, 0.0781)	29 (0.8053, 0.0524)	30 (0.8030, 0.0545)
			$\boldsymbol{\delta}_2$	31 (0.8064, 0.0692)	33 (0.8198, 0.0735)	33 (0.7981, 0.0535)	35 (0.8092, 0.0537)
			$\boldsymbol{\delta}_3$	30 (0.8155, 0.0746)	31 (0.8119, 0.0775)	32 (0.8040, 0.0505)	33 (0.7960, 0.0563)
			$\boldsymbol{\delta}_4$	28 (0.8188, 0.0750)	30 (0.8182, 0.0739)	30 (0.7974, 0.0492)	32 (0.8097, 0.0504)
		MM	$\boldsymbol{\delta}_2$	32 (0.8224, 0.0700)	34 (0.8201, 0.0766)	34 (0.8052, 0.0547)	36 (0.8126, 0.0516)
			$\boldsymbol{\delta}_3$	30 (0.8118, 0.0791)	32 (0.8255, 0.0708)	32 (0.7969, 0.0534)	34 (0.8064, 0.0534)
			$\boldsymbol{\delta}_4$	28 (0.8212, 0.0738)	30 (0.8244, 0.0782)	30 (0.7975, 0.0540)	32 (0.8078, 0.0516)
AR(1)	0.10	IM	$\boldsymbol{\delta}_1$	51 (0.8143, 0.0597)	52 (0.8140, 0.0639)	53 (0.8069, 0.0512)	54 (0.7973, 0.0470)
			$\boldsymbol{\delta}_2$	58 (0.8147, 0.0610)	60 (0.8134, 0.0569)	60 (0.8030, 0.0525)	62 (0.8020, 0.0530)
			$\boldsymbol{\delta}_3$	56 (0.8097, 0.0605)	58 (0.8139, 0.0611)	58 (0.8092, 0.0519)	60 (0.8043, 0.0535)
			$\boldsymbol{\delta}_4$	53 (0.8134, 0.0627)	55 (0.8131, 0.0639)	55 (0.8101, 0.0529)	57 (0.8002, 0.0518)
		MM	$\boldsymbol{\delta}_2$	60 (0.8156, 0.0632)	63 (0.8122, 0.0590)	62 (0.8103, 0.0502)	65 (0.8067, 0.0515)
			$\boldsymbol{\delta}_3$	56 (0.8135, 0.0595)	59 (0.8099, 0.0632)	58 (0.8015, 0.0562)	61 (0.8094, 0.0534)
			$\boldsymbol{\delta}_4$	53 (0.8116, 0.0627)	55 (0.8066, 0.0627)	55 (0.7982, 0.0543)	57 (0.8058, 0.0521)
	0.13	IM	$\boldsymbol{\delta}_1$	30 (0.8235, 0.0762)	31 (0.8153, 0.0689)	32 (0.8156, 0.0541)	33 (0.8084, 0.0544)
			$\boldsymbol{\delta}_2$	34 (0.8139, 0.0688)	35 (0.8125, 0.0713)	36 (0.8000, 0.0517)	37 (0.7992, 0.0512)
			$\boldsymbol{\delta}_3$	33 (0.8194, 0.0761)	34 (0.8177, 0.0725)	35 (0.8131, 0.0571)	36 (0.8074, 0.0536)
			$\boldsymbol{\delta}_4$	31 (0.8116, 0.0747)	32 (0.8169, 0.0764)	33 (0.7966, 0.0527)	34 (0.7969, 0.0528)
		MM	$\boldsymbol{\delta}_2$	35 (0.8207, 0.0721)	37 (0.8230, 0.0714)	37 (0.8022, 0.0516)	39 (0.8155, 0.0504)
			$\boldsymbol{\delta}_3$	33 (0.8165, 0.0707)	34 (0.8116, 0.0736)	35 (0.7996, 0.0554)	36 (0.8035, 0.0591)
			$\boldsymbol{\delta}_4$	31 (0.8137, 0.0750)	32 (0.8149, 0.0722)	33 (0.7980, 0.0549)	34 (0.7974, 0.0533)

Table 2.9. Required number of clusters (empirical power, empirical type I error) for count outcomes with between-subject correlation $\rho_2 = 0.05$

		Missing		GEE		Adjusted GEE	
	ζ	Pattern	$\boldsymbol{\delta}$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.2$	$\rho_1 = 0.4$
CS	0.10	IM	$\boldsymbol{\delta}_1$	56 (0.8113, 0.0623)	59 (0.8062, 0.0632)	58 (0.8014, 0.0524)	61 (0.8105, 0.0515)
			$\boldsymbol{\delta}_2$	64 (0.8063, 0.0602)	67 (0.8141, 0.0611)	66 (0.8096, 0.0574)	69 (0.8094, 0.0528)
			$\boldsymbol{\delta}_3$	61 (0.8064, 0.0605)	64 (0.8121, 0.0606)	63 (0.8020, 0.0498)	66 (0.8116, 0.0511)
			$\boldsymbol{\delta}_4$	59 (0.8105, 0.0604)	62 (0.8083, 0.0595)	61 (0.8115, 0.0549)	64 (0.8091, 0.0528)
		MM	$\boldsymbol{\delta}_2$	65 (0.8194, 0.0604)	68 (0.8219, 0.0589)	67 (0.8061, 0.0523)	70 (0.8118, 0.0561)
			$\boldsymbol{\delta}_3$	62 (0.8149, 0.0616)	65 (0.8123, 0.0626)	64 (0.8068, 0.0532)	67 (0.8058, 0.0511)
			$\boldsymbol{\delta}_4$	59 (0.8111, 0.0625)	62 (0.8114, 0.0630)	61 (0.8103, 0.0540)	64 (0.7954, 0.0536)
	0.13	IM	$\boldsymbol{\delta}_1$	33 (0.8172, 0.0695)	35 (0.8224, 0.0698)	35 (0.8055, 0.0516)	37 (0.8117, 0.0553)
			$\boldsymbol{\delta}_2$	38 (0.8209, 0.0710)	39 (0.8188, 0.0659)	40 (0.8060, 0.0526)	41 (0.8098, 0.0515)
			$\boldsymbol{\delta}_3$	36 (0.8121, 0.0673)	38 (0.8174, 0.0684)	38 (0.8045, 0.0500)	40 (0.8119, 0.0529)
			$\boldsymbol{\delta}_4$	35 (0.8250, 0.0691)	36 (0.8105, 0.0753)	37 (0.8045, 0.0522)	38 (0.8008, 0.0552)
		MM	$\boldsymbol{\delta}_2$	38 (0.8191, 0.0683)	40 (0.8115, 0.0695)	40 (0.7995, 0.0507)	42 (0.8105, 0.0534)
			$\boldsymbol{\delta}_3$	36 (0.8159, 0.0697)	38 (0.8222, 0.0694)	38 (0.8070, 0.0510)	40 (0.8119, 0.0528)
			$\boldsymbol{\delta}_4$	35 (0.8178, 0.0725)	36 (0.8161, 0.0696)	37 (0.8081, 0.0493)	38 (0.7968, 0.0526)
AR(1)	0.10	IM	$\boldsymbol{\delta}_1$	61 (0.8138, 0.0586)	63 (0.8068, 0.0609)	63 (0.7991, 0.0531)	65 (0.8164, 0.0486)
			$\boldsymbol{\delta}_2$	69 (0.8119, 0.0590)	71 (0.8013, 0.0595)	71 (0.8008, 0.0500)	73 (0.8005, 0.0513)
			$\boldsymbol{\delta}_3$	67 (0.8114, 0.0635)	69 (0.8146, 0.0646)	69 (0.8113, 0.0539)	71 (0.8153, 0.0509)
			$\boldsymbol{\delta}_4$	64 (0.8126, 0.0606)	66 (0.8083, 0.0590)	66 (0.8078, 0.0498)	68 (0.8099, 0.0514)
		MM	$\boldsymbol{\delta}_2$	71 (0.8101, 0.0619)	74 (0.8187, 0.0596)	73 (0.8076, 0.0498)	76 (0.8220, 0.0547)
			$\boldsymbol{\delta}_3$	67 (0.8095, 0.0526)	70 (0.8169, 0.0595)	69 (0.8050, 0.0483)	72 (0.8166, 0.0555)
			$\boldsymbol{\delta}_4$	64 (0.8192, 0.0581)	66 (0.8138, 0.0655)	66 (0.8068, 0.0524)	68 (0.8044, 0.0520)
	0.13	IM	$\boldsymbol{\delta}_1$	36 (0.8251, 0.0699)	37 (0.8132, 0.0709)	38 (0.8059, 0.0557)	39 (0.8079, 0.0531)
			$\boldsymbol{\delta}_2$	41 (0.8202, 0.0667)	42 (0.8190, 0.0685)	43 (0.8052, 0.0536)	44 (0.8026, 0.0523)
			$\boldsymbol{\delta}_3$	39 (0.8184, 0.0668)	40 (0.8116, 0.0746)	41 (0.8063, 0.0530)	42 (0.8047, 0.0572)
			$\boldsymbol{\delta}_4$	38 (0.8232, 0.0667)	39 (0.8095, 0.0684)	40 (0.8114, 0.0541)	41 (0.8109, 0.0547)
		MM	$\boldsymbol{\delta}_2$	42 (0.8145, 0.0692)	43 (0.8164, 0.0672)	44 (0.8117, 0.0528)	45 (0.8016, 0.0542)
			$\boldsymbol{\delta}_3$	40 (0.8157, 0.0696)	41 (0.8160, 0.0725)	42 (0.8160, 0.0509)	43 (0.8102, 0.0498)
			$\boldsymbol{\delta}_4$	38 (0.8147, 0.0672)	39 (0.8133, 0.0699)	40 (0.8128, 0.0519)	41 (0.8073, 0.0532)

Table 2.10. Required number of clusters (empirical power, empirical type I error) for cross-sectional SW-CRTs with count outcomes

	GEE		Adjusted GEE	
ζ	$\rho = 0.03$	$\rho = 0.05$	$\rho = 0.03$	$\rho = 0.05$
0.10	43 (0.8091, 0.0688)	55 (0.8149, 0.0598)	45 (0.8005, 0.0506)	57 (0.8013, 0.0483)
0.13	26 (0.8316, 0.0756)	32 (0.8209, 0.0741)	28 (0.8035, 0.0510)	34 (0.7984, 0.0534)

2.7. Discussion

In this study we present closed-form sample size formulas for closed-cohort and cross-sectional SW-CRTs with outcomes from the exponential family. The sample sizes are derived based on the GEE approach, which is flexible in accommodating arbitrary structures of longitudinal and between-subject correlations through the specification of $\mathbf{\Omega}$ and $\mathbf{\Phi}$. Moreover, through the specification of missing data pattern as well as marginal and joint observational probabilities, the proposed sample size method offers great flexibility in accounting for missing data. Due to the prolonged follow-up, the problem of missing data tends to be more pronounced in SW-CRTs than in traditional CRTs. By introducing $\boldsymbol{v}_s$ and p_s, the proposed sample sizes also accommodate arbitrary crossover schemes. For example, unbalanced randomization is accommodated by unequal p_s $(s = 1, \cdots, S)$, while setting $p_s = 0$ represents an irregular crossover scheme where no cluster switch to intervention at Step $s + 1$.

