Summer 8-6-2019

Sample Size Calculation of Clinical Trials with Correlated Outcomes

Dateng Li
Southern Methodist University, datengl@smu.edu

SAMPLE SIZE CALCULATION OF CLINICAL TRIALS WITH CORRELATED OUTCOMES

Approved by:

Dr. Jing Cao
Associate Professor in Department of Statistical Science, SMU

Dr. Daniel Heitjan
Professor in Department of Statistical Science, SMU & Population & Data Sciences, UTSW

Dr. Hon Keung Tony Ng
Professor in Department of Statistical Science, SMU

Dr. Song Zhang
Associate Professor in Department of Population & Data Sciences, UTSW

SAMPLE SIZE CALCULATION OF CLINICAL TRIALS WITH CORRELATED OUTCOMES

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 Statistical Science

by

Dateng Li

B.S., Mathematics, Oklahoma State University

August 6, 2019

ACKNOWLEDGMENTS

Firstly of all, I would like to express my most sincere gratitude to my advisors Drs. Jing Cao and Song Zhang. They are the greatest mentors I could imagine: worked with me closely on my research and always provided the help and advice I needed. Their enthusiasm for statistics and science inspired me in all the time of my research. Without their training, I would not be able to become a statistician. They do not only train me in statistics, but also showed me how to treat people with respect and kindness.

Secondly, I would like to thank the rest of my thesis committee members Drs. Daniel Heitjan and Hon Keung Tony Ng, for their time and academic efforts to help me greatly improve my dissertation. I also want to thank all the other faculty members in the Department of Statistical Science at SMU for their excellent teaching and help during my PhD study.

I would also like to take this moment to express my deep appreciation to all of my friends. Thank you for listening to me, offering me advice and supporting me throughout the journey. You have made my life in Dallas so enjoyable.

Finally, I would like to acknowledge with gratitude, the support and love of my parents and my family. Thank you all for encouraging me in all of my pursuits and inspiring me to follow my dreams. You are the most important people to me in the world and I dedicate this dissertation to you.

Li, Dateng B.S., Mathematics, Oklahoma State University

Sample Size Calculation of Clinical Trials with Correlated Outcomes

Advisor: Dr. Jing Cao

Doctor of Philosophy degree conferred August 6, 2019
Dissertation completed July 10, 2019

In this thesis, we investigate sample size calculation for three kinds of clinical trials: (1). Randomized controlled trials (RCTs) with longitudinal count outcomes; (2). Cluster randomized trials (CRTs) with count outcomes; (3). CRTs with multiple binary co-primary endpoints.

Statistical inference based on RCTs with longitudinal count measurements are frequently performed in biomedical studies. Most of existing sample size calculation methods for count outcomes are developed under the Poisson model. Deviation from the Poisson assumption (equality of mean and variance) has been widely documented in practice, which makes it more desirable to have sample size methods with more realistic assumptions to ensure valid experimental design. In this thesis we investigate sample size calculation for clinical trials with longitudinal count measurements based on the negative binomial distribution. This approach is flexible to accommodate over-dispersion and unequal measurement intervals, as well as arbitrary randomization ratios, missing data patterns, and correlation structures. In addition, the derived sample size formulas have closed forms for the comparison of both slopes and time-averaged responses, which greatly reduces the burden of implementation in practice.

Pragmatic clinical trials are designed to test intervention in real-world health system practice in order to maximize the applicability and generalizability [45]. One common feature of pragmatic clinical trials is the use of clustered randomization, where clusters of patients (formed by physicians or clinics, for example) are the units of randomization to

avoid “contamination” between intervention and control participants. For example, in studies of dietary change, participants in the control group might learn about the experimental diet and adopt to the experimental diet themselves. For CRTs with count outcomes, we propose to directly incorporate pragmatic issues (e.g., over-dispersion, variability of cluster sizes, etc.) into sample size calculation. The proposed method is developed based on the GEE approach and it is advantageous in that the sample size formula has a closed form, which facilitates its implementation in pragmatic CRTs. We also show in theory that ignoring these pragmatic features will under-estimate the sample size, which leads to an under-powered trial.

Recently, with increasing complexity of medical therapies and technological advances in monitoring multiple outcomes, many clinical trials attempt to evaluate multiple co-primary endpoints. In this study we also present a sample size calculation method for CRTs with $K \geq 2$ binary co-primary endpoints. Three types of correlation structures are considered: inter-subject correlation within endpoint, intra-subject correlation across endpoints, and inter-subject correlation across endpoints. A closed-form joint distribution of the K test statistics is derived, which facilitates the evaluation of power and type I error for arbitrarily constructed hypotheses. We further present a theorem that characterizes the relationship between the correlation structures and testing power.

Extensive simulations are conducted to demonstrate that the proposed methods can maintain the nominal levels of power and type I error over a wide range of design configurations. We also illustrate the applications of the proposed approaches using real clinical trials.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

This dissertation is dedicated to my family and friends.

CHAPTER 1

INTRODUCTION

At the design stage of a clinical trial, one of the key considerations is sample size calculation. On one hand, insufficient sample size will lead to under-powered clinical trials to detect a difference between experiment groups. On the other hand, redundant sample size will introduce unnecessary expenditure of resources or expose participants to possible clinical risks. In this thesis, we investigate sample size calculation for three kinds of clinical trials: (1). Randomized controlled trials (RCTs) with longitudinal count outcomes; (2). Cluster randomized trials (CRTs) with count outcomes; (3). CRTs with multiple binary co-primary endpoints. The rest of this chapter is organized as follows. In Sections 1.1 - 1.3, we provide the literature review on sample size calculation for each of the three kinds of clinical trials, respectively. The limitations of current methods are also discussed. In Section 1.4, we give an overview on our proposed work.

1.1. Sample Size Calculation for RCTs with Longitudinal Count Outcomes

Statistical inference based on longitudinal count measurements are frequently conducted in biomedical studies [2]. For example, in oncology trials the number of adverse events are closely monitored throughout the follow-up period [56], while in epilepsy trials the number of seizure episodes are periodically recorded from each patient [19]. The goal in such longitudinal studies is to determine whether the slopes (i.e., rates of change) are significantly different between the control and treatment groups. On the other hand, researchers may measure the response multiple times, hoping that the time-averaged responses (TAR) can be more accurate than a single measurement [34]. Most of existing sample size calculation methods for count outcomes are developed under the Poisson

model. Ogungbenro & Aarons [42] and Amatya et al. [3] employed mixed-effects Poisson models to compute sample sizes for repeated count outcomes. Note that under the mixed-effects model approach, the various correlation structures (specified through random effects) are assumed at the level of the latent log-link function, not at the level of observed counts. Alternatively, the generalized estimating equation (GEE) approach [33] has been employed in sample size calculation, where correlation is directly modeled in the distribution of count measurements. Patel & Rowe [44] developed a GEE sample size formula for comparing two linear curves of longitudinal count outcomes under the Poisson model, which requires complete observations from every patient. Lou et al. [35] proposed a more flexible sample size method based on GEE that accommodates arbitrary missing data patterns and correlation structures under the Poisson assumption. All the aforementioned sample size methods are derived for the comparison of slopes between two experiment groups. Sample size calculation for the comparison of TAR has received relatively less attention. In the context of continuous and binary outcomes, sample size calculation for the comparison of TAR has been investigated by Liu & Wu [34] and Zhang & Ahn [62]. In terms of count outcomes, Asendorf et al. [5] proposed a sample size approach for longitudinal count data based on a binomial thinning model. This approach accommodates marginal Poisson or negative binomial (NB) distribution, but it only considers the first order autoregressive correlation structure.

One limitation of the Poisson model stems from its underlying assumption that the mean and variance of the count outcome are equal. In practice this assumption might be too restrictive. Deviation from this assumption has been widely reported in biomedical research [11, 12]. Specifically, researchers have frequently encountered the phenomenon of over-dispersion where the variance of a count variable is greater than the mean. Imposing a Poisson model on over-dispersed data will result in under-estimated variance and incorrect conclusion. As an alternative to the Poisson model, the NB model is gaining popularity due to its flexibility to accommodate over-dispersion. Zhu & Lakkis [63] developed sample size calculation methods to compare the rates of two NB distributions. Tang

[54] investigated sample size methods for comparing NB rates in noninferiority and equivalence trials with unequal follow-up times. These two methods, however, only consider independent NB measurements. They are not applicable to clinical trials where longitudinal design is employed.

1.2. Sample Size Calculation for CRTs with Count Outcomes

Pragmatic clinical trials are designed to test intervention in the full spectrum of realistic clinical settings in order to maximize the applicability and generalizability [45]. The research question of interest is whether an intervention actually works in real-world health system practice. One common feature of pragmatic clinical trials is the use of clustered randomization, where clusters of patients (formed by physicians or clinics, for example) are the units of randomization to avoid "contamination" and to allow the intervention to be applied as it would be in real-world practice [48]. One example of contamination is that when a physician simultaneously provides care to patients enrolled in different arms, leakage of treatments might occur between two arms, resulting in an observed intervention effect that is diluted and biased toward the null. Thus in CRTs, an individual physician would only provide care to patients in one treatment arm.

When conducted in pragmatic settings, CRTs create unique design and analytic challenges to biostatisticians. A well recognized statistical issue in CRTs is that responses tend to be more similar within clusters than those across clusters. This within-cluster similarity is quantified by the intraclass correlation coefficient (ICC) [39], and there has been extensive investigation in design methods for CRTs to properly account for ICC. See for example, Roberts & Roberts [50] and Eldridge et al. [15]. Amatya et al. [3] proposed a sample size calculation method for CRTs with a count outcome. It was developed based on the Poisson regression model under the assumption that the number of patients (cluster size) is equal across all clusters. Such an assumption might be too restrictive, especially for pragmatic CRTs conducted in realistic clinical settings. The clusters are

usually formed naturally with different cluster size due to practice scale, patient base, and logistics, etc. Many researchers have shown that ignoring cluster size variability in sample size calculation can lead to under-powered studies [1, 20]. Wang et al. [60] proposed to incorporate cluster size variability into sample size calculation for CRTs with count outcomes, where a correction term defined based on the coefficient of variation in cluster size is included [37].