The closed-form formula for SW-CRTs with continuous outcomes shows that the time-specific intercepts do not affect the required number of clusters, which is consistent with the results in Li et al. (2018). However, it affects the required number of clusters in SW-CRTs with binary and count outcomes. Moreover, the closed-form formulas for SW-CRTs with continuous outcomes enable us to theoretically show that under $T = 3$ steps, the within-subject (longitudinal) correlation has no impact on sample size, and that under $T = 4$ steps, as long as the correlation between measurements at Time 2 and 3 are equal, different longitudinal correlation structures (such as CS and AR(1)) will lead to the same sample size requirement. Simulation results demonstrate that the proposed sample size method and the adjustment method perform well in maintaining the nominal power and type I error under a wide range of design configurations in SW-CRTs with different types of outcomes from the exponential family.

Chapter 3

A BAYESIAN ADAPTIVE DESIGN FOR SW-CRTS

3.1. Introduction

Adaptive designs have been increasingly considered for years in clinical research and development due to their flexibility and efficiency based on the accumulated data (Montaner et al., 1990; Rosenberger et al., 2001; Jennison and Turnbull, 2005). Specifically, they enable researchers to modify the design for determining the optimal intervention under investigation based on the review of interim data without destroying the validity and integrity of their intended studies (Chow et al., 2005). One of the popular adaptive designs is the group sequential design that attracted much attention in clinical trials. For example, Reiertsen et al. (1997) conducted a group sequential design to compare the diagnostic and therapeutic laparoscopy with conventional appedicectomy. Vahedi et al. (2007) did a group sequential trial for assessment of the efficacy of early decompressive craniectomy in patients with malignant middle cerebral artery infarction. The purpose of group sequential design is to shorten the length of clinical trials in a more efficient way without compromising the treatment safety and efficacy by maximizing the power for identifying the optimal intervention. Specifically, group sequential design employs stopping rules that will allow researchers to make a decision regarding whether to stop the trial early in case of overwhelming efficacy or futility based on the results of interim analysis. In other words, flexibility can be increased, as well as patient exposure, cost, and trial duration can be reduced if the intervention has been tested either exceedingly good or exceedingly poor.

Group sequential design with Bayesian approach is called Bayesian group sequential design, which has been investigated by many researchers (Heitjan, 1997; Zhou et al., 2008; Lee et al., 2010; Zhu and Yu, 2017). It can stop the trial early due to efficacy or futility as

that in frequentist group sequential designs, but with its own advantages: (1) it provides a natural way to make statistical inferences by combining information from the observed data and previous studies; (2) it obeys the likelihood principle without constrained by the design and doesn't require large sample theory for valid inference; (3) it gives interpretable results through a decision theoretical framework (Carlin and Louis, 2008; Gelman et al., 2013).

Stepped wedge cluster randomized trial (SW-CRT) is a relatively new trial that is frequently adopted to overcome the limitations of the traditional cluster randomized trials on practical considerations, such as large-scale research studies, availability of resources, ethical considerations, and cost-effectiveness. There are two popular frequentist statistical analysis methods for SW-CRTs, which are generalized linear mixed model and generalized estimating equations (Barker et al., 2016). Beside frequentist methods, recently the Bayesian approach has been implemented for SW-CRTs in several applications. For example, Reuther et al. (2014) used a Bayesian mixed effect model to evaluate the reduction of challenging behavior among two types of case conferences for people with dementia. Camacho et al. (2015) investigated the chance of demonstrating Ebola vaccine efficacy in the declining Ebola epidemic using a Bayesian analysis approach. However, there is few discussion regarding Bayesian design for SW-CRTs, particularly the Bayesian group sequential design. In this study, we propose to investigate the incorporation of the Bayesian group sequential approach into SW-CRTs through the predictive probability. The discussion of predictive probability for aiding a decision making process can be found in many works (Choi and Pepple, 1989; Choi et al., 1985; Gould, 2005; Spiegelhalter et al., 1986; Herson, 1979; Grieve et al., 1991; Johns and Andersen, 1999). Our proposed method continuously examines the results from updated data and determines whether the researchers can stop the study with a solid decision on efficacy/futility or should continue the study. A distinct advantage of this method is that it mimics the clinical decision-making process to make a rational decision. Based on interim data, the chance (predictive probability) that the trial will show a conclusive result at the end of study is evaluated. The decision to continue or stop the trial is made according to the strength of this predictive probability.

In this chapter, (1) we describe the proposed Bayesian predictive probability method and algorithms for making decisions in cross-sectional SW-CRTs; (2) we describe the criteria for determining the design parameters subject to the constraints of design properties. For example, the design parameters are searched within the given constraints such as type I error and power of the test can be guaranteed; (3) we conduct extensive simulation studies to examine the operating characteristics; (4) we illustrate the proposed method with a real example; and (5) we conclude our proposed method and discuss the practical issues.

3.2. A Bayesian predictive probability approach

We illustrate our proposed method by considering a cross-sectional SW-CRT, which can be easily extended to other types of SW-CRTs. Suppose there are S sequences. Within each sequence, there are $T = S + 1$ time points for a standard SW-CRT and n clusters at each time period ("step"). We assume the continuous measurements are collected from different cohort of subjects within each cluster at predefined time points. Let Y_{ijt} be the measurement obtained at time t $(t = 1, \ldots, T)$ from subject j $(j = 1, \ldots, J)$ of the ith cluster $(i = 1, \ldots, n)$. Thus JT is the cluster size and $N = nJT$ is the total number of subjects in this study. Let p_s be the proportion of clusters assigned to the sth sequence with $\sum_{s=1}^{S} p_s = 1$. A random effect model assumed for Y_{ijt} is

$$Y_{ijt} = \mu_{it} + e_{ijt}, \tag{3.1}$$

where $\mu_{it} = a_t\lambda + u_{it}\zeta + \alpha_i$ and

$$\alpha_i \sim N(0, \tau^2),$$

$$e_{ijt} \sim N(0, \sigma^2).$$

a_t is the indicator to specify the trend of time-specific intercept, for example, $a_t = t$ for linear trend and $a_t = t^2$ for quadratic trend. λ is the slope of time effect. u_{it} is the treatment indicator for subjects within the ith cluster at time t. ζ is the intervention effect. α_i is the random effect that describe the variation between clusters. e_{ijt} is the random error. It will

be convenient to proof that the variance of Y_{ijt} is $\sigma^2 + \tau^2$ and the intracluster correlation coefficient (ICC) is $\frac{\tau^2}{\sigma^2+\tau^2}$. Note that the observations are independent across clusters.

Suppose $\boldsymbol{Y}^{(t)} = (Y_{111}, \ldots, Y_{11t}, \ldots, Y_{nJ1}, \ldots, Y_{nJt})^{'}$ be the accumulated observations up to time t. We also define $\boldsymbol{Y}^{(-t)}$ such that

$$\boldsymbol{Y}^{(T)} = \boldsymbol{Y}^{(t)} \cup \boldsymbol{Y}^{(-t)}.$$

That is, $\boldsymbol{Y}^{(-t)}$ contains the collection of outcomes observed from time $(t+1)$ to T and $\boldsymbol{Y}^{(T)}$ is the full observation at the end of the study.

If we rewrite the model (3.1) using observed data at time t in matrix form, it will be

$$\boldsymbol{Y}^{(t)} = \boldsymbol{X_1}\boldsymbol{\beta_1} + \boldsymbol{X_2}\boldsymbol{\beta_2} + \boldsymbol{e}, \tag{3.2}$$

where

$$\boldsymbol{\beta_2} \sim N\left(\boldsymbol{0}, \tau^2 \boldsymbol{I}_n\right),$$

$$\boldsymbol{e} \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{nJt}\right),$$

$\boldsymbol{\beta_1} = (\lambda, \zeta)'$, $\boldsymbol{\beta_2} = (\alpha_1, \ldots, \alpha_n)'$. $\boldsymbol{X_1}$ is a $nJt \times 2$ matrix with each row (a_t, u_{it}) corresponding to the element in $\boldsymbol{Y}^{(t)}$. $\boldsymbol{X_2}$ is a $nJt \times n$ cluster-specific matrix with each row being a vector of length n with elements being 0 except the ith element being 1, for example, the first row of $\boldsymbol{X_2}$ is $(1, 0, \ldots, 0)'$ corresponding to Y_{111} and the last row of $\boldsymbol{X_2}$ is $(0, \ldots, 0, 1)'$ corresponding to Y_{nJt}. $\boldsymbol{e}$ is a vector of random error with elements corresponding to elements in $\boldsymbol{Y}^{(t)}$. $\boldsymbol{I}_c$ is a $c \times c$ identity matrix, where c is an integer.

Suppose we design a cross-sectional SW-CRT to test the hypothesis $H_1 : \zeta > \zeta_0$, where ζ_0 is the minimal target intervention effect. To apply the proposed approach to the SW-CRT at every predefined time point, the predictive probability of declaring the intervention effective at the end of the study given the current observations $\boldsymbol{Y}^{(t)}$ will be calculated to support the decision. Here we describe the steps for calculation of predictive probability as follows.

First, based on observed $\boldsymbol{Y}^{(t)}$, we update knowledge about the parameters under the Bayesian framework. Conceptually, prior distribution represents researchers' opinion about parameters before the study, likelihood represents learning from the study, and posterior distribution represents updated opinion about the parameters after the study. The likelihood

model is shown in Equation (3.2). To further develop our Bayesian model, the priors for the parameters of interest and nuisance parameters need to be specified. When no reliable information regarding the parameters at the design stage or the inference based only on observed data is desired, noninformative priors are preferred. Otherwise, the informative priors including information from previous studies are recommended (Zhang et al., 2007). To demonstrate our proposed method, we assume that observed data is highly informative about the parameters of interest. And we can afford to be vague about the priors with considering the noninformative independent prior distributions as suggested by Gelman (2006) and Gelman et al. (2013):

$$
\begin{aligned}
&p\left(\boldsymbol{\beta_1}\right) \propto 1; \\
&p\left(log(\sigma)\right) \propto 1 \text{ with } \sigma > 0, \text{which is equivalent to } p\left(\sigma^2\right) \propto \frac{1}{\sigma^2}; \\
&p(\tau) \propto 1 \text{ with } \tau > 0, \text{which is equivalent to } p(\tau^2) \propto \frac{1}{\tau}
\end{aligned}
$$

and can be interpreted as a limit of the folded-t family. Then we derive the joint posterior distribution based on the above specification, which is given by

$$
p\left(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2, \tau^2 | \boldsymbol{Y}^{(t)}\right) \propto p(\boldsymbol{Y}^{(t)} | \boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2) \times p(\boldsymbol{\beta_2} | \tau^2) \times p\left(\boldsymbol{\beta_1}, \sigma^2, \tau^2\right),
$$

where

$$
\begin{aligned}
p(\boldsymbol{Y}^{(t)} | \boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2) &= \frac{1}{\left(2\pi\sigma^2\right)^{\frac{nJt}{2}}} \exp\left\{ \frac{-1}{2\sigma^2} \left(\boldsymbol{Y}^{(t)} - \boldsymbol{X_1}\boldsymbol{\beta_1} - \boldsymbol{X_2}\boldsymbol{\beta_2}\right)' \left(\boldsymbol{Y}^{(t)} - \boldsymbol{X_1}\boldsymbol{\beta_1} - \boldsymbol{X_2}\boldsymbol{\beta_2}\right) \right\}, \\
p(\boldsymbol{\beta_2} | \tau^2) &= \frac{1}{\left(2\pi\tau^2\right)^{n/2}} \exp\left\{ -\frac{1}{2\tau^2} \boldsymbol{\beta_2}' \boldsymbol{\beta_2} \right\}.
\end{aligned}
$$

Because the closed-form marginal posterior distributions don't exist, one of the most popular Markov Chain Monte Carlo (MCMC) methods, Gibbs sampler, can be used here to obtain posterior samples (Geman and Geman, 1987). All the full conditionals can be derived from

the joint posterior distribution as follows:

$$p(\boldsymbol{\beta_1}|\boldsymbol{\beta_2},\sigma^2,\tau^2,\boldsymbol{Y}^{(t)}) \sim N\left(\left(\boldsymbol{X_1'X_1}\right)^{-1}\boldsymbol{X_1'}\left(\boldsymbol{Y^{(t)}}-\boldsymbol{X_2\beta_2}\right),\sigma^2\left(\boldsymbol{X_1'X_1}\right)^{-1}\right),$$

$$p(\boldsymbol{\beta_2}|\boldsymbol{\beta_1},\sigma^2,\tau^2,\boldsymbol{Y}^{(t)}) \sim N\left(\left(\frac{\boldsymbol{X_2'X_2}}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\frac{\boldsymbol{X_2'}\left(\boldsymbol{Y^{(t)}}-\boldsymbol{X_1\beta_1}\right)}{\sigma^2},\left(\frac{\boldsymbol{X_2'X_2}}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\right),$$

$$p(\sigma^2|\boldsymbol{\beta_1},\boldsymbol{\beta_2},\boldsymbol{Y}^{(t)}) \sim scaledInv\chi^2\left(nJt,\frac{\left(\boldsymbol{Y}^{(t)}-\boldsymbol{X_1\beta_1}-\boldsymbol{X_2\beta_2}\right)'\left(\boldsymbol{Y}^{(t)}-\boldsymbol{X_1\beta_1}-\boldsymbol{X_2\beta_2}\right)}{nJt}\right),$$

$$p(\tau^2|\boldsymbol{\beta_1},\boldsymbol{\beta_2},\sigma^2,\boldsymbol{Y}^{(t)}) \sim scaledInv\chi^2\left(n-1,\frac{1}{n-1}\boldsymbol{\beta_2'\beta_2}\right).$$

Second, we use updated knowledge of parameters to predict future observations $\boldsymbol{Y}^{(-t)}$. The posterior predictive distribution of $\boldsymbol{Y}^{(-t)}$ is given by

$$p(\boldsymbol{Y}^{(-t)}|\boldsymbol{\beta_1},\boldsymbol{\beta_2},\sigma^2,\tau^2,\boldsymbol{Y}^{(t)}) \sim N\left(\boldsymbol{X_1^{(-t)}\beta_1}+\boldsymbol{M_{21}M_{11}^{-1}}\left(\boldsymbol{Y}^{(t)}-\boldsymbol{X_1\beta_1}\right),\boldsymbol{M_{22}}-\boldsymbol{M_{21}M_{11}^{-1}M_{12}}\right), \tag{3.3}$$

where $\boldsymbol{Y}^{(-t)}=\left(Y_{11(t+1)},\ldots,Y_{11T},\ldots,Y_{nJ(t+1)},\ldots,Y_{nJT}\right)'$ is a vector of predicted future observations. $\boldsymbol{M_{11}}$ and $\boldsymbol{M_{22}}$ are $nJt\times nJt$ and $nJ(T-t)\times nJ(T-t)$ matrices with diagonal elements being $\sigma^2+\tau^2$ and off-diagonal elements being τ^2, respectively. $\boldsymbol{M_{12}}$ is a $nJt\times nJ(T-t)$ matrix with all elements being τ^2, $\boldsymbol{M_{21}}$ is the transpose of $\boldsymbol{M_{12}}$.

Third, we have a "full" set of observations by combining observed $\boldsymbol{Y}^{(t)}$ with predicted $\boldsymbol{Y}^{(-t)}$, based on which we determine whether to declare the intervention effective. Specifically, given the "full" observations $(\boldsymbol{Y}^{(t)},\boldsymbol{Y}^{(-t)})$, conclusions are made based on the posterior probability of ζ exceeding the prespecified level ζ_0, such as

$$P\left(\zeta>\zeta_0|\boldsymbol{Y}^{(t)},\boldsymbol{Y}^{(-t)}\right)=\int p\left(\zeta|\boldsymbol{Y}^{(t)},\boldsymbol{Y}^{(-t)}\right)I(\zeta>\zeta_0)d\zeta.$$

Here $I(\cdot)$ is an indicator function and $p\left(\zeta|\boldsymbol{Y}^{(t)},\boldsymbol{Y}^{(-t)}\right)$ is the marginal posterior distribution of ζ. We compare $P\left(\zeta>\zeta_0|\boldsymbol{Y}^{(t)},\boldsymbol{Y}^{(-t)}\right)$ with a threshold θ_U to consider that whether the intervention is efficacious at the end of the study. θ_U plays a pivotal role in screening out the inefficacious intervention based on "complete" data at the end of the study.

Fourth, we repeat previous three steps to evaluate the predictive probability of the trial being success at the end of study. Let PP denote the proportion of times that intervention

is declared effective, indicating the chance of getting a statistically significant result at the end of the study given current evidence. The PP at time t is

$$PP_t = \int I\left(P\left(\zeta > \zeta_0 | \boldsymbol{Y}^{(t)}, \boldsymbol{Y}^{(-t)}\right) > \theta_U\right) p\left(\boldsymbol{Y}^{(-t)} | \boldsymbol{Y}^{(t)}\right) d\boldsymbol{Y}^{(-t)}$$

In other words, PP_t is the weighted average of positive trials with respect to the current belief about the unknown parameters.

Fifth, we can use PP_t to determine whether the trial should be stopped early due to overwhelming evidence of efficacy/futility or continued because the current data are not yet conclusive, based on the following decision rules:

- if $PP_t < \pi_L$, stop the trial and conclude the intervention ineffective;
- if $PP_t > \pi_U$, stop the trial and conclude the intervention effective;
- otherwise continue to time $(t+1)$ until reaching the end of study.

Predictive probabilities formalize the decision making process by comparing them with thresholds. In practice, we choose θ_U, π_U, and π_L as positive numbers between 0 and 1. Then the decision for termination of the trial for efficacy or futility can be made if the trial currently show superiority or inferiority and are likely to maintain it after remaining data are collected. If $\pi_U = 1$, the trial will stop for futility, but not for efficacy. Moreover, a lower θ_U indicates that the null hypothesis will be rejected easily with increased power and type I error. In contrast, a higher θ_U indicates that the null hypothesis will be rejected less likely with decreased power and type I error. π_U and π_L are related to early stopping due to efficacy and futility, respectively. A higher π_L leads to easier stopping for lack of efficacy and then causes decreased power and type I error. Similarly, a lower π_U leads to easier stopping for efficacy and then causes increased power and type I error.

3.3. Simulation studies

To illustrate our proposed predictive probability method, we conducted extensive simulation studies. We assumed a cross-sectional SW-CRT with $T = 5$ time points (correspondingly

$S = 4$ for a standard SW-CRT) and n clusters with $J = 10$ at each time point. We set intervention effect $\zeta = 0.4$ and slope of time effect $\lambda = 0.1$ with a linear trend indicator $a_t = t$ for $t = 1, \ldots, T$. We also assumed an even transition scheme with $p_s = \frac{1}{S}$ for $s = 1, \ldots, S$. Let $\sigma^2 = 1$ for random error and $\rho = 0.05$ for ICC. The hypothesis test we are interested in is $H_1 : \zeta > 0.1$.

Here the detailed algorithm to conduct a Bayesian group sequential design based on the predictive probability is described. In practice, it has been recommended to start the Bayesian adaptive scheme after enough data have been collected to avoid premature decisions based on spurious results. Without loss of generality, we assume that the adaptation start from time $t_0 = 3$. The trial proceeds as follows.

1. At Step t ($t = t_0, \cdots, T-1$), collect observation $\boldsymbol{Y}^{(t)}$ to prepare for adaptation.

   (a) Simulate $\boldsymbol{Y}^{(-t)}$ from posterior predictive distribution shown in (3.3) using drawed parameter samplers.

   (b) Construct a "full" set $\boldsymbol{Y}^* = \left(\boldsymbol{Y}^{(t)}, \boldsymbol{Y}^{(-t)}\right)$. Conduct MCMC simulation to sample from posterior distribution $p(\zeta|\boldsymbol{Y}^*)$. Let $\zeta^{(l)}$ ($l = 1, \cdots, L = 1000$) be the posterior samples of intervention effect from L iterations of simulation.

   (c) Numerically evaluate $P\left(\zeta > \zeta_0|\boldsymbol{Y}^*\right) \approx \frac{1}{L}\sum_{l=1}^{L} I\left(\zeta^{(l)} > \zeta_0\right)$.

   (d) If $P\left(\zeta > \zeta_0|\boldsymbol{Y}^*\right) > \theta_U$, declare the intervention effective and set decision indicator $D = 1$, otherwise $D = 0$.

   (e) Repeat (a)-(d) M times. Let D_m ($m = 1, \cdots, M = 1000$) be the mth decision indicator. The predictive probability of the trial being success at the end of study given $\boldsymbol{Y}^{(t)}$ is numerically evaluated by

   $$PP_t \approx \frac{1}{M}\sum_{m=1}^{M} D_m.$$

   (f) Make adaptive decision to terminate or continue the trial at Step t according to decision rules:

      i. if $PP_t < \pi_L$, stop the trial early and conclude the intervention ineffective;

ii. if $PP_t > \pi_U$, stop the trial early and conclude the intervention effective;

iii. otherwise continue to Step $(t+1)$.

2. At Step T, collect the full observation $\boldsymbol{Y}^{(T)}$.

   (a) Conduct MCMC simulation to sample from posterior distribution $p\left(\zeta|\boldsymbol{Y}^{(T)}\right)$. Let $\zeta^{(l)}$ $(l = 1, \cdots, L = 1000)$ be the posterior sample from lth iteration of simulation.

   (b) Numerically evaluate probability $P\left(\zeta > \zeta_0|\boldsymbol{Y}^{(T)}\right)$ by

$$P\left(\zeta > \zeta_0|\boldsymbol{Y}^{(T)}\right) \approx \frac{1}{L}\sum_{l=1}^{L} I\left(\zeta^{(l)} > \zeta_0\right).$$

   (c) Stop trial. Declare the intervention effective if $P\left(\zeta > \zeta_0|\boldsymbol{Y}^{(T)}\right) > \theta_U$, declare the intervention ineffective if $P\left(\zeta > \zeta_0|\boldsymbol{Y}^{(T)}\right) < \theta_L$, otherwise declare the trial inconclusive.

To design a trial using proposed predictive probability method, we should search the optimal design parameters $(n, \theta_U, \theta_L, \pi_U, \pi_L)$ to attain desirable design properties. Specifically, the search can be implemented as follows.