It is noteworthy that the methods in Amatya et al. [3] and Wang et al. [60] were both developed under the Poisson model, which by definition imposes the restriction that the mean and variance of the count outcome are equal [17]. In practice, however, the phenomenon over-dispersion has been widely reported in biomedical research [11, 12]. In its presence, employing a mis-specified Poisson model will lead to under-estimated sample sizes. One common approach to account for over-dispersion is to model the count data using the quasi Poisson distribution [61], which assumes the variance is a linear function of the mean, with a slope greater than one to accommodate over-dispersion. Alternatively, methods based on the negative binomial (NB) model, which assume the variance follows a quadratic function of the mean [21], are gaining popularity.

Finally, both Amatya et al. [3] and Wang et al.[60] assume that all patients contribute an equal length of follow-up, during which the counts of a certain event are measured. In pragmatic trials, patients may experience treatment discontinuation or dropouts, leads to different length of follow-up periods. For example, in a phase III trial [9] of a novel phosphodiesterase 4 (PDE4) inhibitor for chronic obstructive pulmonary disease (COPD) treatment, the primary endpoint was the incidence count of moderate or severe COPD exacerbation. The enrolled patients were initially scheduled have a follow-up period of 52 weeks. However, roughly 30% of them withdrew early from the studies. As a result, the incidence counts are measured over different lengths of follow-up across patients. In the context of individual randomization trials, there has been some development in sample size methods for count outcomes measured over different follow-up periods [31, 54]. To the best of our knowledge, there is no such development for CRTs.

1.3. Sample Size Calculation for CRTs with Multiple Binary Co-Primary Endpoints

Extensive research on power/sample size calculation has been conducted for CRTs with a single primary outcome. For example, Lake et al. [28] investigated sample size re-estimation for CRTs; Manatunga et al. [37] and Wang et al. [59] explored the impact of random variability in cluster size on sample size requirement; Raudenbush [49] studied the optimal allocation of resources within and between clusters. More comprehensive reviews can be found in Gao et al. [18] and Murray et al. [40].

Recently, the increasing complexity of medical therapies and technological advances in obtaining a wider variety of measurements from study subjects have made multiple endpoints and multiple testing increasingly important in clinical trials [8]. Sample size calculation for RCTs with multiple co-primary endpoints have been studied by many researchers. For example, Sozu et al. [52, 53] investigated sample size calculation approaches for RCTs with multiple binary endpoints or a mixture of continuous and binary endpoints. Lafaye de Micheaux et al. [27] considered the case where multiple continuous correlated endpoints are of primary interest. However, to the best of our knowledge, there has been very limited research on investigating sample size calculation for CRTs when multiple co-primary endpoints are evaluated. The challenge lies in the fact that when co-primary endpoints are measured in a CRT, it gives rise to multiple types of dependence. For example, the measurements of two endpoints from the same subject are dependent in a way different from that of two measurements of the same endpoint from two subjects within the same cluster. Therefore, an exchangeable correlation structure which only accounts for the ICC becomes inadequate to model the correlation structure in CRTs involving multiple endpoints. The multiple layers of the correlation structure requests proper accomodation in data analysis and experimental design.

1.4. The Proposed Work

The thesis contains three parts, which are respectively presented in Chapter 2-4.

In Chapter 2, we investigate sample size calculation for RCTs with longitudinal count measurements based on the NB distribution under the GEE approach framework. Our proposed approach is flexible to accommodate over-dispersion and unequal measurement intervals, as well as arbitrary randomization ratios, missing data patterns, and correlation structures. The derived sample size formulas have closed forms for the comparison of both slopes and time-averaged responses, which greatly reduces the burden of implementation in practice. We have conducted extensive simulation study to demonstrate that the proposed method can maintain the nominal levels of power and type I error over a wide range of design configurations. We illustrate the application of this approach using a real epileptic trial.

In Chapter 3, we consider the problem of sample size calculation for CRTs with a count outcome. Particularly, we propose to directly incorporate pragmatic issues (e.g., over-dispersion, varying cluster sizes, etc.) into the calculation. The proposed method is developed based on the GEE approach and it is advantageous in that the sample size formula has a closed form, which facilitates its implementation in pragmatic CRTs. We also show in theory that ignoring these pragmatic features will under-estimate the sample size, which leads to an under-powered trial. We assess the performance of the proposed sample size method through extensive simulation studies. An application example based on a real clinical trial is presented.

In Chapter 4, we present a sample size calculation method for CRTs with $K \geq 2$ binary co-primary endpoints. It is developed based on the GEE approach, where three types of correlation was considered: inter-subject correlation within an endpoint, intra-subject correlation across endpoints, and inter-subject correlation across endpoints. A closed-form joint distribution of the K test statistics was derived, which can be used to evaluate power and type I error for arbitrarily constructed hypotheses. We further present a theorem that

characterizes the relationship between the correlation structures and testing power. We assess the performance of our proposed method based on extensive simulation studies, finished with application to a real clinical trial is presented.

Finally, some discussion and concluding remarks are presented in Chapter 5.

CHAPTER 2
SAMPLE SIZE CALCULATION FOR CLINICAL TRIALS WITH LONGITUDINAL COUNT MEASUREMENTS BASED ON THE NEGATIVE BINOMIAL DISTRIBUTION

In this chapter, we investigate sample size calculation for clinical trials with longitudinal count measurements based on the negative binomial distribution. This approach is flexible to accommodate over-dispersion and unequal measurement intervals, as well as arbitrary randomization ratios, missing data patterns, and correlation structures.

2.1. Statistical Model and Sample Size Approach for Comparing TAR

Suppose in a clinical trial patients are randomized to the treatment and control groups. Let $\boldsymbol{t} = (t_1, \cdots, t_m)$ be the time of patient visits. For the ith ($i = 1, \cdots, n$) subject, the primary outcome, denoted by y_{ij}, is the count of a certain event (eg., epilepsy episode or adverse drug reaction) during an interval of length T_j. Here n is the total sample size, m is the number of measurements observed from each subject, and the lengths of measurement intervals (T_j's) can be unequal. We model y_{ij} by a negative binomial distribution,

$$P(y_{ij}) = \frac{\Gamma(y_{ij} + v^{-1})}{y_{ij}!\,\Gamma(v^{-1})} \left(\frac{v\mu_{ij}}{1 + v\mu_{ij}}\right)^{y_{ij}} \left(\frac{1}{1 + v\mu_{ij}}\right)^{v^{-1}}, \tag{2.1}$$

where μ_{ij} and v are model parameters and $\Gamma(\cdot)$ is the gamma function. Under this specification, we have $E(y_{ij}) = \mu_{ij}$ and $Var(y_{ij}) = \mu_{ij} + v\mu_{ij}^2$. Parameter v controls how much the variance deviates from the mean, which is called the dispersion parameter. It has been shown that the Poisson model is a limiting case of a NB model with $v \to 0$ [46]. When making inference about TAR, we assume that the mean of y_{ij} to be constant over time,

and model μ_{ij} by

$$\log(\mu_{ij}) = \log(T_j) + \beta_1 + \beta_2 r_i,$$

where $r_i = 1/0$ indicates that subject i is assigned to the treatment/control group, $\log(T_j)$ is the offset to account for potentially non-equal measurement intervals across visits, β_1 is the rate of the control group on the log scale, and β_2 models the ratio between the treatment and control groups, representing the treatment effect. Define $\boldsymbol{Z}_{ij} = (1, r_i)' = \boldsymbol{Z}_i$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, we have

$$\mu_{ij}(\boldsymbol{\beta}) = T_j \exp(\boldsymbol{Z}_{ij}'\boldsymbol{\beta}) = T_j \exp(\boldsymbol{Z}_i'\boldsymbol{\beta}).$$

Let the within-subject correlation be $Corr(y_{ij}, y_{ij'}) = \rho_{jj'}$ with $\rho_{jj'} = 1$ for $j = j'$. Observations from different subjects are assumed to be independent. We are interested in testing $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$. In clinical trials researchers are frequently faced with the problem of missing data due to missed clinic visits or patient dropout, etc. We directly incorporate missing data into sample size calculation. First we define indicator Δ_{ij}, which takes value 1 if the jth outcome of the ith subject is observed, and value 0 if it is missing. Then $\delta_j = E(\Delta_{ij})$ is the probability of the jth outcome being observed, which is assumed to be equal across all subjects. It's also reasonable to assume $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_m$. Furthermore, $\delta_{jj'} = E(\Delta_{ij}\Delta_{ij'})$ is the joint probability that a subject has observations at both times t_j and $t_{j'}$. Note that $\delta_{jj} = \delta_j$. The introduction of δ_j and $\delta_{jj'}$ allows us to accommodate a wide range of missing patterns. For example, under the independent missing pattern (IM), missing data occur independently over time, so we have $\delta_{jj'} = \delta_j\delta_{j'}$ ($j \neq j'$). Under the monotone missing pattern (MM), a subject missing the observation at t_j misses all subsequent observations, hence $\delta_{jj'} = \delta_{\max\{j,j'\}}$. The following derivations are presented under the missing completely at random (MCAR) assumption.