1. Determine the range of n using frequentist methods.

2. Determine the range of thresholds $(\theta_U, \theta_L, \pi_U, \pi_L)$.

3. For each combination of n and thresholds, summarize the empirical power, empirical type I error, probability of early stopping, and expected number of subjects. The expected number of subjects $(E(N))$ is the average of required numbers of subjects from all simulations and can be calculated by nJT', where T' is the average of steps needed in the simulations.

4. Search all combinations to identify the optimal design with desirable design characteristics. The optimization criteria can be, but not limited to

   - subject to desired power and type I error, find the optimal design with minimal number of clusters (n) and expected number of subjects $(E(N))$;

- subject to fixed number of clusters (n) and desired type I error, find the optimal design with maximal power.

To identify the optimal design, we set $n = (15, 20, 25)$ because frequentist methods suggest that $n = 19$ (Li et al., 2018) and $n = 22$ (Hooper et al., 2016). We set $\pi_U = 1$ because we don't want to stop the trial if the intervention shows early signs of efficacy. We also set $\theta_L = (0.05, 0.10, 0.15, 0.20)$ even it only matters when trial continues to the last step. We chose the following values for other thresholds: θ_U from 0.80 to 1.00 by 0.1 intervals and π_L from 0.05 to 0.70 by 0.05 intervals, as discussed by Zhou et al. (2008).

By searching over all combinations, we can identify the optimal design parameters to achieve desirable design properties. Tables 3.1-3.3 show the empirical powers, empirical type I errors, and expected numbers of subjects using proposed predictive probability method when $\theta_L = 0.10$. Actually, the effect of θ_L is very small because it only affects the study when the study doesn't stop early and continues to the last step. The results for $\theta_L = (0.05, 0.15, 0.20)$ are very similar to results in Tables 3.1-3.3 and not shown here. Suppose we desire a design with power at least 0.8 and type I error at most 0.05, the optimal choice is $(n = 20, \theta_U = 0.87, \pi_L = 0.70)$ in Table 3.2 with smallest expected number of subjects being 925. The empirical power and type I error are 0.800 and 0.038, respectively. Using frequentist method with above parameter settings and type I error of 0.038, the calculated power is 0.796 (Li et al., 2018). Therefore, the proposed Bayesian method is comparable to the frequentist method. On the other hand, suppose we desire a design with fixed number of clusters $n = 15$ and type I error at most 0.05, the optimal choice subject to maximal power is $(\theta_U = 0.96, \pi_L = 0.05)$ in Table 3.1 with empirical power and type I error being 0.728 and 0.043, respectively.

Table 3.1. Empirical power/type I error/expected number of subjects using predictive probability method with $n = 15$

θ_U	$\pi_L = 0.05$	$\pi_L = 0.20$	$\pi_L = 0.40$	$\pi_L = 0.70$
0.80	0.927/0.204/744	0.905/0.185/732	0.853/0.145/714	0.758/0.094/682
0.81	0.925/0.194/744	0.903/0.175/732	0.847/0.137/713	0.747/0.087/679
0.82	0.921/0.188/743	0.896/0.170/730	0.837/0.133/710	0.740/0.082/677
0.83	0.914/0.177/743	0.890/0.159/729	0.825/0.125/707	0.726/0.077/673
0.84	0.904/0.165/742	0.881/0.148/729	0.819/0.117/706	0.716/0.073/671
0.85	0.900/0.153/742	0.874/0.136/726	0.811/0.109/704	0.702/0.068/668
0.86	0.896/0.145/740	0.866/0.132/724	0.804/0.100/701	0.696/0.065/666
0.87	0.884/0.138/740	0.855/0.124/723	0.795/0.095/699	0.679/0.060/661
0.88	0.875/0.128/739	0.846/0.115/720	0.786/0.088/697	0.665/0.056/657
0.89	0.869/0.114/737	0.835/0.101/718	0.775/0.077/694	0.649/0.050/653
0.90	0.860/0.103/737	0.824/0.090/715	0.763/0.068/691	0.640/0.044/650
0.91	0.847/0.094/735	0.811/0.082/711	0.748/0.061/688	0.630/0.037/646
0.92	0.836/0.085/733	0.801/0.073/710	0.732/0.055/683	0.615/0.029/642
0.93	0.825/0.072/730	0.784/0.063/706	0.724/0.047/680	0.591/0.025/635
0.94	0.793/0.063/729	0.743/0.057/698	0.685/0.042/673	0.567/0.022/630
0.95	0.765/0.053/724	0.721/0.050/695	0.662/0.036/669	0.536/0.019/621
0.96	0.728/0.043/720	0.689/0.037/688	0.628/0.026/661	0.505/0.014/613
0.97	0.689/0.030/714	0.650/0.026/681	0.591/0.022/651	0.464/0.009/600
0.98	0.629/0.018/704	0.582/0.015/669	0.529/0.013/639	0.394/0.008/581
0.99	0.525/0.011/690	0.491/0.011/652	0.439/0.010/617	0.300/0.005/557

Table 3.2. Empirical power/type I error/expected number of subjects using predictive probability method with $n = 20$

θ_U	$\pi_L = 0.05$	$\pi_L = 0.20$	$\pi_L = 0.40$	$\pi_L = 0.70$
0.80	0.959/0.174/995	0.945/0.152/985	0.923/0.119/975	0.850/0.062/943
0.81	0.956/0.164/995	0.939/0.142/984	0.920/0.114/974	0.846/0.058/942
0.82	0.954/0.154/994	0.937/0.130/983	0.917/0.103/973	0.841/0.055/940
0.83	0.951/0.143/993	0.935/0.120/983	0.913/0.094/972	0.836/0.050/938
0.84	0.948/0.134/993	0.932/0.116/982	0.905/0.086/969	0.829/0.048/936
0.85	0.944/0.127/993	0.925/0.110/981	0.897/0.076/967	0.821/0.047/933
0.86	0.942/0.119/993	0.921/0.103/980	0.894/0.070/966	0.808/0.046/928
0.87	0.940/0.112/992	0.918/0.099/980	0.891/0.064/965	0.800/0.038/925
0.88	0.930/0.101/991	0.914/0.091/979	0.887/0.058/964	0.790/0.035/922
0.89	0.928/0.092/990	0.910/0.080/978	0.882/0.052/963	0.777/0.030/916
0.90	0.922/0.080/990	0.901/0.068/975	0.869/0.045/958	0.766/0.028/911
0.91	0.916/0.075/988	0.895/0.060/973	0.856/0.042/954	0.750/0.026/905
0.92	0.899/0.067/986	0.881/0.053/971	0.836/0.042/950	0.721/0.024/895
0.93	0.891/0.061/985	0.874/0.051/970	0.825/0.038/945	0.701/0.020/889
0.94	0.875/0.053/984	0.855/0.043/966	0.804/0.033/941	0.666/0.018/877
0.95	0.854/0.046/981	0.827/0.037/962	0.777/0.025/935	0.631/0.015/864
0.96	0.832/0.040/978	0.808/0.029/957	0.757/0.021/928	0.606/0.012/853
0.97	0.806/0.025/974	0.776/0.019/951	0.720/0.013/914	0.568/0.010/840
0.98	0.761/0.017/969	0.725/0.013/936	0.655/0.010/893	0.499/0.005/816
0.99	0.649/0.008/957	0.619/0.008/913	0.542/0.006/857	0.406/0.003/780

Table 3.3. Empirical power/type I error/expected number of subjects using predictive probability method with $n = 25$

θ_U	$\pi_L = 0.05$	$\pi_L = 0.20$	$\pi_L = 0.40$	$\pi_L = 0.70$
0.80	0.993/0.214/1249	0.988/0.193/1246	0.968/0.158/1235	0.912/0.105/1208
0.81	0.989/0.202/1249	0.985/0.182/1246	0.966/0.153/1234	0.906/0.097/1204
0.82	0.985/0.185/1249	0.981/0.165/1246	0.963/0.140/1233	0.899/0.091/1201
0.83	0.985/0.178/1248	0.977/0.159/1244	0.962/0.136/1232	0.896/0.087/1199
0.84	0.983/0.171/1248	0.975/0.153/1244	0.957/0.128/1230	0.888/0.080/1196
0.85	0.982/0.162/1248	0.974/0.145/1243	0.954/0.125/1228	0.884/0.073/1194
0.86	0.975/0.152/1247	0.968/0.136/1241	0.948/0.116/1226	0.877/0.072/1191
0.87	0.974/0.140/1247	0.966/0.126/1241	0.942/0.105/1225	0.869/0.065/1188
0.88	0.973/0.132/1247	0.965/0.119/1240	0.937/0.099/1222	0.864/0.060/1186
0.89	0.966/0.121/1246	0.958/0.108/1239	0.933/0.093/1222	0.856/0.057/1182
0.90	0.966/0.111/1246	0.958/0.101/1237	0.928/0.087/1219	0.843/0.051/1175
0.91	0.958/0.098/1246	0.949/0.090/1235	0.921/0.076/1217	0.835/0.048/1171
0.92	0.952/0.091/1245	0.942/0.085/1232	0.911/0.071/1213	0.825/0.045/1167
0.93	0.946/0.083/1245	0.936/0.076/1231	0.898/0.064/1209	0.814/0.036/1162
0.94	0.937/0.071/1244	0.922/0.066/1226	0.883/0.051/1202	0.780/0.027/1147
0.95	0.926/0.066/1244	0.906/0.058/1222	0.871/0.045/1196	0.763/0.026/1139
0.96	0.906/0.049/1239	0.886/0.044/1217	0.847/0.036/1188	0.726/0.020/1123
0.97	0.895/0.035/1237	0.868/0.031/1210	0.826/0.025/1179	0.693/0.014/1108
0.98	0.851/0.024/1229	0.825/0.021/1198	0.781/0.017/1165	0.632/0.010/1080
0.99	0.790/0.014/1218	0.763/0.012/1177	0.704/0.010/1134	0.546/0.004/1042

There are several observations about the effect of n, θ_U, and π_L on empirical power and expected number of subjects, as shown in Figure 3.1. First, a lower θ_U leads to a higher empirical power (red). This is reasonable because a lower θ_U indicates less restrictive limits to make the decision for efficacious intervention and then leads to a higher power. Second, θ_U has a greater impact on empirical power and expected number of clusters than π_L. For example, the empirical powers are 0.927 and 0.525 when θ_U is 0.80 and 0.99 with fixed $\pi_L = 0.05$, respectively. On the other hand, the empirical powers are 0.927 and 0.758 when π_L is 0.05 and 0.70 with fixed $\theta_U = 0.80$. Third, the empirical power increases with larger number of clusters (n), which is consistent with the frequentist conclusion. Fourth, as the number of clusters increases, the expected number of subjects increases noticeably because it is a increasing function of n, even T' varies but within a small range between 3 and 5 in this simulation study and T' is close to T because we set $\pi_U = 1$ for not stopping early when intervention is efficacious. Fifth, when $\theta_U = 1$, the power is 0 based on the formula of PP_t and the trial stops at step t_0 with expected number of subjects being nJt_0.

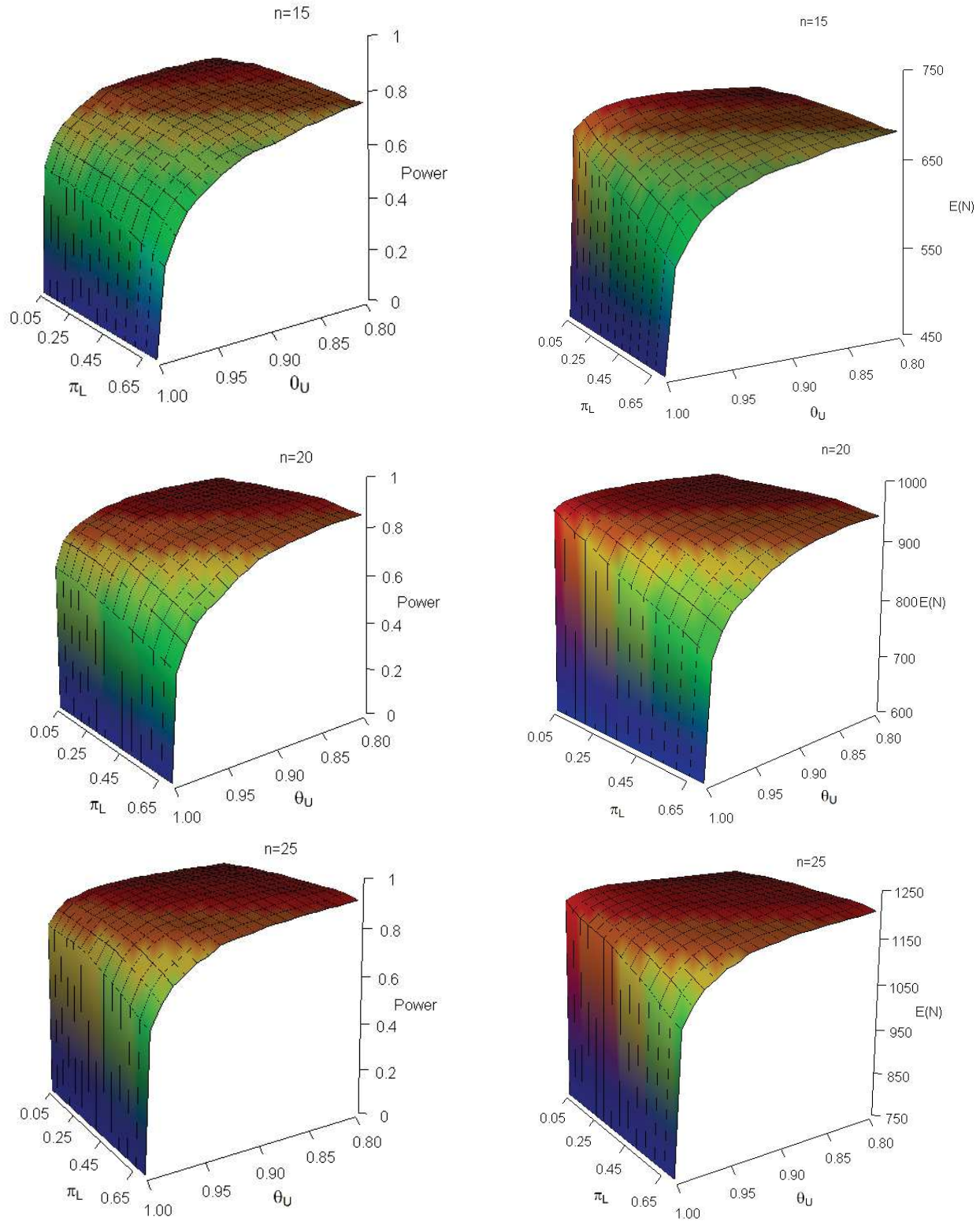

Figure 3.1. Empirical power (left) and expected number of subjects (right) using predictive probability method with different combination of design parameters

Figure 3.2 shows the trend of empirical power, probability of early stopping, and expected number of subjects with increasing intervention effect using predictive probability method with fixed $\theta_U = 0.96$ and $\pi_L = 0.05$. In the first panel, the gray dashed line is corresponding to 0.05. The points at intervention effect 0.1 are the empirical type I errors for the trials with different numbers of clusters (n). As intervention effect increases, the trial with larger n attains a higher power. The second panel shows the probability of early stopping, which decreases with the increased intervention effect. When intervention effect is small, say 0.1, the intervention is inefficacious and the trial tends to have smaller PP_t and easier stopping with high probability. When intervention effect is large, say 0.4, the intervention is efficacious and the trial is more likely to continue to the end of the study. Moreover, the probability of early stopping for the trial with a higher power is smaller. In the third panel, a smaller probability of early stopping leads to a larger expected number of subjects.

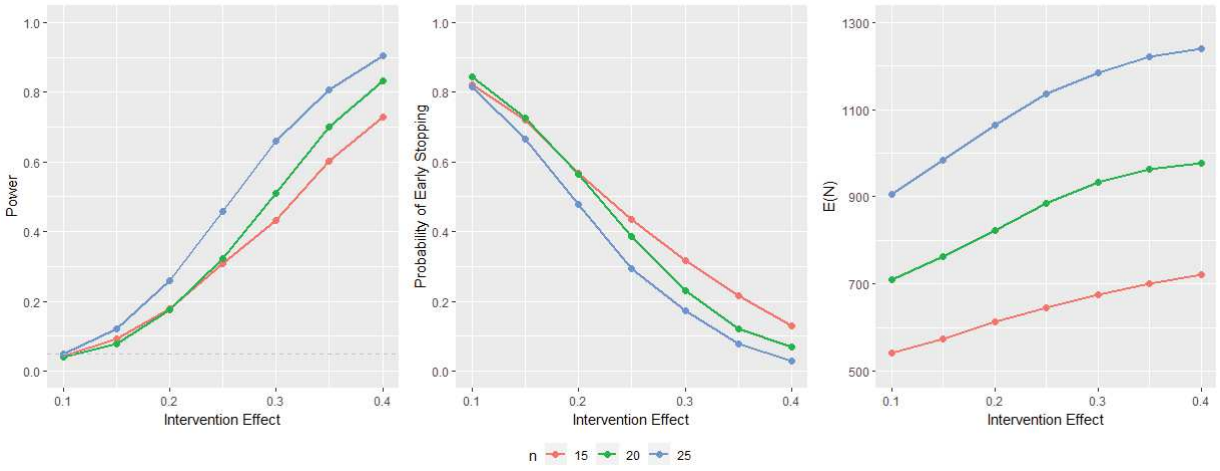

Figure 3.2. Empirical power, probability of early stopping, and expected number of subjects using predictive probability method with different intervention effects

3.4. A real application example

We illustrate our proposed method using a pragmatic cross-sectional SW-CRT, which is established to evaluate the effect of combining interventions in preventing malnutrition and reducing weight loss in hospitalized patients with acute tertiary care (Kitson et al., 2013). The combining interventions incorporate nutrition screening, nutritional supplements, and feeding assistance. The primary outcome is the weekly rate of change in patient's body weight. Suppose the wards are randomized into $S = 4$ sequences with $J = 20$ patients at each of $T = 5$ time periods. The measurements will be collected at baseline and four follow-ups. We assume the standardized intervention effect $\zeta = 0.3$ and $ICC = 0.03$ based on the Schultz et al. (2014). To choose the best design for this study by controlling power and type I error, we set the required number of clusters $n = (10, 15, 20)$ based on frequentist method. Moreover, we set $\pi_U = 1$ to avoid early stopping for efficacy and $\theta_L = 0.1$. For each n, we search the θ_U and π_L space to find the optimal design with the desirable properties. The results with $n = 10$ are shown in Table 3.4. Within each cell, there are empirical power, empirical type I error, and expected number of subjects. If a trial with power ≥ 0.9 and type I error ≤ 0.1 is desired, $(\theta_U, \pi_L) = (0.89, 0.20)$ will be chosen with regards to the smallest number of subjects 978. The corresponding empirical power, empirical type I error, and probability of early stopping are 0.911, 0.092, 6.27%, respectively. If a trial with power $\geq$ 0.8 and type I error ≤ 0.05 is desired, $(\theta_U, \pi_L) = (0.93, 0.40)$ is the best choice with the smallest number of subjects 937. The corresponding empirical power, empirical type I error, and probability of early stopping are 0.804, 0.047, and 17.4%, respectively.

Table 3.4. Empirical power/type I error/expected number of subjects for a SW-CRT with $n = 10$ and $\zeta = 0.3$

θ_U	$\pi_L = 0.05$	$\pi_L = 0.20$	$\pi_L = 0.40$	$\pi_L = 0.70$
0.8	0.971/0.195/997	0.951/0.177/988	0.921/0.147/972	0.829/0.088/935
0.81	0.971/0.184/997	0.949/0.170/987	0.918/0.135/970	0.822/0.082/932
0.82	0.969/0.177/997	0.947/0.161/986	0.912/0.127/968	0.815/0.075/929
0.83	0.968/0.165/997	0.943/0.146/985	0.908/0.119/967	0.806/0.072/925
0.84	0.964/0.153/997	0.940/0.134/984	0.901/0.110/965	0.801/0.067/923
0.85	0.962/0.140/996	0.937/0.125/983	0.891/0.101/961	0.791/0.062/919
0.86	0.957/0.132/996	0.931/0.119/982	0.884/0.096/959	0.778/0.059/915
0.87	0.954/0.121/996	0.923/0.106/979	0.875/0.083/956	0.769/0.052/911
0.88	0.943/0.112/995	0.916/0.097/979	0.869/0.077/954	0.758/0.049/907
0.89	0.936/0.105/994	0.911/0.092/978	0.861/0.073/952	0.747/0.046/904
0.90	0.924/0.094/993	0.897/0.080/974	0.847/0.066/949	0.731/0.045/900
0.91	0.913/0.085/991	0.887/0.072/971	0.838/0.061/946	0.719/0.041/894
0.92	0.899/0.080/989	0.873/0.070/967	0.817/0.057/940	0.699/0.035/887
0.93	0.887/0.073/987	0.858/0.059/963	0.804/0.047/937	0.674/0.030/878
0.94	0.867/0.063/986	0.835/0.052/960	0.787/0.040/930	0.648/0.025/869
0.95	0.840/0.050/982	0.814/0.039/956	0.762/0.033/921	0.625/0.021/860
0.96	0.820/0.040/978	0.790/0.033/949	0.729/0.025/912	0.593/0.015/850
0.97	0.782/0.030/972	0.753/0.025/940	0.689/0.022/899	0.544/0.009/832
0.98	0.725/0.018/963	0.692/0.016/925	0.626/0.013/881	0.492/0.008/813
0.99	0.639/0.011/949	0.607/0.010/896	0.542/0.009/854	0.405/0.007/779

In simulation studies and example, we set $\pi_U = 1$ to avoid early stopping due to efficacy. However, π_U can be any number between 0 and 1 theoretically. Here we investigate the effect of π_U on empirical power, probability of early stopping, and expected number of subjects. Suppose early stopping due to efficacy is allowed and (θ_U, π_L) are fixed at $(0.93, 0.40)$, which lead to the optimal design in the example with power ≥ 0.8 and type I error ≤ 0.05. We set π_U from 0.80 to 1 by 0.02 intervals. The result in Figure 3.3 shows that empirical power does not change much at different values of π_U. However, the probability of early stopping and expected number of subjects change noticeably at different values of π_U. Specifically, the probability of early stopping is decreasing with increased π_U because a higher π_U indicates a stricter condition for early stopping for efficacy. Correspondingly, the expected number of subjects is increasing.