Utilizing an independent working correlation structure, it can be shown that the GEE estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)$ is solved from the following estimating equation,

$$S_n(\boldsymbol{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m} \Delta_{ij} \frac{1}{1 + v\mu_{ij}(\boldsymbol{\beta})} [y_{ij} - \mu_{ij}(\boldsymbol{\beta})] \boldsymbol{Z}_i = 0, \tag{2.2}$$

which, with the introduction of Δ_{ij}, can accommodate missing data. Equation (2.2) is generally solved through the Newton-Raphson algorithm. At the lth iteration,

$$\hat{\boldsymbol{\beta}}^{(l)} = \hat{\boldsymbol{\beta}}^{(l-1)} + n^{-\frac{1}{2}} \boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\beta}}^{(l-1)}) \boldsymbol{S}_n(\hat{\boldsymbol{\beta}}^{(l-1)}), \tag{2.3}$$

where

$$\boldsymbol{A}_n(\hat{\boldsymbol{\beta}}) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \Delta_{ij} \boldsymbol{Z}_i \boldsymbol{Z}_i' \frac{\mu_{ij}(\hat{\boldsymbol{\beta}})}{1 + v\mu_{ij}(\hat{\boldsymbol{\beta}})}. \tag{2.4}$$

Under the special case of equal measurement intervals, $T_1 = ... = T_m = T$, $\hat{\boldsymbol{\beta}}$ can be solved analytically,

$$\begin{aligned} \hat{\beta}_1 &= \log \left(\frac{\sum_{i=1}^{n} (1 - r_i) \sum_{j=1}^{m} \Delta_{ij} y_{ij}}{T \sum_{i=1}^{n} (1 - r_i) \sum_{j=1}^{m} \Delta_{ij}} \right), \\ \hat{\beta}_2 &= \log \left(\frac{\sum_{i=n}^{n} r_i \sum_{j=1}^{m} \Delta_{ij} y_{ij}}{\sum_{i=n1}^{n} r_i \sum_{j=1}^{m} \Delta_{ij}} \right) - \log \left(\frac{\sum_{i=1}^{n} (1 - r_i) \sum_{j=1}^{m} \Delta_{ij} y_{ij}}{\sum_{i=1}^{n} (1 - r_i) \sum_{j=1}^{m} \Delta_{ij}} \right). \end{aligned} \tag{2.5}$$

Derivation of Equation (2.5) is presented in Appendix A.1.

Note that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ approximately follows a normal distribution with mean 0 and variance $\boldsymbol{\Sigma}_n = \boldsymbol{A}_n^{-1} \boldsymbol{V}_n \boldsymbol{A}_n^{-1}$ [33], where

$$\boldsymbol{V}_n(\hat{\boldsymbol{\beta}}) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{j'=1}^{m} \Delta_{ij} \Delta_{ij'} \hat{\epsilon}_{ij} \hat{\epsilon}_{ij'} \frac{1}{[1 + v\mu_{ij}(\hat{\boldsymbol{\beta}})]^2} \boldsymbol{Z}_i \boldsymbol{Z}_i'.$$

Here, $\hat{\epsilon}_{ij} = y_{ij} - \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{Z}_i)$. Let $\hat{\sigma}_2^2$ be the (2,2)th element of $\boldsymbol{\Sigma}_n$. We reject $H_0 : \beta_2 = 0$ if $|\frac{\sqrt{n}\hat{\beta}_2}{\hat{\sigma}_2}|$ is greater than $z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$th percentile of the standard

normal distribution. Let $\boldsymbol{A}$ and $\boldsymbol{V}$ denote the limits of $\boldsymbol{A}_n$ and $\boldsymbol{V}_n$ as $n \to \infty$, it follows that $\boldsymbol{\Sigma}_n$ converges to $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}$. Let σ_2^2 denotes the (2,2)th element of $\boldsymbol{\Sigma}$. Then given the true treatment effect $\beta_2 = \beta_{20}$, the sample size to achieve power $1-\gamma$ at two-sided significance level α can be calculated by

$$n = \frac{\sigma_2^2(z_{1-\alpha/2} + z_{1-\gamma})^2}{\beta_{20}^2}.$$

Here we show that a closed-form expression of σ_2^2 can be derived, which results in a closed-form sample size formula. Specifically, as $n \to \infty$, we have

$$\begin{aligned}
\boldsymbol{A} &= E\left[\sum_{j=1}^{m} \Delta_{ij} \frac{\mu_{ij}(\boldsymbol{\beta})}{1+v\mu_{ij}(\boldsymbol{\beta})} \begin{pmatrix} 1 & r_i \\ r_i & r_i^2 \end{pmatrix}\right] \\
&= (1-\bar{r}) \sum_{j=1}^{m} \delta_j \frac{T_j\mu_1}{1+vT_j\mu_1} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \bar{r} \sum_{j=1}^{m} \delta_j \frac{T_j\mu_2}{1+vT_j\mu_2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} (1-\bar{r}) \sum_{j=1}^{m} \delta_j \frac{T_j\mu_1}{1+vT_j\mu_1} + \bar{r} \sum_{j=1}^{m} \delta_j \frac{T_j\mu_2}{1+vT_j\mu_2} & \bar{r} \sum_{j=1}^{m} \delta_j \frac{T_j\mu_2}{1+vT_j\mu_2} \\ \bar{r} \sum_{j=1}^{m} \delta_j \frac{T_j\mu_2}{1+vT_j\mu_2} & \bar{r} \sum_{j=1}^{m} \delta_j \frac{T_j\mu_2}{1+vT_j\mu_2} \end{pmatrix},
\end{aligned}$$

where $\mu_1 = \exp(\beta_1)$ and $\mu_2 = \exp(\beta_1 + \beta_2)$. Similarly,

$$\begin{aligned}
\boldsymbol{V} &= E\left[\sum_{j=1}^{m}\sum_{j'=1}^{m} \Delta_{ij}\Delta_{ij'}\epsilon_{ij}\epsilon_{ij'} \frac{1}{[1+v\mu_{ij}(\boldsymbol{\beta})]^2} \begin{pmatrix} 1 & r_{j'} \\ r_j & r_j r_{j'} \end{pmatrix}\right] \qquad (2.6) \\
&= (1-\bar{r}) \sum_{j=1}^{m}\sum_{j'=1}^{m} \frac{\delta_{jj'}\rho_{jj'}\mu_1\sqrt{T_jT_{j'}}}{\sqrt{(1+vT_j\mu_1)(1+vT_{j'}\mu_1)}} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \bar{r} \sum_{j=1}^{m}\sum_{j'=1}^{m} \frac{\delta_{jj'}\rho_{jj'}\mu_2\sqrt{T_jT_{j'}}}{\sqrt{(1+vT_j\mu_2)(1+vT_{j'}\mu_2)}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.
\end{aligned}$$

Here we use the fact that $E(\epsilon_{ij}\epsilon_{ij'}) = \rho_{jj'}\sqrt{(1+vT_j\mu_i)(1+vT_{j'}\mu_i)}$. Because $\boldsymbol{A}$ and $\boldsymbol{V}$ are 2×2 matrices, it is easy to show that

$$\sigma_2^2 = \frac{\sum_{j=1}^m\sum_{j'=1}^m \frac{\delta_{jj'}\rho_{jj'}\mu_1\sqrt{T_jT_{j'}}}{\sqrt{(1+vT_j\mu_1)(1+vT_{j'}\mu_1)}}}{(1-\bar{r})(\sum_{j=1}^m \delta_j\frac{T_j\mu_1}{1+vT_j\mu_1})^2} + \frac{\sum_{j=1}^m\sum_{j'=1}^m \frac{\delta_{jj'}\rho_{jj'}\mu_2\sqrt{T_jT_{j'}}}{\sqrt{(1+vT_j\mu_2)(1+vT_{j'}\mu_2)}}}{\bar{r}(\sum_{j=1}^m \delta_j\frac{T_j\mu_2}{1+vT_j\mu_2})^2}. \tag{2.7}$$

It follows that the closed-form sample size formula for the comparison of TAR is:

$$\left[\frac{\sum_{j=1}^m\sum_{j'=1}^m \frac{\delta_{jj'}\rho_{jj'}\mu_1\sqrt{T_jT_{j'}}}{\sqrt{(1+vT_j\mu_1)(1+vT_{j'}\mu_1)}}}{(1-\bar{r})(\sum_{j=1}^m \delta_j\frac{T_j\mu_1}{1+vT_j\mu_1})^2} + \frac{\sum_{j=1}^m\sum_{j'=1}^m \frac{\delta_{jj'}\rho_{jj'}\mu_2\sqrt{T_jT_{j'}}}{\sqrt{(1+vT_j\mu_2)(1+vT_{j'}\mu_2)}}}{\bar{r}(\sum_{j=1}^m \delta_j\frac{T_j\mu_2}{1+vT_j\mu_2})^2}\right]\frac{(z_{1-\alpha/2}+z_{1-\gamma})^2}{\beta_{20}^2}. \tag{2.8}$$

When $T_1 = ... = T_m = T$, Formula (2.8) can be simplified to

$$n = \frac{[(1-\bar{r})T\mu_1 + \bar{r}T\mu_2 + vT^2\mu_1\mu_2](\sum_{j=1}^m\sum_{j'=1}^m \delta_{jj'}\rho_{jj'})(z_{1-\alpha/2}+z_{1-\gamma})^2}{(1-\bar{r})\bar{r}T^2\mu_1\mu_2\beta_{20}^2(\sum_{j=1}^m\delta_j)^2}. \tag{2.9}$$

Formula (2.9) explicitly shows that the sample size is an increasing function of the dispersion parameter v. Here are some useful facts about the sample size formulas (2.8) and (2.9):

- Longitudinal observations are usually positively correlated. In such scenarios, stronger correlation (larger $\rho_{jj'}$) is associated with larger sample size requirement for the comparison of TAR.
- When within-subject correlation is non-negative ($\rho_{jj'} \geq 0$), given the same set of marginal observation probabilities $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_m)$, missing data pattern MM is associated with a larger sample size than IM, because when $j = j'$ we have $\delta_{jj}^{(IM)} = \delta_{jj}^{(MM)} = \delta_j$, and for $j \neq j'$

$$\delta_{jj'}^{(MM)} = \delta_{\max\{j,j'\}} \geq \delta_{jj}^{(IM)} = \delta_j\delta_{j'}.$$

For sample size calculation, we assume that the dispersion parameter v is known. In practice, we need to obtain an estimate of v which we denote as $\hat{v}$. Note that $E(\epsilon_{ij}^2) = E[(y_{ij} - \mu_{ij})^2] = Var(y_{ij}) = \mu_{ij} + v\mu_{ij}^2$, resulting in $E(\epsilon_{ij}^2 - \mu_{ij} - v\mu_{ij}^2) = 0$. We follow the approach of Kong et al. [25], which finds $\hat{v}$ by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \Delta_{ij}(\hat{\epsilon}_{ij}^2 - \mu_{ij}(\hat{\boldsymbol{\beta}}) - \mu_{ij}(\hat{\boldsymbol{\beta}})^2 v)^2,$$

given $\hat{\boldsymbol{\beta}}$. It is easy to verify that

$$\hat{v} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} \Delta_{ij}[\hat{\epsilon}_{ij}^2 - \mu_{ij}(\hat{\boldsymbol{\beta}})]\mu_{ij}(\hat{\boldsymbol{\beta}})^2}{\sum_{i=1}^{n}\sum_{j=1}^{m} \Delta_{ij}\mu_{ij}(\hat{\boldsymbol{\beta}})^4}. \tag{2.10}$$

A complete estimation procedure for $(\boldsymbol{\beta}, v)'$ is listed as follows:

1. Initial values of $\hat{\boldsymbol{\beta}}$ and $\hat{v}$ can be obtained through the maximum likelihood estimation by treating all observations as independent. We denote the initial values as $\hat{\boldsymbol{\beta}}^{(0)}$ and $\hat{v}^{(0)}$.

2. Plug $\hat{\boldsymbol{\beta}}^{(0)}$ and $\hat{v}^{(0)}$ into Equations (2.3)-(2.4) to update $\hat{\boldsymbol{\beta}}$ which we denote as $\hat{\boldsymbol{\beta}}^{(1)}$. Given $\hat{\boldsymbol{\beta}}^{(1)}$, we use Equation (2.10) to update $\hat{v}$, denoted as $\hat{v}^{(1)}$.

3. Repeat Step 2 until $\hat{\boldsymbol{\beta}}$ and $\hat{v}$ converge.

2.2. Statistical Model and Sample Size Approach for Comparing Slopes

For comparing the slopes between two experimental groups, we assume that

$$\log(\mu_{ij}) = \log(T_j) + \beta_1 + \beta_2 r_i + \beta_3 t_j + \beta_4 r_i t_j.$$

Note that β_4 is the difference in slope between the control and treatment groups, representing the treatment effect. We are interested in testing $H_0 : \beta_4 = 0$ vs $H_1 : \beta_4 \neq 0$. Let

$\boldsymbol{Z}_{ij} = (1, r_i, t_j, r_i t_j)'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)'$, hence,

$$\mu_{ij}(\boldsymbol{\beta}) = T_j \exp(\boldsymbol{\beta}' \boldsymbol{Z}_{ij}).$$

The GEE estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)'$ is solved from

$$\boldsymbol{S}_n(\boldsymbol{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m} \Delta_{ij} \frac{1}{1 + v\mu_{ij}(\boldsymbol{\beta})} [y_{ij} - \mu_{ij}(\boldsymbol{\beta})] \boldsymbol{Z}_{ij} = 0, \tag{2.11}$$

similarly based on the Newton-Raphson algorithm as described in Section 2.1. $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ approximately follows a normal distribution with mean $\mathbf{0}$ and variance matrix $\boldsymbol{\Sigma}_n = \boldsymbol{A}_n^{-1} \boldsymbol{V}_n \boldsymbol{A}_n^{-1}$, where

$$\boldsymbol{A}_n(\hat{\boldsymbol{\beta}}) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \Delta_{ij} \boldsymbol{Z}_{ij} \boldsymbol{Z}_{ij}' \frac{\mu_{ij}(\hat{\boldsymbol{\beta}})}{1 + v\mu_{ij}(\hat{\boldsymbol{\beta}})},$$

and

$$\boldsymbol{V}_n(\hat{\boldsymbol{\beta}}) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{j'=1}^{m} \Delta_{ij} \Delta_{ij'} \hat{\epsilon}_{ij} \hat{\epsilon}_{ij'} \frac{1}{1 + v\mu_{ij}(\hat{\boldsymbol{\beta}})} \frac{1}{1 + v\mu_{ij'}(\hat{\boldsymbol{\beta}})} \boldsymbol{Z}_{ij} \boldsymbol{Z}_{ij'}'.$$

Let $\hat{\sigma}_4^2$ be the (4,4)th element of $\boldsymbol{\Sigma}_n$. We reject $H_0 : \beta_4 = 0$ if $|\frac{\sqrt{n}\hat{\beta}_4}{\hat{\sigma}_4}|$ is greater than $z_{1-\alpha/2}$. Let σ_4^2 denote the (4,4)th element of $\boldsymbol{\Sigma}$. Under the true treatment effect $\beta_4 = \beta_{40}$, with a predetermined type I error α and power $1 - \gamma$, the required sample size is

$$n = \frac{\sigma_4^2 (z_{1-\alpha/2} + z_{1-\gamma})^2}{\beta_{40}^2}. \tag{2.12}$$

In order to calculate the sample size, we need to derive the expression of $\boldsymbol{A}$ and $\boldsymbol{V}$. As $n \to \infty$, we have

$$
\begin{aligned}
\boldsymbol{A} &= E\left[\sum_{j=1}^{m}\Delta_{ij}\frac{\mu_{ij}(\boldsymbol{\beta})}{1+v\mu_{ij}(\boldsymbol{\beta})}\begin{pmatrix} 1 & r_i & t_j & r_it_j \\ r_i & r_i^2 & r_it_j & r_i^2t_j \\ t_j & r_it_j & t_j^2 & r_it_j^2 \\ r_it_j & r_i^2t_j & r_it_j^2 & r_i^2t_j^2 \end{pmatrix}\right] \\
&= (1-\bar{r})\sum_{j=1}^{m}\delta_j\frac{T_j\xi_{1j}}{1+vT_j\xi_{1j}}\begin{pmatrix} 1 & 0 & t_j & 0 \\ 0 & 0 & 0 & 0 \\ t_j & 0 & t_j^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \bar{r}\sum_{j=1}^{m}\delta_j\frac{T_j\xi_{2j}}{1+vT_j\xi_{2j}}\begin{pmatrix} 1 & 1 & t_j & t_j \\ 1 & 1 & t_j & t_j \\ t_j & t_j & t_j^2 & t_j^2 \\ t_j & t_j & t_j^2 & t_j^2 \end{pmatrix},
\end{aligned}
$$

where $\xi_{1j} = \exp(\beta_1 + \beta_3 t_j)$ and $\xi_{2j} = \exp[\beta_1 + \beta_2 + (\beta_3 + \beta_4)t_j]$ are the time-specific rates in the control and treatment groups, respectively. Similarly we have

$$
\boldsymbol{V} = E\left[\sum_{j=1}^{m}\sum_{j'=1}^{m}\Delta_{ij}\Delta_{ij'}\epsilon_{ij}\epsilon_{ij'}\frac{1}{1+v\mu_{ij}(\boldsymbol{\beta})}\frac{1}{1+v\mu_{ij'}(\boldsymbol{\beta})}\begin{pmatrix} 1 & r_i & t_{j'} & r_it_{j'} \\ r_i & r_i^2 & r_it_{j'} & r_i^2t_{j'} \\ t_j & r_it_j & t_jt_{j'} & r_it_jt_{j'} \\ r_it_j & r_i^2t_j & r_it_jt_{j'} & r_i^2t_jt_{j'} \end{pmatrix}\right].
$$

It follows that, $\boldsymbol{V} = (1-\bar{r})\boldsymbol{V}_1 + \bar{r}\boldsymbol{V}_2$, where

$$\boldsymbol{V}_1 = \sum_{j=1}^{m}\sum_{j'=1}^{m} \delta_{jj'}\rho_{jj'} \frac{\sqrt{T_jT_{j'}\xi_{1j}\xi_{1j'}}}{\sqrt{(1+vT_j\xi_{1j})(1+vT_{j'}\xi_{1j'})}} \begin{pmatrix} 1 & 0 & t_{j'} & 0 \\ 0 & 0 & 0 & 0 \\ t_j & 0 & t_jt_{j'} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$\boldsymbol{V}_2 = \sum_{j=1}^{m}\sum_{j'=1}^{m} \delta_{jj'}\rho_{jj'} \frac{\sqrt{T_jT_{j'}\xi_{2j}\xi_{2j'}}}{\sqrt{(1+vT_j\xi_{2j})(1+vT_{j'}\xi_{2j'})}} \begin{pmatrix} 1 & 1 & t_{j'} & t_{j'} \\ 1 & 1 & t_{j'} & t_{j'} \\ t_j & t_j & t_jt_{j'} & t_jt_{j'} \\ t_j & t_j & t_jt_{j'} & t_jt_{j'} \end{pmatrix}.$$

With some algebra, we can show that

$$\sigma_4^2 = \frac{\sum_{j=1}^{m}\sum_{j'=1}^{m}\delta_{jj'}\rho_{jj'}\eta_{1j}\eta_{1j'}(t_j-\bar{t}_1)(t_{j'}-\bar{t}_1)}{(1-\bar{r})[\sum_{j=1}^{m}\delta_j\eta_{1j}^2(t_j-\bar{t}_1)^2]^2} + \frac{\sum_{j=1}^{m}\sum_{j'=1}^{m}\delta_{jj'}\rho_{jj'}\eta_{2j}\eta_{2j'}(t_j-\bar{t}_2)(t_{j'}-\bar{t}_2)}{\bar{r}[\sum_{j=1}^{m}\delta_j\eta_{2j}^2(t_j-\bar{t}_2)^2]^2}, \tag{2.13}$$

where $\eta_{1j} = \frac{\sqrt{T_j\xi_{1j}}}{\sqrt{1+vT_j\xi_{1j}}}$, $\eta_{2j} = \frac{\sqrt{T_j\xi_{2j}}}{\sqrt{1+vT_j\xi_{2j}}}$, $\bar{t}_1 = \frac{\sum_{j=1}^{m}\delta_j\eta_{1j}^2t_j}{\sum_{j=1}^{m}\delta_j\eta_{1j}^2}$, and $\bar{t}_2 = \frac{\sum_{j=1}^{m}\delta_j\eta_{2j}^2t_j}{\sum_{j=1}^{m}\delta_j\eta_{2j}^2}$. Therefore, the closed-form sample size formula for comparing slopes is

$$n = \frac{\left[\frac{\sum_{j=1}^{m}\sum_{j'=1}^{m}\delta_{jj'}\rho_{jj'}\eta_{1j}\eta_{1j'}(t_j-\bar{t}_1)(t_{j'}-\bar{t}_1)}{(1-\bar{r})(\sum_{j=1}^{m}\delta_j\eta_{1j}^2(t_j-\bar{t}_1)^2)^2} + \frac{\sum_{j=1}^{m}\sum_{j'=1}^{m}\delta_{jj'}\rho_{jj'}\eta_{2j}\eta_{2j'}(t_j-\bar{t}_2)(t_{j'}-\bar{t}_2)}{\bar{r}(\sum_{j=1}^{m}\delta_j\eta_{2j}^2(t_j-\bar{t}_2)^2)^2}\right](z_{1-\alpha/2}+z_{1-\gamma})^2}{\beta_{40}^2}. \tag{2.14}$$

The derivation of Equation (2.13) is presented in Appendix A.2.