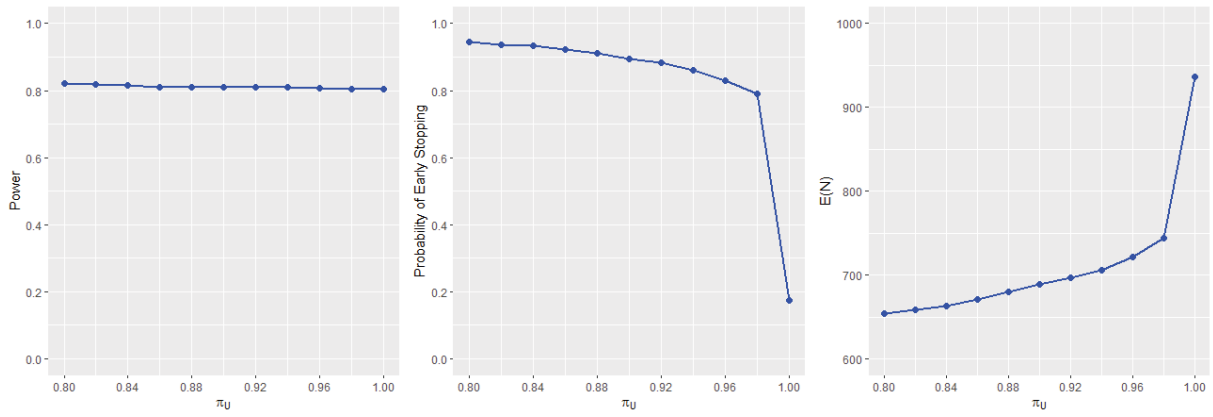

Figure 3.3. Empirical power, probability of early stopping, and expected number of subjects with varying π_U

3.5. Discussion

In this study we proposed a predictive probability method to incorporate Bayesian adaptive approach to SW-CRTs, unlike frequentist design, which is a novel and ethical method with flexibility of early stopping and efficiency of identification of effective or elimination of ineffective treatment. This predictive probability method allows researchers to make conclu-

sive result at the end of the study based on the current collected data. The proposed method requires the specification of the number of clusters and thresholds $(n, \theta_U, \theta_L, \pi_U, \pi_L)$. The optimal choice should balance the practical considerations and trial efficiency. The design parameters can be determined through a numerical search algorithm to yield desired frequentist operating characteristics such as power, type I error, and expected number of subjects. We conducted simulation studies to examining the performance of proposed predictive probability approach in cross-sectional SW-CRTs. The results show that the proposed method achieves good operating characteristics. When evidence accumulates that the treatment is effective or ineffective, it is wise to have a rule to stop the trial early.

Bayesian predictive probability framework allows existing knowledge about the parameters from previous studies or literatures to be formally incorporated into design through prior specification, as well as updates information of parameters of interest using accumulated data observed over time. For prior selection, if the prior information is reliable, we can combine it with current study through informative priors. If the prior informative is not reliable, we can apply noninformative priors, even it leads to similar or same estimates as maximum likelihood method, it provides a natural way for making inference based on likelihood and without the restriction of large sample theory and predefined design settings (Zhou et al., 2008). In this study, we employed noninformative priors and limit inference to the trial itself with no broader scope intended. However, it should not be considered as the only alternative when prior knowledge is not available. Other priors leading to proper posterior distributions can also worthy of consideration.

In this study, the financial constraints are not taken into consideration. Actually, expected cost for recruiting a subject and collecting one measurement at different time can be employed to determine an optimal Bayesian design under practical considerations. For example, we can calculate the expected cost for each combination of parameter settings and then determine which design is the best through comparing costs. The cost function might be set as $C(T') = nJT'$, where we assume fixed cluster size (J) and equal weight for subjects from different clusters. This is equivalent to the expected number of clusters in our study

($E(N) = nJT'$). If different weight and varying cluster size are considered, the cost function could be rewritten as $C(T') = n\sum_{i=1}^{n} J_i w_i T'$, where w_i is the weight for subjects from the ith clusters with $\sum_{i=1}^{n} w_i = 1$ and J_i is the cluster size for the ith cluster. If the relationship between cost and stopping step T' is not linear, say quadratic, the cost function could be expressed as $C(T') = n\sum_{i=1}^{n} J_i w_i T'^2$.

Appendix A

Derivation of equations

A.1. Derivation of Equation (2.3)

First we have

$$\begin{aligned}\boldsymbol{A}_n &= n^{-1}J\sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{X}_i \\ &= n^{-1}J\sum_{i=1}^{n}\begin{bmatrix} \boldsymbol{I_T} & \boldsymbol{u_i} \\ \boldsymbol{u_i'} & \boldsymbol{u_i'u_i} \end{bmatrix}.\end{aligned}$$

As $n \to \infty$, $\boldsymbol{A}_n$ approaches

$$\begin{aligned}\boldsymbol{A} &= J\sum_{s=1}^{S}p_s\begin{bmatrix} \boldsymbol{I_T} & \boldsymbol{v_s} \\ \boldsymbol{v_s'} & \boldsymbol{v_s'v_s} \end{bmatrix} \\ &= J\sum_{s=1}^{S}p_s\boldsymbol{W_s'W_s}.\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\boldsymbol{E}_n &= n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\right)^{\otimes 2} \\ &= n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\right)\left(\sum_{j=1}^{J}\hat{\boldsymbol{e}}_{ij}'\boldsymbol{X}_i\right) \\ &= n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\sum_{j'=1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\hat{\boldsymbol{e}}_{ij'}'\boldsymbol{X}_i\right) \\ &= n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\hat{\boldsymbol{e}}_{ij}'\boldsymbol{X}_i + 2\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\hat{\boldsymbol{e}}_{ij'}'\boldsymbol{X}_i\right).\end{aligned}$$

As $n \to \infty$, $\boldsymbol{E}_n$ approaches

$$\boldsymbol{E} = \sigma^2 J \sum_{s=1}^{S} p_s \boldsymbol{W}_{\boldsymbol{s}}^{'} \left(\boldsymbol{\Omega} + (J-1)\, \boldsymbol{\Phi} \right) \boldsymbol{W}_{\boldsymbol{s}}.$$

We are only interested in σ_ζ^2, which is the $(T+1, T+1)$-component of $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1} \boldsymbol{E} \boldsymbol{A}^{-1}$. The last row of $\boldsymbol{A}^{-1}$ can be simplified as

$$\left[J \sum_{t=1}^{T} \bar{u}_t \left(1 - \bar{u}_t\right) \right]^{-1} \begin{bmatrix} -\bar{\boldsymbol{u}}' & 1 \end{bmatrix}.$$

Then we have

$$\begin{aligned}
\sigma_\zeta^2 &= \left[J \sum_{t=1}^{T} \bar{u}_t \left(1 - \bar{u}_t\right) \right]^{-2} \begin{bmatrix} -\bar{\boldsymbol{u}}' & 1 \end{bmatrix} \boldsymbol{E} \begin{bmatrix} -\bar{\boldsymbol{u}}' & 1 \end{bmatrix}^{'} \\
&= \left[J \sum_{t=1}^{T} \bar{u}_t \left(1 - \bar{u}_t\right) \right]^{-2} \begin{bmatrix} -\bar{\boldsymbol{u}}' & 1 \end{bmatrix} \left(\sigma^2 J \sum_{s=1}^{S} p_s \boldsymbol{W}_{\boldsymbol{s}}^{'} \left(\boldsymbol{\Omega} + (J-1)\, \boldsymbol{\Phi} \right) \boldsymbol{W}_{\boldsymbol{s}} \right) \begin{bmatrix} -\bar{\boldsymbol{u}}' & 1 \end{bmatrix}^{'} \\
&= \frac{\sigma^2 \sum_{s=1}^{S} p_s \left(\boldsymbol{v}_{\boldsymbol{s}} - \bar{\boldsymbol{u}} \right)^{'} \left[\boldsymbol{\Omega} + (J-1)\, \boldsymbol{\Phi} \right] \left(\boldsymbol{v}_{\boldsymbol{s}} - \bar{\boldsymbol{u}} \right)}{J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right) \right]^2}.
\end{aligned}$$

The required number of clusters is

$$\begin{aligned}
n &= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma} \right)^2 \sigma_\zeta^2}{\zeta_0^2} \\
&= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma} \right)^2 \sigma^2 \sum_{s=1}^{S} p_s \left(\boldsymbol{v}_{\boldsymbol{s}} - \bar{\boldsymbol{u}} \right)^{'} \left[\boldsymbol{\Omega} + (J-1)\, \boldsymbol{\Phi} \right] \left(\boldsymbol{v}_{\boldsymbol{s}} - \bar{\boldsymbol{u}} \right)}{\zeta_0^2 J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right) \right]^2}.
\end{aligned}$$

When $T = 3$, we have $\boldsymbol{v_1} - \bar{\boldsymbol{u}} = \begin{pmatrix} 0 & 1 - p_1 & 0 \end{pmatrix}^{'}$ and $\boldsymbol{v_2} - \bar{\boldsymbol{u}} = \begin{pmatrix} 0 & -p_1 & 0 \end{pmatrix}^{'}$.

Suppose $\boldsymbol{\Omega} = \begin{bmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{bmatrix}$ and $\boldsymbol{\Phi} = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{12} & \rho_{22} & \rho_{23} \\ \rho_{13} & \rho_{23} & \rho_{33} \end{bmatrix}$, then the required number of clusters is

$$\begin{aligned}
n &= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma} \right)^2 \sigma^2 \sum_{s=1}^{S} p_s \left(\boldsymbol{v}_{\boldsymbol{s}} - \bar{\boldsymbol{u}} \right)^{'} \left[\boldsymbol{\Omega} + (J-1)\, \boldsymbol{\Phi} \right] \left(\boldsymbol{v}_{\boldsymbol{s}} - \bar{\boldsymbol{u}} \right)}{\zeta_0^2 J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right) \right]^2} \\
&= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma} \right)^2 \sigma^2 \left[1 + (J-1)\, \rho_{22} \right]}{\zeta_0^2 J p_1 \left(1 - p_1\right)}.
\end{aligned}$$

When $T = 4$, we have $\boldsymbol{v_1} - \bar{\boldsymbol{u}} = \begin{pmatrix} 0 & 1-p_1 & p_3 & 0 \end{pmatrix}'$, $\boldsymbol{v_2} - \bar{\boldsymbol{u}} = \begin{pmatrix} 0 & -p_1 & p_3 & 0 \end{pmatrix}'$ and $\boldsymbol{v_3} - \bar{\boldsymbol{u}} = \begin{pmatrix} 0 & -p_1 & p_3 - 1 & 0 \end{pmatrix}'$. The required number of clusters is

$$
\begin{aligned}
n &= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma^2 \sum_{s=1}^{S} p_s \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[\boldsymbol{\Omega} + (J-1)\,\boldsymbol{\Phi}\right] \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)}{\zeta_0^2 J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right)\right]^2} \\
&= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma^2 \left[p_1 \left(1-p_1\right) w_{11} + 2 p_1 p_3 w_{12} + p_3 \left(1-p_3\right) w_{22}\right]}{\zeta_0^2 J \left[p_1 \left(1-p_1\right) + p_3 \left(1-p_3\right)\right]^2},
\end{aligned}
$$

where $\begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix} = \boldsymbol{\Omega^*} + (J-1)\,\boldsymbol{\Phi^*}$ with $\boldsymbol{Y}_{ij}^* = (Y_{ij2}, Y_{ij3})'$, $\boldsymbol{\Omega}^* = \mathrm{Corr}(\boldsymbol{Y}_{ij}^*)$, and $\boldsymbol{\Phi}^* = \mathrm{Corr}(\boldsymbol{Y}_{ij}^*, \boldsymbol{Y}_{ij'}^*)$, where $j \neq j'$.

A.2. Derivation of Equations (2.4) and (2.5)

When $p_1 = \cdots = p_S = \frac{1}{S}$, $\bar{u}_t = \sum_{s=1}^{S} p_s v_{st} = \frac{1}{S}\sum_{s=1}^{S} v_{st} = \frac{t-1}{T-1}$ because $T = S+1$. Then

$$\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right) = \sum_{t=1}^{T} \left(\frac{t-1}{T-1}\right)\left(1 - \frac{t-1}{T-1}\right) = \frac{T\left(T-2\right)}{6\left(T-1\right)} = \frac{S^2-1}{6S}.$$

The required number of clusters is

$$\begin{aligned} n &= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma^2 \sum_{s=1}^{S} p_s \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[\boldsymbol{\Omega} + \left(J-1\right)\boldsymbol{\Phi}\right]\left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)}{\zeta_0^2 J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right)\right]^2} \\ &= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma^2 \frac{1}{S} \sum_{s=1}^{S} \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[\boldsymbol{\Omega} + \left(J-1\right)\boldsymbol{\Phi}\right]\left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)}{\zeta_0^2 J \left[\frac{S^2-1}{6S}\right]^2} \\ &= \frac{36\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 S\sigma^2 \sum_{s=1}^{S} \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[\boldsymbol{\Omega} + \left(J-1\right)\boldsymbol{\Phi}\right]\left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)}{\zeta_0^2 J \left(S^2-1\right)^2}. \end{aligned}$$

A linear mixed-effect model with cluster and patient random effects would motivate a correlation structure such that the within-subject (longitudinal) correlation $\boldsymbol{\Omega}$ is compound symmetric (CS) with off-diagonal elements being ρ_1 and the between-subject correlation is $\boldsymbol{\Phi} = \mathbf{1}\mathbf{1}'\rho_2$. In this case we have

$$\begin{aligned} &\sum_{s=1}^{S} \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[\boldsymbol{\Omega} + \left(J-1\right)\boldsymbol{\Phi}\right]\left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right) \\ &= \sum_{s=1}^{S} \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[\begin{pmatrix} 1 & \cdots & \rho_1 \\ \vdots & \ddots & \vdots \\ \rho_1 & \cdots & 1 \end{pmatrix} + \left(J-1\right)\begin{pmatrix} \rho_2 & \cdots & \rho_2 \\ \vdots & \ddots & \vdots \\ \rho_2 & \cdots & \rho_2 \end{pmatrix}\right]\left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right) \\ &= \sum_{s=1}^{S} \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[\left(\rho_1 + \left(J-1\right)\rho_2\right)\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} + \left(1-\rho_1\right)\begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}\right]\left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right) \\ &= \left(\rho_1 + \left(J-1\right)\rho_2\right)\sum_{s=1}^{S}\left[\left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)'\mathbf{1}\right]^2 + \left(1-\rho_1\right)\sum_{s=1}^{S}\left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)'\left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right) \\ &= \left(\rho_1 + \left(J-1\right)\rho_2\right)\frac{1}{12}S\left(S^2-1\right) + \left(1-\rho_1\right)\frac{1}{6}\left(S^2-1\right) \\ &= \frac{1}{12}\left(S^2-1\right)\left[\left(S-2\right)\rho_1 + S\left(J-1\right)\rho_2 + 2\right]. \end{aligned}$$

The above derivation uses the fact that

$$\sum_{s=1}^{S}\left[\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)'\mathbf{1}\right]^2=\sum_{s=1}^{S}\left(\frac{S+1}{2}-s\right)^2=\frac{1}{12}S\left(S^2-1\right),$$

and

$$\sum_{s=1}^{S}\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)'\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)=2\sum_{s=1}^{S-1}\left(S-s\right)\left(\frac{s}{S}\right)^2=\frac{1}{6}\left(S^2-1\right).$$

Then the required number of clusters is

$$\begin{aligned}n&=\frac{36\left(z_{1-\alpha/2}+z_{1-\gamma}\right)^2S\sigma^2\sum_{s=1}^{S}\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)'\left[\boldsymbol{\Omega}+\left(J-1\right)\boldsymbol{\Phi}\right]\left(\boldsymbol{v_s}-\bar{\boldsymbol{u}}\right)}{\zeta_0^2J\left(S^2-1\right)^2}\\&=\frac{36\left(z_{1-\alpha/2}+z_{1-\gamma}\right)^2S\sigma^2\frac{1}{12}\left(S^2-1\right)\left[\left(S-2\right)\rho_1+S\left(J-1\right)\rho_2+2\right]}{\zeta_0^2J\left(S^2-1\right)^2}\\&=\frac{3\left(z_{1-\alpha/2}+z_{1-\gamma}\right)^2S\sigma^2\left[\left(S-2\right)\rho_1+S\left(J-1\right)\rho_2+2\right]}{\zeta_0^2J\left(S^2-1\right)}.\end{aligned}$$

A.3. Sample size for cross-sectional SW-CRTs

For cross-sectional SW-CRTs, the correlation structure can be modeled by $\mathbf{\Omega} = \mathbf{1}\mathbf{1}'\rho + (1-\rho)\,\boldsymbol{I}$ and $\mathbf{\Phi} = \mathbf{1}\mathbf{1}'\rho$. The required number of clusters can be expressed as

$$
\begin{aligned}
n &= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma^2 \sum_{s=1}^{S} p_s \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[\mathbf{\Omega} + (J-1)\,\mathbf{\Phi}\right] \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)}{\zeta_0^2 J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right)\right]^2} \\
&= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma^2 \sum_{s=1}^{S} p_s \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)' \left[(1-\rho)\,\boldsymbol{I} + J\mathbf{1}\mathbf{1}'\rho\right] \left(\boldsymbol{v_s} - \bar{\boldsymbol{u}}\right)}{\zeta_0^2 J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right)\right]^2} \\
&= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma^2 \sum_{s=1}^{S} p_s \left\{ J\rho \left[\sum_{t=1}^{T} \left(v_{st} - \overline{u}_t\right)\right]^2 + (1-\rho) \sum_{t=1}^{T} \left(v_{st} - \overline{u}_t\right)^2 \right\}}{\zeta_0^2 J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right)\right]^2}.
\end{aligned}
$$

Under the special case of $p_s = \frac{1}{S}$ for $s = 1, \ldots, S$, we have

$$
\begin{aligned}
n &= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma^2 \sum_{s=1}^{S} p_s \left\{ J\rho \left[\sum_{t=1}^{T} \left(v_{st} - \overline{u}_t\right)\right]^2 + (1-\rho) \sum_{t=1}^{T} \left(v_{st} - \overline{u}_t\right)^2 \right\}}{\zeta_0^2 J \left[\sum_{t=1}^{T} \overline{u}_t \left(1 - \overline{u}_t\right)\right]^2} \\
&= \frac{\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 \sigma^2 \frac{1}{12} \left(S^2 - 1\right) \left[(SJ-2)\,\rho + 2\right]}{\zeta_0^2 J \left[\frac{S^2-1}{6S}\right]^2} \\
&= \frac{3\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 S\sigma^2 \left[(SJ-2)\,\rho + 2\right]}{\zeta_0^2 J \left(S^2 - 1\right)}.
\end{aligned}
$$

A.4. Derivation of Equation (2.11)

We have

$$\begin{aligned}\boldsymbol{A}_n &= n^{-1}J\sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{B}_{ij}\boldsymbol{X}_i \\ &= n^{-1}J\sum_{i=1}^{n}\begin{bmatrix}\boldsymbol{B}_{ij} & \boldsymbol{B}_{ij}\boldsymbol{u}_i \\ \boldsymbol{u}_i'\boldsymbol{B}_{ij} & \boldsymbol{u}_i'\boldsymbol{B}_{ij}\boldsymbol{u}_i\end{bmatrix}\end{aligned}$$

As $n \to \infty$, $\boldsymbol{A}_n$ approaches

$$\begin{aligned}\boldsymbol{A} &= J\sum_{s=1}^{S}p_s\begin{bmatrix}\boldsymbol{B}_s & \boldsymbol{B}_s\boldsymbol{v}_s \\ \boldsymbol{v}_s'\boldsymbol{B}_s & \boldsymbol{v}_s'\boldsymbol{B}_s\boldsymbol{v}_s\end{bmatrix} \\ &= J\sum_{s=1}^{S}p_s\boldsymbol{W}_s'\boldsymbol{B}_s\boldsymbol{W}_s\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\boldsymbol{E}_n &= n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\right)^{\otimes 2} \\ &= n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\right)\left(\sum_{j=1}^{J}\hat{\boldsymbol{e}}_{ij}'\boldsymbol{X}_i\right) \\ &= n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\sum_{j'=1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\hat{\boldsymbol{e}}_{ij'}'\boldsymbol{X}_i\right) \\ &= n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\hat{\boldsymbol{e}}_{ij}'\boldsymbol{X}_i + 2\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\boldsymbol{X}_i'\hat{\boldsymbol{e}}_{ij}\hat{\boldsymbol{e}}_{ij'}'\boldsymbol{X}_i\right)\end{aligned}$$

As $n \to \infty$, $\boldsymbol{E}_n$ approaches

$$\boldsymbol{E} = J\sum_{s=1}^{S}p_s\boldsymbol{W}_s'\boldsymbol{G}_s\left[\boldsymbol{\Omega} + (J-1)\,\boldsymbol{\Phi}\right]\boldsymbol{G}_s\boldsymbol{W}_s$$

BIBLIOGRAPHY

Ahn, C., Heo, M. and Zhang, S. (2014), *Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research*, CRC Press.

Bacchieri, G., Barros, A. J., dos Santos, J. V., Gonçalves, H. and Gigante, D. P. (2010), 'A community intervention to prevent traffic accidents among bicycle commuters', *Revista de Saude Publica* **44**(5), 867–875.

Bailet, L. L., Repper, K. K., Piasta, S. B. and Murphy, S. P. (2009), 'Emergent literacy intervention for prekindergarteners at risk for reading failure', *Journal of Learning Disabilities* **42**(4), 336–355.

Baio, G., Copas, A., Ambler, G., Hargreaves, J., Beard, E. and Omar, R. Z. (2015), 'Sample size calculation for a stepped wedge trial', *Trials* **16**(1), 354.

Barker, D., McElduff, P., D'Este, C. and Campbell, M. (2016), 'Stepped wedge cluster randomised trials: a review of the statistical methodology used and available', *BMC Medical Research Methodology* **16**(1), 69.

Bashour, H. N., Kanaan, M., Kharouf, M. H., Abdulsalam, A. A., Tabbaa, M. A. and Cheikha, S. A. (2013), 'The effect of training doctors in communication skills on women's satisfaction with doctor–woman relationship during labour and delivery: a stepped wedge cluster randomised trial in damascus', *BMJ Open* **3**(8), e002674.