To compute a sample size using Formulas (2.8) or (2.14), we need to specify the number of measurements m, the visit times $\boldsymbol{t}$, the measurement intervals $\boldsymbol{T}$, the randomization scheme $\bar{r}$, the parameter vector $\boldsymbol{\beta}$, the overdispersion parameter v, the correlation structure and the correlation parameter ρ, the observational probability vector $\boldsymbol{\delta}$ and the missing pattern, and pre-determined levels of type I error α and power $1-\gamma$.

2.3. Simulation

We conduct the following simulation study to evaluate the performance of the proposed sample size methods. First we assume that each patient is scheduled to contribute m=5 measurements with visit times $\boldsymbol{t} = (1, ..., 5)$. Without loss of generality, we explore two types of measurement intervals: equal-length $\boldsymbol{T} = (1, 1, 1, 1, 1)$, and unequal-length $\boldsymbol{T} = (0.8, 1, 1, 1, 1.2)$. The nominal levels of type I error and power are set at $\alpha = 0.05$ and $1 - \gamma$ = 0.8, respectively. The marginal observation probabilities are set at $\boldsymbol{\delta} = (1, 0.95, 0.9, 0.85, 0.8)$. That is, we assume no missing data at the initial measurement, and an additional 5% missing at each of the subsequent visits. We set $\bar{r} = 0.5$, i.e., a balanced randomization. We consider two missing data patterns (IM and MM), and two correlation structures, compound symmetry (CS) and AR(1). Under CS, within-subject correlation is constant regardless of time, $\rho_{jj'} = \rho$ for $j \neq j'$. Under AR(1), the correlation is assumed to decay over time, $\rho_{jj'} = \rho^{|j'-j|}$. Three values of the correlation parameter are evaluated: $\rho = 0.1, 0.3, 0.5$. For the comparison of TAR between two groups, we set intercept at $\beta_1 = 0.2$ and explore two levels of true treatment effects β_{20}= 0.25 and 0.3. For the comparison of slopes, we set $(\beta_1, \beta_2, \beta_3) = (0.1, 0, 0.1)$ and explore two treatment effects $\beta_{40} = 0.1$ and 0.15. Finally, we consider two values of the dispersion parameter, v = 0.5 and 1. For each combination of design parameters, we compute sample size, and then conduct simulation to evaluate the empirical power and type I error. The simulation algorithm for a particular design configuration is described as follows:

1. For the comparison of TAR, we calculate sample size n by plugging the set of design parameters into Equation (2.8). For the comparison of slopes, we calculate sample size n by Equation (2.14).

2. We run 5000 iterations. For each iteration:

   (a) Generate a dataset of size n under the the alternative hypothesis ($\beta_2 = \beta_{20}$ or $\beta_4 = \beta_{40}$). Every patient has a vector of correlated count outcomes, $(y_{i1}, \cdots, y_{im})$.

Each outcome marginally follows a negative binomial distribution with parameters (μ_{ij}, v), while jointly they are correlated with assumed correlation structure and ρ. The correlated negative binomial data are generated using the R package "corcounts" [16].

(b) For each subject, we generate missing indicators $(\Delta_{i1}, \cdots, \Delta_{im})$ based on the specified marginal observation probabilities and missing pattern.

(c) For comparing the TAR between groups, we obtain $\hat{\beta}_2$ and $\hat{\sigma_2}$; for comparing the slopes between groups, we obtain $\hat{\beta}_4$ and $\hat{\sigma_4}$.

3. The empirical power is estimated as the proportion of iterations that reject the null hypothesis. That is, the proportion of iterations $|\sqrt{n}\frac{\hat{\beta}_2}{\hat{\sigma_2}}| > z_{1-0.05/2}$ for comparing the TAR, and $|\sqrt{n}\frac{\hat{\beta}_4}{\hat{\sigma_4}}| > z_{1-0.05/2}$ for comparing the slopes.

4. We repeat Steps 2-3 to assess the empirical type I error by setting the value of β_2 or β_4 to 0 in Step 2(a).

Tables $2.1-2.2$ present the sample size, empirical type I error, and empirical power under different combinations of design parameters for comparing TAR, while Tables $2.3-2.4$ are for comparing slopes. Under the combinations of design parameters considered, the sample sizes range from 111 to 691. The empirical powers and type I errors are generally close to the nominal levels, which indicates good performance of the proposed sample size method. Tables $2.1-2.2$ confirm the theoretical properties of sample size for comparing TAR, that the sample size increases as v increases, a stronger correlation is associated with a larger sample size, and that the sample size under MM is greater than that under IM. For comparing slopes, Tables $2.3-2.4$ help us empirically assess the impact of different design parameters. We observe that a larger v is associated with a larger sample size requirement, and that the MM missing pattern results in a larger sample size than IM. It can be analytically shown that under the CS correlation structure, there is a linear relationship between sample size and parameter ρ. The relationship is less clear under the AR(1) structure. In Figure 2.1 we graphically explore the association between

sample size and correlation parameter (ρ). It shows that under the CS correlation structure, the sample size always decreases as ρ increases. Note that the sample size for TAR has an opposite direction of association with ρ. Under the AR(1) correlation structure, the association between sample size and ρ is no longer monotone. As ρ increases, the sample size first increases and then decreases, following a $\bigcap$ shape. The above observation suggests that at the design stage, it is important to have a good understanding about the correlation structure and the strength of correlation, which could be learned from prior experiments.

2.4. Application

We apply the proposed sample size method to an epileptic study [55]. A total of 59 epileptic patients were enrolled, among which 31 patients received the anti-epileptic drug (treatment) and the other 28 patients received the placebo (control). The outcome is the number of epileptic seizure episodes recorded at four time points over two-week intervals. Suppose researchers want to design a new randomized trial where the design parameters are based on the data generated from this study. The research question is to compare the slopes between the treatment versus control group.

We code the visit times as $\boldsymbol{t} = (1, 2, 3, 4)$ with $\boldsymbol{T} = (1, 1, 1, 1)$. Using the GEE approach, the regression coefficients are estimated as $\hat{\boldsymbol{\beta}} = (2.257, 0.006, -0.043, -0.033)'$ and dispersion parameter $\hat{v} = 2.07$. The difference in slope ($\hat{\beta}_4$) is not significantly different from 0, with a p-value of 0.617. Based on the QIC criteria [43], we determine that the CS structure provides a better fit and the correlation parameter (ρ) is estimated to be 0.8059.

Suppose for the new study, the treatment is considered clinically meaningful if the magnitude of β_4 is at least 0.2. Therefore, we assume $\boldsymbol{\beta}' = (2.257, 0.006, -0.043, -0.2)'$ and $v = 2.07$ for sample size calculation. The correlation structure is assumed to be CS with $\rho = 0.8059$. Other design parameters are specified as $\bar{r} = 0.5$ (a balanced random-

ization), two-sided type I error $\alpha = 0.05$, power 0.8, and marginal observation probabilities $\boldsymbol{\delta} = (1, 0.95, 0.9, 0.85)$. The corresponding sample sizes are 98, 103, 69 under IM, MM, and complete data, respectively.

Figure 2.1: Association between sample size and correlation parameter ρ with IM missing pattern, $\boldsymbol{\beta} = (0.1, 0, 0.1, 0.1)$, $\boldsymbol{t} = (0, 1, 2, 3, 4, 5)$, $\boldsymbol{T} = (1, 1, 1, 1, 1)$, $v = 0.5$

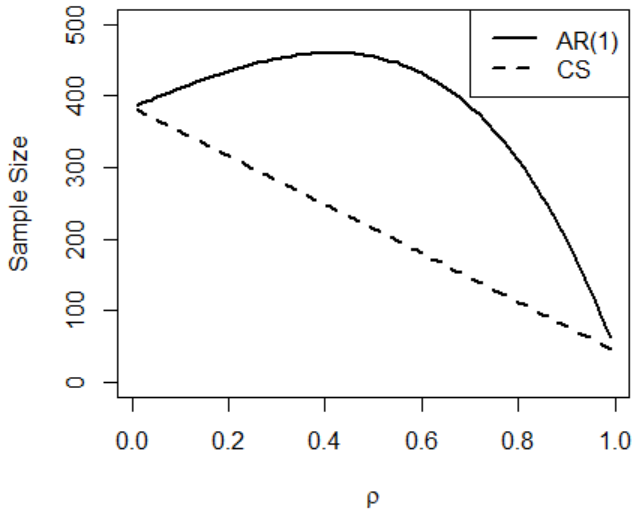

Table 2.1: Sample size (empirical type I error, empirical power) for comparing TAR with $v = 0.5$, $\boldsymbol{T}_1 = (1, 1, 1, 1, 1)$, and $\boldsymbol{T}_2 = (0.8, 1, 1, 1, 1.2)$

			β_2	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$
$\boldsymbol{T}_1$	CS	Complete	0.25	173(0.053,0.797)	272(0.057,0.804)	371(0.057,0.806)
			0.3	119(0.054,0.812)	187(0.056,0.800)	254(0.049,0.799)
		IM	0.25	187(0.053,0.806)	285(0.058,0.806)	384(0.058,0.799)
			0.3	128(0.054,0.799)	196(0.050,0.801)	263(0.057,0.797)
		MM	0.25	189(0.055,0.803)	293(0.051,0.803)	397(0.057,0.803)
			0.3	130(0.054,0.803)	201(0.057,0.801)	272(0.053,0.800)
	AR(1)	Complete	0.25	145(0.053,0.806)	200(0.052,0.802)	275(0.059,0.803)
			0.3	100(0.056,0.803)	137(0.054,0.810)	189(0.054,0.805)
		IM	0.25	159(0.053,0.801)	213(0.054,0.797)	289(0.060,0.808)
			0.3	109(0.054,0.817)	146(0.054,0.815)	198(0.052,0.807)
		MM	0.25	161(0.054,0.822)	219(0.052,0.809)	299(0.055,0.797)
			0.3	110(0.059,0.813)	150(0.053,0.809)	205(0.055,0.802)
$\boldsymbol{T}_2$	CS	Complete	0.25	174(0.053,0.809)	273(0.058,0.804)	372(0.052,0.800)
			0.3	119(0.060,0.806)	187(0.056,0.792)	255(0.060,0.792)
		IM	0.25	189(0.052,0.806)	288(0.053,0.808)	387(0.056,0.799)
			0.3	129(0.054,0.799)	198(0.050,0.806)	266(0.053,0.790)
		MM	0.25	191(0.059,0.808)	296(0.052,0.796)	401(0.051,0.783)
			0.3	131(0.051,0.814)	203(0.057,0.805)	275(0.051,0.797)
	AR(1)	Complete	0.25	146(0.052,0.811)	200(0.055,0.795)	276(0.055,0.795)
			0.3	100(0.060,0.813)	137(0.053,0.803)	189(0.057,0.814)
		IM	0.25	160(0.054,0.800)	215(0.053,0.807)	292(0.051,0.802)
			0.3	110(0.054,0.804)	148(0.053,0.807)	200(0.051,0.803)
		MM	0.25	162(0.052,0.804)	221(0.050,0.806)	302(0.051,0.805)
			0.3	111(0.055,0.809)	152(0.055,0.807)	207(0.058,0.811)

Table 2.2: Sample size (empirical type I error, empirical power) for comparing TAR with $v = 1$, $\boldsymbol{T}_1 = (1, 1, 1, 1, 1)$, and $\boldsymbol{T}_2 = (0.8, 1, 1, 1, 1.2)$

			β_2	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$
$\boldsymbol{T}_1$	CS	Complete	0.25	244(0.048,0.811)	382(0.052,0.785)	521(0.057,0.790)
			0.3	168(0.056,0.804)	263(0.060,0.798)	359(0.048,0.791)
		IM	0.25	263(0.052,0.793)	401(0.053,0.816)	540(0.059,0.799)
			0.3	181(0.053,0.801)	276(0.055,0.790)	372(0.051,0.804)
		MM	0.25	266(0.057,0.799)	412(0.054,0.797)	558(0.055,0.792)
			0.3	183(0.052,0.805)	284(0.052,0.803)	384(0.054,0.804)
	AR(1)	Complete	0.25	204(0.055,0.808)	281(0.058,0.807)	387(0.056,0.806)
			0.3	141(0.058,0.813)	193(0.060,0.808)	266(0.057,0.795)
		IM	0.25	224(0.055,0.807)	300(0.055,0.807)	406(0.051,0.812)
			0.3	154(0.053,0.807)	207(0.053,0.814)	280(0.054,0.813)
		MM	0.25	226(0.052,0.811)	307(0.054,0.814)	420(0.055,0.799)
			0.3	156(0.058,0.809)	212(0.054,0.806)	289(0.055,0.800)
$\boldsymbol{T}_2$	CS	Complete	0.25	244(0.051,0.808)	384(0.055,0.785)	523(0.062,0.793)
			0.3	168(0.052,0.813)	264(0.052,0.796)	360(0.057,0.799)
		IM	0.25	265(0.053,0.809)	405(0.056,0.795)	544(0.057,0.797)
			0.3	182(0.054,0.810)	279(0.058,0.794)	375(0.053,0.799)
		MM	0.25	268(0.056,0.800)	416(0.062,0.800)	563(0.057,0.792)
			0.3	185(0.058,0.805)	286(0.050,0.797)	387(0.057,0.792)
	AR(1)	Complete	0.25	205(0.052,0.805)	282(0.053,0.818)	388(0.055,0.810)
			0.3	141(0.055,0.813)	194(0.053,0.805)	267(0.054,0.801)
		IM	0.25	225(0.063,0.805)	302(0.055,0.807)	410(0.054,0.809)
			0.3	155(0.059,0.813)	208(0.053,0.809)	282(0.054,0.812)
		MM	0.25	228(0.056,0.803)	310(0.051,0.808)	424(0.055,0.820)
			0.3	157(0.055,0.804)	214(0.055,0.820)	292(0.057,0.796)

Table 2.3: Sample size (empirical type I error, empirical power) for comparing slopes with $v = 0.5$, $\boldsymbol{T}_1 = (1, 1, 1, 1, 1)$, and $\boldsymbol{T}_2 = (0.8, 1, 1, 1, 1.2)$

			β_4	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$
$\boldsymbol{T}_1$	CS	Complete	0.1	311(0.053,0.801)	243(0.053,0.804)	175(0.059,0.800)
			0.15	135(0.057,0.805)	105(0.057,0.802)	76(0.062,0.803)
		IM	0.1	350(0.046,0.804)	282(0.058,0.808)	214(0.059,0.808)
			0.15	152(0.054,0.807)	122(0.056,0.810)	93(0.062,0.799)
		MM	0.1	351(0.056,0.806)	285(0.055,0.799)	219(0.059,0.794)
			0.15	152(0.054,0.803)	124(0.061,0.799)	95(0.059,0.799)
	AR(1)	Complete	0.1	371(0.046,0.818)	412(0.050,0.808)	415(0.052,0.803)
			0.15	161(0.050,0.806)	178(0.051,0.818)	180(0.056,0.817)
		IM	0.1	411(0.053,0.811)	452(0.054,0.803)	455(0.054,0.808)
			0.15	178(0.054,0.814)	196(0.054,0.813)	197(0.049,0.815)
		MM	0.1	413(0.056,0.806)	458(0.047,0.807)	464(0.052,0.813)
			0.15	179(0.056,0.799)	198(0.058,0.799)	201(0.057,0.809)
$\boldsymbol{T}_2$	CS	Complete	0.1	319(0.050,0.808)	251(0.051,0.800)	183(0.060,0.799)
			0.15	138(0.053,0.809)	109(0.059,0.811)	80(0.062,0.800)
		IM	0.1	359(0.051,0.805)	291(0.053,0.794)	223(0.060,0.804)
			0.15	156(0.061,0.814)	126(0.064,0.812)	97(0.058,0.814)
		MM	0.1	360(0.050,0.794)	293(0.052,0.801)	226(0.055,0.793)
			0.15	156(0.058,0.795)	127(0.059,0.802)	98(0.058,0.808)
	AR(1)	Complete	0.1	380(0.054,0.803)	423(0.052,0.809)	426(0.053,0.795)
			0.15	165(0.053,0.806)	183(0.060,0.802)	185(0.057,0.808)
		IM	0.1	421(0.049,0.806)	464(0.054,0.808)	468(0.052,0.806)
			0.15	183(0.053,0.798)	201(0.056,0.813)	203(0.056,0.806)
		MM	0.1	423(0.050,0.804)	469(0.052,0.807)	476(0.050,0.802)
			0.15	183(0.059,0.815)	203(0.055,0.809)	206(0.058,0.809)

Table 2.4: Sample size (empirical type I error, empirical power) for comparing slopes with $v = 1$, $\boldsymbol{T}_1 = (1, 1, 1, 1, 1)$, and $\boldsymbol{T}_2 = (0.8, 1, 1, 1, 1.2)$

			β_4	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$
$\boldsymbol{T}_1$	CS	Complete	0.1	453(0.051,0.805)	353(0.061,0.802)	253(0.060,0.799)
			0.15	198(0.056,0.804)	154(0.059,0.805)	111(0.061,0.800)
		IM	0.1	511(0.052,0.801)	411(0.056,0.793)	311(0.061,0.799)
			0.15	223(0.049,0.805)	180(0.055,0.808)	136(0.064,0.810)
		MM	0.1	513(0.051,0.814)	416(0.056,0.802)	319(0.058,0.791)
			0.15	224(0.051,0.798)	182(0.056,0.803)	140(0.054,0.791)
	AR(1)	Complete	0.1	542(0.052,0.810)	601(0.046,0.809)	604(0.053,0.792)
			0.15	237(0.051,0.809)	262(0.048,0.805)	264(0.054,0.811)
		IM	0.1	600(0.049,0.815)	659(0.055,0.811)	662(0.051,0.802)
			0.15	262(0.053,0.808)	288(0.052,0.804)	289(0.051,0.802)
		MM	0.1	603(0.052,0.805)	668(0.055,0.810)	676(0.048,0.804)
			0.15	263(0.052,0.807)	292(0.049,0.806)	295(0.055,0.800)
$\boldsymbol{T}_2$	CS	Complete	0.1	462(0.054,0.814)	362(0.056,0.795)	261(0.056,0.799)
			0.15	202(0.055,0.816)	158(0.059,0.809)	114(0.063,0.802)
		IM	0.1	522(0.052,0.809)	421(0.052,0.799)	320(0.067,0.800)
			0.15	228(0.057,0.803)	184(0.061,0.806)	140(0.065,0.804)
		MM	0.1	523(0.058,0.797)	425(0.055,0.807)	326(0.054,0.807)
			0.15	229(0.054,0.808)	186(0.057,0.799)	143(0.056,0.808)
	AR(1)	Complete	0.1	553(0.052,0.800)	614(0.050,0.804)	618(0.055,0.801)
			0.15	242(0.058,0.800)	268(0.055,0.804)	270(0.055,0.813)
		IM	0.1	613(0.049,0.810)	674(0.052,0.805)	678(0.050,0.803)
			0.15	268(0.055,0.810)	295(0.053,0.819)	297(0.050,0.811)
		MM	0.1	616(0.054,0.801)	683(0.053,0.804)	691(0.052,0.804)
			0.15	269(0.058,0.808)	298(0.056,0.813)	302(0.053,0.806)

CHAPTER 3
INCORPORATING PRAGMATIC FEATURES INTO SAMPLE SIZE CALCULATION FOR CLUSTERED RANDOMIZATION TRIALS WITH A COUNT OUTCOME

In this chapter, we propose a sample size calculation method for CRT with a count outcome, which is capable of handling pragmatic issues such as over-dispersion and varying cluster size.