Beard, E., Lewis, J. J., Copas, A., Davey, C., Osrin, D., Baio, G., Thompson, J. A., Fielding, K. L., Omar, R. Z., Ononge, S. et al. (2015), 'Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014', *Trials* **16**(1), 353.

Brown, C. A. and Lilford, R. J. (2006), 'The stepped wedge trial design: a systematic review', *BMC Medical Research Methodology* **6**(1), 54.

Camacho, A., Eggo, R. M., Funk, S., Watson, C. H., Kucharski, A. J. and Edmunds, W. J. (2015), 'Estimating the probability of demonstrating vaccine efficacy in the declining ebola epidemic: a bayesian modelling approach', *BMJ Open* **5**(12), e009346.

Carlin, B. P. and Louis, T. A. (2008), *Bayesian Methods for Data Analysis*, CRC Press.

Choi, S. C., Smith, P. J. and Becker, D. P. (1985), 'Early decision in clinical trials when the treatment differences are small: experience of a controlled trial in head trauma', *Controlled Clinical Trials* **6**(4), 280–288.

Choi, S. and Pepple, P. (1989), 'Monitoring clinical trials based on predictive probability of significance', *Biometrics* pp. 317–323.

Chow, S.-C., Chang, M. and Pong, A. (2005), 'Statistical consideration of adaptive methods in clinical development', *Journal of Biopharmaceutical Statistics* **15**(4), 575–591.

Copas, A. J., Lewis, J. J., Thompson, J. A., Davey, C., Baio, G. and Hargreaves, J. R. (2015), 'Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches', *Trials* **16**(1), 352.

Crowder, M. (1995), 'On the use of a working correlation matrix in using generalised linear models for repeated measures', *Biometrika* **82**(2), 407–410.

Cunanan, K. M., Carlin, B. P. and Peterson, K. A. (2016), 'A practical bayesian stepped wedge design for community-based cluster-randomized clinical trials: The british columbia telehealth trial', *Clinical Trials* **13**(6), 641–650.

Donner, A. and Klar, N. (2000), *Design and Analysis of Cluster Randomization Trials in Health Research*, London: Arnold.

Edwards, S. J. (2013), 'Ethics of clinical science in a public health emergency: drug discovery at the bedside', *The American Journal of Bioethics* **13**(9), 3–14.

Fan, C., Zhang, D. and Zhang, C. H. (2013), 'A comparison of bias-corrected covariance estimators for generalized estimating equations', *Journal of Biopharmaceutical Statistics* **23**(5), 1172–1187.

Fay, M. P. and Graubard, B. I. (2001), 'Small-sample adjustments for wald-type tests using sandwich estimators', *Biometrics* **57**(4), 1198–1206.

Gelman, A. (2006), 'Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)', *Bayesian Analysis* **1**(3), 515–534.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013), *Bayesian Data Analysis*, Chapman and Hall/CRC.

Geman, S. and Geman, D. (1987), Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *in* 'Readings in Computer Vision', Elsevier, pp. 564–584.

Girling, A. J. and Hemming, K. (2016), 'Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models', *Statistics in Medicine* **35**(13), 2149–2166.

Gould, A. L. (2005), 'Timing of futility analyses for proof of concept trials', *Statistics in medicine* **24**(12), 1815–1835.

Grieve, A., Choi, S. and Pepple, P. (1991), 'Predictive probability in clinical trials', *Biometrics* pp. 323–330.

Hargreaves, J. R., Prost, A., Fielding, K. L. and Copas, A. J. (2015), 'How important is randomisation in a stepped wedge trial?', *Trials* **16**(1), 359.

Heitjan, D. F. (1997), 'Bayesian interim analysis of phase II cancer clinical trials', *Statistics in Medicine* **16**(16), 1791–1802.

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J. and Lilford, R. J. (2015), 'The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting', *BMJ* **350**, h391.

Herson, J. (1979), 'Predictive probability early termination plans for phase II clinical trials', *Biometrics* pp. 775–783.

Hooper, R., Teerenstra, S., de Hoop, E. and Eldridge, S. (2016), 'Sample size calculation for stepped wedge and other longitudinal cluster randomised trials', *Statistics in Medicine* **35**(26), 4718–4728.

Hussey, M. A. and Hughes, J. P. (2007), 'Design and analysis of stepped wedge cluster randomized trials', *Contemporary Clinical Trials* **28**(2), 182–191.

Jennison, C. and Turnbull, B. W. (2005), 'Meta-analyses and adaptive group sequential designs in the clinical development process', *Journal of Biopharmaceutical Statistics* **15**(4), 537–558.

Johns, D. and Andersen, J. S. (1999), 'Use of predictive probabilities in phase II and phase III clinical trials', *Journal of Biopharmaceutical Statistics* **9**(1), 67–79.

Kauermann, G. and Carroll, R. J. (2001), 'A note on the efficiency of sandwich covariance matrix estimation', *Journal of the American Statistical Association* **96**(456), 1387–1396.

Kitson, A. L., Schultz, T. J., Long, L., Shanks, A., Wiechula, R., Chapman, I. and Soenen, S. (2013), 'The prevention and reduction of weight loss in an acute tertiary care setting: protocol for a pragmatic stepped wedge randomised cluster trial (the PRoWL project)', *BMC Health Services Research* **13**(1), 299.

Lee, J. J., Gu, X. and Liu, S. (2010), 'Bayesian adaptive randomization designs for targeted agent development', *Clinical Trials* **7**(5), 584–596.

Li, F., Turner, E. L. and Preisser, J. S. (2018), 'Sample size determination for GEE analyses of stepped wedge cluster randomized trials', *Biometrics* **74**(4), 1450–1458.

Liang, K. Y. and Zeger, S. L. (1986), 'Longitudinal data analysis for discrete and continuous outcomes using generalized linear models', *Biometrika* **84**, 3–32.

Lou, Y., Cao, J., Zhang, S. and Ahn, C. (2017), 'Sample size estimation for a two-group comparison of repeated count outcomes using GEE', *Communications in Statistics-Theory and Methods* **46**(14), 6743–6753.

Mancl, L. A. and DeRouen, T. A. (2001), 'A covariance estimator for GEE with improved small-sample properties', *Biometrics* **57**(1), 126–134.

Martin, J., Taljaard, M., Girling, A. and Hemming, K. (2016), 'Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials', *BMJ Open* **6**(2), e010166.

McCaffrey, D. F. and Bell, R. M. (2006), 'Improved hypothesis testing for coefficients in generalized estimating equations with small samples of clusters', *Statistics in Medicine* **25**(23), 4081–4098.

McDonald, B. W. (1993), 'Estimating logistic regression parameters for bivariate binary data', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 391–397.

Mhurchu, C. N., Gorton, D., Turley, M., Jiang, Y., Michie, J., Maddison, R. and Hattie, J. (2013), 'Effects of a free school breakfast programme on children's attendance, academic achievement and short-term hunger: results from a stepped-wedge, cluster randomised controlled trial', *Journal of Epidemiology & Community Health* **67**(3), 257–264.

Montaner, J. S., Lawson, L. M., Levitt, N., Belzberg, A., Schechter, M. T. and Ruedy, J. (1990), 'Corticosteroids prevent early deterioration in patients with moderately severe pneumocystis carinii pneumonia and the acquired immunodeficiency syndrome (AIDS)', *Annals of Internal Medicine* **113**(1), 14–20.

Morel, J. G., Bokossa, M. and Neerchal, N. K. (2003), 'Small sample correction for the variance of GEE estimators', *Biometrical Journal* **45**(4), 395–409.

Moulton, L. H., Golub, J. E., Durovni, B., Cavalcante, S. C., Pacheco, A. G., Saraceni, V., King, B. and Chaisson, R. E. (2007), 'Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention', *Clinical Trials* **4**(2), 190–199.

Pan, W. and Wall, M. M. (2002), 'Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations', *Statistics in Medicine* **21**(10), 1429–1441.

Reiertsen, O., Larsen, S., Trondsen, E., Edwin, B., Faerden, A. and Rosseland, A. (1997), 'Randomized controlled trial with sequential design of laparoscopic versus conventional appendicectomy', *British Journal of Surgery* **84**(6), 842–847.

Reuther, S., Holle, D., Buscher, I., Dortmann, O., Müller, R., Bartholomeyczik, S. and Halek, M. (2014), 'Effect evaluation of two types of dementia-specific case conferences in German nursing homes (FallDem) using a stepped-wedge design: study protocol for a randomized controlled trial', *Trials* **15**(1), 319.

Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C. N. and Ricks, M. L. (2001), 'Optimal adaptive designs for binary response trials', *Biometrics* **57**(3), 909–913.

Schultz, T. J., Kitson, A. L., Soenen, S., Long, L., Shanks, A., Wiechula, R., Chapman, I. and Lange, K. (2014), 'Does a multidisciplinary nutritional intervention prevent nutritional decline in hospital patients? A stepped wedge randomised cluster trial', *e-SPEN Journal* **9**(2), e84–e90.

Spiegelhalter, D. J., Freedman, L. S. and Blackburn, P. R. (1986), 'Monitoring clinical trials: conditional or predictive power?', *Controlled Clinical Trials* **7**(1), 8–17.

Vahedi, K., Vicaut, E., Mateo, J., Kurtz, A., Orabi, M., Guichard, J.-P., Boutron, C., Couvreur, G., Rouanet, F., Touzé, E. et al. (2007), 'Sequential-design, multicenter, randomized, controlled trial of early decompressive craniectomy in malignant middle cerebral artery infarction (DECIMAL Trial)', *Stroke* **38**(9), 2506–2517.

van Holland, B. J., de Boer, M. R., Brouwer, S., Soer, R. and Reneman, M. F. (2012), 'Sustained employability of workers in a production environment: design of a stepped wedge trial to evaluate effectiveness and cost-benefit of the POSE program', *BMC Public Health* **12**(1), 1003.

Woertman, W., de Hoop, E., Moerbeek, M., Zuidema, S. U., Gerritsen, D. L. and Teerenstra, S. (2013), 'Stepped wedge designs could reduce the required sample size in cluster randomized trials', *Journal of Clinical Epidemiology* **66**(7), 752–758.

Zhang, S., Cao, J. and Ahn, C. (2018), 'Sample size calculation for before–after experiments with partially overlapping cohorts', *Contemporary clinical trials* **64**, 274–280.

Zhang, Z., Hamagami, F., Lijuan Wang, L., Nesselroade, J. R. and Grimm, K. J. (2007), 'Bayesian analysis of longitudinal data using growth curve models', *International Journal of Behavioral Development* **31**(4), 374–383.

Zhou, X., Liu, S., Kim, E. S., Herbst, R. S. and Lee, J. J. (2008), 'Bayesian adaptive design for targeted therapy development in lung cancer—a step toward personalized medicine', *Clinical Trials* **5**(3), 181–193.

Zhu, H. and Yu, Q. (2017), 'A bayesian sequential design using alpha spending function to control type I error', *Statistical Methods in Medical Research* **26**(5), 2184–2196.

Zhu, H., Zhang, S. and Ahn, C. (2017), 'Sample size considerations for split-mouth design', *Statistical Methods in Medical Research* **26**(6), 2543–2551.

Ziegler, A. and Vens, M. (2010), 'Generalized estimating equations', *Methods of Information in Medicine* **49**(05), 421–425.