3.1. Statistical Model and Sample Size Approach

Suppose in a CRT, N clusters are randomized to a control or a treatment arm. In the following, we use $\mathcal{C}$ and $\mathcal{T}$ to denote the set of clusters assigned to the control and treatment arm, respectively. We further use $m_i\,(i = 1, ..., N)$ to denote the cluster size and m_i is assumed to independently follow a certain discrete distribution $Prob(m_i = m) = g(m)$ with outcome space $\mathcal{M}$. We further define mean $\eta_m = E(m_i)$ and variance $\tau^2 = Var(m_i)$. The primary outcome, denoted by y_{ij} , is the count of a certain event (e.g. infections, exacerbation episodes, hospital visits, etc.) for the jth patient in the ith cluster during a follow-up period of length t_{ij} . We assume that there is a common planned follow-up time (denoted by t^*) for every patient. However, each patient has a probability to drop out from the study. The time to drop out from the study, which we denote as d_{ij} , is independent and identically distributed with a certain continuous distribution. Hence the true follow-up time is $t_{ij} = min(t^*, d_{ij})$, with $f(t_{ij})$ as the density function. We model y_{ij} by a quasi Poisson distribution [61], where the first two moments are defined by $E(y_{ij}) = \mu_{ij}$ and $Var(y_{ij}) = \theta\mu_{ij}$. Here $\theta \geq 1$ is the over-dispersion parameter. The mean parameter μ_{ij} is modeled by

$$\log(\mu_{ij}) = \log(t_{ij}) + \beta_1 + \beta_2 r_i. \tag{3.1}$$

In the model, $r_i = 1/0$ indicates that cluster i is assigned to the treatment/control arm, $\log(t_{ij})$ is the offset to account for variable lengths of follow-up, β_1 is the event rate under control on the log scale, and β_2 models the difference between the treatment and control arms, representing the treatment effect. A cluster is assigned to the treatment arm with probability $\bar{r} = E(r_i)$. Intraclass correlation (ICC) is denoted by $\rho = \text{Corr}(y_{ij}, y_{ij'})$ for $j \neq j'$ and observations are assumed to be independent across clusters. The primary interest is to test $H_0 : \beta_2 = 0$ vs $H_1 : \beta_2 \neq 0$. Let $\boldsymbol{Z}_{ij} = (1, r_i)' = \boldsymbol{Z}_i$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, we have

$$\mu_{ij}(\boldsymbol{\beta}) = t_{ij} \exp(\boldsymbol{Z}_{ij}'\boldsymbol{\beta}) = t_{ij} \exp(\boldsymbol{Z}_i'\boldsymbol{\beta}).$$

Let $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{im_i})'$ be the response vector of the ith cluster and similarly define $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), \cdots, \mu_{im_i}(\boldsymbol{\beta}))'$. Utilizing an independent working correlation structure, the GEE estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)'$ is obtained from

$$S_N(\boldsymbol{\beta}) = N^{-\frac{1}{2}} \sum_{i=1}^{N} \boldsymbol{D}_i' \boldsymbol{W}_i^{-1} [\boldsymbol{y}_i - \boldsymbol{\mu}_{\boldsymbol{i}}(\boldsymbol{\beta})] = \boldsymbol{0}, \tag{3.2}$$

where $\boldsymbol{D}_i$ is a $m_i \times 2$ gradient matrix defined as $\boldsymbol{D}_i = \frac{\partial \boldsymbol{\mu}_{\boldsymbol{i}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ and $\boldsymbol{W}_i = \theta \boldsymbol{S}_i^{0.5} \boldsymbol{I}_{m_i} \boldsymbol{S}_i^{0.5}$. Here $\boldsymbol{S}_i = \text{diag}[\boldsymbol{\mu}_i(\boldsymbol{\beta})]$ is a $m_i \times m_i$ diagonal matrix and $\boldsymbol{I}_{m_i}$ is the $m_i \times m_i$ identity matrix. With some algebra, we can simplify (3.2) to

$$S_N(\boldsymbol{\beta}) = N^{-\frac{1}{2}} \theta^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} [y_{ij} - \mu_{ij}(\boldsymbol{\beta})] \boldsymbol{Z}_i = 0. \tag{3.3}$$

Solving Equation (3.3), we obtain the GEE estimator:

$$\begin{aligned} \hat{\beta}_1 &= \log\left(\frac{\sum_{i \in \mathcal{C}} \sum_{j=1}^{m_i} y_{ij}}{\sum_{i \in \mathcal{C}} \sum_{j=1}^{m_i} t_{ij}} \right), \\ \hat{\beta}_2 &= \log\left(\frac{\sum_{i \in \mathcal{T}} \sum_{j=1}^{m_i} y_{ij}}{\sum_{i \in \mathcal{T}} \sum_{j=1}^{m_i} t_{ij}} \right) - \log\left(\frac{\sum_{i \in \mathcal{C}} \sum_{j=1}^{m_i} y_{ij}}{\sum_{i \in \mathcal{C}} \sum_{j=1}^{m_i} t_{ij}} \right). \end{aligned} \tag{3.4}$$

$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ approximately follows a normal distribution with $\mathbf{0}$ mean, and the variance matrix can be estimated by $\boldsymbol{\Sigma}_n = \boldsymbol{A}_n^{-1}\boldsymbol{V}_n\boldsymbol{A}_n^{-1}$, with

$$\boldsymbol{A}_N(\hat{\boldsymbol{\beta}}) = N^{-1}\theta^{-1}\sum_{i=1}^{N}\sum_{j=1}^{m_i}\mu_{ij}(\hat{\boldsymbol{\beta}})\boldsymbol{Z}_i\boldsymbol{Z}_i',$$

and

$$\boldsymbol{V}_N(\hat{\boldsymbol{\beta}}) = N^{-1}\theta^{-2}\sum_{i=1}^{N}\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i}\hat{\epsilon}_{ij}\hat{\epsilon}_{ij'}\boldsymbol{Z}_i\boldsymbol{Z}_i'.$$

where $\hat{\epsilon}_{ij} = y_{ij} - \mu_{ij}(\hat{\boldsymbol{\beta}})$ denotes the residual. It is easy to show that the over-dispersion parameter θ in $\boldsymbol{A}_n(\hat{\boldsymbol{\beta}})$ and $\boldsymbol{V}_n(\hat{\boldsymbol{\beta}})$ are cancelled when calculating $\boldsymbol{\Sigma}_n$, hence we can rewrite $\boldsymbol{A}_n(\hat{\boldsymbol{\beta}})$ and $\boldsymbol{V}_n(\hat{\boldsymbol{\beta}})$ as

$$\begin{aligned}\boldsymbol{A}_N(\hat{\boldsymbol{\beta}}) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\mu_{ij}(\hat{\boldsymbol{\beta}})\boldsymbol{Z}_i\boldsymbol{Z}_i', \\ \boldsymbol{V}_N(\hat{\boldsymbol{\beta}}) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j'=1}^{m}\hat{\epsilon}_{ij}\hat{\epsilon}_{ij'}\boldsymbol{Z}_i\boldsymbol{Z}_{i'}'.\end{aligned} \tag{3.5}$$

Let $\hat{\sigma}_2^2$ be the (2,2)th element of $\boldsymbol{\Sigma}_n$. With straightforward matrix algebra, it can be shown that

$$\hat{\sigma}_2^2 = \frac{\sum_{i\in\mathcal{C}}\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i}\hat{\epsilon}_{ij}\hat{\epsilon}_{ij'}}{[\sum_{i\in\mathcal{C}}\sum_{j=1}^{m_i}\mu_{ij}(\hat{\boldsymbol{\beta}})]^2} + \frac{\sum_{i\in\mathcal{T}}\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i}\hat{\epsilon}_{ij}\hat{\epsilon}_{ij'}}{[\sum_{i\in\mathcal{T}}\sum_{j=1}^{m_i}\mu_{ij}(\hat{\boldsymbol{\beta}})]^2}. \tag{3.6}$$

Furthermore, define $y_{i\cdot} = \sum_{j=1}^{m_i} y_{ij}$, $t_{i\cdot} = \sum_{j=1}^{m_i} t_{ij}$ and $\mu_{i\cdot}(\hat{\boldsymbol{\beta}}) = \sum_{j=1}^{m_i}\mu_{ij}(\hat{\boldsymbol{\beta}})$ as the cluster-level aggregated data, then estimators (3.4) and (3.6) can be re-written as

$$\begin{aligned}\hat{\beta}_1 &= \log\left(\frac{\sum_{i\in\mathcal{C}} y_{i\cdot}}{\sum_{i\in\mathcal{C}} t_{i\cdot}}\right), \\ \hat{\beta}_2 &= \log\left(\frac{\sum_{i\in\mathcal{T}} y_{i\cdot}}{\sum_{i\in\mathcal{T}} t_{i\cdot}}\right) - \log\left(\frac{\sum_{i\in\mathcal{C}} y_{i\cdot}}{\sum_{i\in\mathcal{C}} t_{i\cdot}}\right),\end{aligned} \tag{3.7}$$

and

$$\hat{\sigma}_2^2 = \frac{\sum_{i\in\mathcal{C}}[y_{i\cdot} - \mu_{i\cdot}(\hat{\boldsymbol{\beta}})]^2}{[\sum_{i\in\mathcal{C}} \mu_{i\cdot}(\hat{\boldsymbol{\beta}})]^2} + \frac{\sum_{i\in\mathcal{T}}[y_{i\cdot} - \mu_{i\cdot}(\hat{\boldsymbol{\beta}})]^2}{[\sum_{i\in\mathcal{T}} \mu_{i\cdot}(\hat{\boldsymbol{\beta}})]^2}. \tag{3.8}$$

It is noteworthy that Equations (3.4)-(3.8) do not include the over-dispersion parameter θ or the correlation parameter ρ. Furthermore, Equations (3.7)-(3.8) suggest that although the model (3.1) is specified at the patient level, the treatment effect estimate $\hat{\beta}_2$ and its variance $\hat{\sigma}_2^2$ can be obtained using cluster-level aggregated data.

We reject $H_0 : \beta_2 = 0$ if $|\frac{\sqrt{N}\hat{\beta}_2}{\hat{\sigma}_2}|$ is greater than $z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the standard normal distribution.

To derive the sample size formula, first we define $\boldsymbol{A}$ and $\boldsymbol{V}$ to be the limits of $\boldsymbol{A}_N$ and $\boldsymbol{V}_N$ as $N \to \infty$, then $\boldsymbol{\Sigma}_N$ converges to $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}$. Let σ_2^2 denote the (2,2)th element of $\boldsymbol{\Sigma}$. Then, given $H_1 : \beta_2 = \beta_{20}$, with specified two-sided type I error α and power $1-\gamma$, the required number of clusters is calculated by

$$N = \frac{\sigma_2^2(z_{1-\alpha/2} + z_{1-\gamma})^2}{\beta_{20}^2}. \tag{3.9}$$

In the following, we show that a closed-form expression of σ_2^2 can be derived, even after we take into account pragmatic features of over-dispersion, unbalance randomization, as well as random variability in cluster size and length of follow-up. To facilitate the derivation, we denote $\gamma_t = E(t_{ij})$, $\epsilon_t^2 = Var(t_{ij})$, and $\kappa_t = E(t_{ij}^{1/2})$ as the mean, variance, and expected value of square root under the density function $f(t_{ij})$ assumed for length of follow-up. First we have

$$\boldsymbol{A}_N(\hat{\boldsymbol{\beta}}) = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{m_i}\mu_{ij}(\boldsymbol{\beta})\begin{pmatrix} 1 & r_i \\ r_i & r_i^2 \end{pmatrix} + o_p(1).$$

As $N \to \infty$, $\boldsymbol{A}_N(\hat{\boldsymbol{\beta}})$ approaches

$$
\begin{aligned}
\boldsymbol{A} &= E\left[\sum_{j=1}^{m_i}\mu_{ij}(\boldsymbol{\beta})\begin{pmatrix} 1 & r_i \\ r_i & r_i^2 \end{pmatrix}\right] \qquad (3.10)\\
&= (1-\bar{r})\sum_{m\in\mathcal{M}}\mu_1\gamma_t m g(m)\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \bar{r}\sum_{m\in\mathcal{M}}\mu_2\gamma_t m g(m)\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\\
&= (1-\bar{r})\mu_1\gamma_t\eta_m\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \bar{r}\mu_2\gamma_t\eta_m\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\\
&= \begin{pmatrix} (1-\bar{r})\mu_1\gamma_t\eta_m + \bar{r}\mu_2\gamma_t\eta_m & \bar{r}\mu_2\gamma_t\eta_m \\ \bar{r}\mu_2\gamma_t\eta_m & \bar{r}\mu_2\gamma_t\eta_m \end{pmatrix},
\end{aligned}
$$

where $\mu_1 = \exp(\beta_1)$ and $\mu_2 = \exp(\beta_1+\beta_2)$. For $\boldsymbol{V}_N(\hat{\boldsymbol{\beta}})$, we have

$$
\boldsymbol{V}_n(\hat{\boldsymbol{\beta}}) = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i}[y_{ij}-\mu_{ij}(\boldsymbol{\beta})][y_{ij'}-\mu_{ij'}(\boldsymbol{\beta})]\begin{pmatrix} 1 & r_i \\ r_i & r_i^2 \end{pmatrix} + o_p(1).
$$

As $N\to\infty$, $\boldsymbol{V}_N(\hat{\boldsymbol{\beta}})$ approaches

$$
\begin{aligned}
\boldsymbol{V} &= E\left[\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i}(y_{ij}-\mu_{ij}(\boldsymbol{\beta}))(y_{ij'}-\mu_{ij'}(\boldsymbol{\beta}))\begin{pmatrix} 1 & r_i \\ r_i & r_i^2 \end{pmatrix}\right]. \qquad (3.11)\\
&= (1-\bar{r})\boldsymbol{V}_1 + \bar{r}\boldsymbol{V}_2,
\end{aligned}
$$

where

$$\boldsymbol{V}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} [\eta_m(\theta\mu_1\gamma_t + \mu_1^2\epsilon_t^2) + (\tau_m^2 + \eta_m^2 - \eta_m)\rho\theta\mu_1\kappa_t^2], \tag{3.12}$$

and

$$\boldsymbol{V}_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} [\eta_m(\theta\mu_2\gamma_t + \mu_1^2\epsilon_t^2) + (\tau_m^2 + \eta_m^2 - \eta_m)\rho\theta\mu_2\kappa_t^2]. \tag{3.13}$$

More details of the derivation for $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ can be found in Appendix B.1.

Using matrix algebra, we can obtain the (2,2)th element of $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}$:

$$\sigma_2^2 = \frac{\eta_m\theta\gamma_t[(1-\bar{r})\mu_1 + \bar{r}\mu_2] + \eta_m\epsilon_t^2\mu_1\mu_2 + (\tau_m^2 + \eta_m^2 - \eta_m)\rho\theta\kappa_t^2[(1-\bar{r})\mu_1 + \bar{r}\mu_2]}{(1-\bar{r})\bar{r}\mu_1\mu_2\eta_m^2\gamma_t^2}. \tag{3.14}$$

Plug (3.14) into Equation (3.9) and we obtain the closed-form sample size formula

$$N = \frac{\{\eta_m\theta\gamma_t[(1-\bar{r})\mu_1 + \bar{r}\mu_2] + \eta_m\epsilon_t^2\mu_1\mu_2 + (\tau_m^2 + \eta_m^2 - \eta_m)\rho\theta\kappa_t^2[(1-\bar{r})\mu_1 + \bar{r}\mu_2]\}(z_{1-\alpha/2} + z_{1-\gamma})^2}{(1-\bar{r})\bar{r}\mu_1\mu_2\eta_m^2\gamma_t^2\beta_{20}^2}. \tag{3.15}$$

The pragmatic features accommodated by sample size formula (3.15) include: unbalanced randomization (through parameter $\bar{r}$), over dispersion (θ), varying cluster size (η_m and τ_m^2), and varying length of follow-up (γ_t, ϵ_t^2, κ_t). To account for random cluster size only the first two moments are required, while to account for random length of follow-up in addition the expected value of $\sqrt{t_{ij}}$ is needed. In practice, it would be easier to supply such information based on prior data than fully specifying the distributions of cluster size

and length of follow-up.

In the following, we summarize the impact of different design parameters on sample size (3.15):

- The sample size is an increasing linear function of ICC ρ (because $m \geq 1$ and $\mu_1 > 0$ and $\mu_2 > 0$).
- In practice we usually have ICC $\rho > 0$, and the sample size is an increasing linear function of over-dispersion θ.
- Given the same averaged cluster size (η_m), larger variability (measured by variance τ_m^2) is associated with a larger sample size.
- Given the same averaged length of follow-up (γ_t), larger values of ϵ_t^2 and κ_t^2 are associated with a larger sample size.
- For individual randomized trial where $\rho = 0$, varying lengths of follow-up affect sample size through the first two moments: γ_t and ϵ_t^2.

If there is no variability in cluster size (i.e., $\tau_m^2 = 0$ and $m_i = \eta_m$), the sample size formula (3.15) can be simplified to

$$N_m = \frac{\{\eta_m\theta\gamma_t[(1-\bar{r})\mu_1 + \bar{r}\mu_2] + \eta_m\epsilon_t^2\mu_1\mu_2 + (\eta_m^2 - \eta_m)\rho\theta\kappa_t^2[(1-\bar{r})\mu_1 + \bar{r}\mu_2]\}(z_{1-\alpha/2} + z_{1-\gamma})^2}{(1-\bar{r})\bar{r}\mu_1\mu_2\eta_m^2\gamma_t^2\beta_{20}^2}. \tag{3.16}$$

It is obvious that $N > N_m$ for $\tau_m^2 > 0$. That is, the variability in cluster size leads to larger sample size. The magnitude of increase, however, is jointly determined by all the other design parameters.

On the other hand, if we ignore drop-out (i.e., $t_{ij} = t^* \; \forall \; i, j$), then $\epsilon_t^2 = 0$ and $\kappa_t^2 = t^*$, and the sample size is simplified to

$$N_t = \frac{\theta[(1-\bar{r})\mu_1 + \bar{r}\mu_2](z_{1-\alpha/2} + z_{1-\gamma})^2}{(1-\bar{r})\bar{r}\mu_1\mu_2 t^*\beta_{20}^2} \cdot \left\{\frac{1-\rho}{\eta_m} + \left(1 + \frac{\tau_m^2}{\eta_m^2}\right)\rho\right\}. \tag{3.17}$$

N_t is proportional to $1/t^*$ and θ, and it can be shown that $N_t < N$ (see Appendix B.2). Furthermore, the second term is identical to the correction term derived by Manatunga et al. [37] for conventional cluster randomization trials with varying cluster sizes where τ_m/η_m is the coefficient of variation.

Finally, if there is no variability in cluster size or length of follow-up, the sample size can be further simplified to

$$N_{mt} = \frac{\theta[1+(\eta_m-1)\rho][(1-\bar{r})\mu_1+\bar{r}\mu_2](z_{1-\alpha/2}+z_{1-\gamma})^2}{(1-\bar{r})\bar{r}\mu_1\mu_2\eta_m t^* \beta_{20}^2}. \tag{3.18}$$

It is easy to verify that $N_{mt} < N_m < N$ and $N_{mt} < N_t < N$.

In CRTs, the number of clusters is usually limited. For example, Ivers et al. [23] reviewed a random selection of 285 CRTs and reported that the median number of clusters was 21. When N is small, the sandwich-type variance estimator (3.8) is known to be biased downwards, leading to an inflated type I error [32]. Alternatively, σ_2^2 can be estimated using re-sampling based methods [14]. Sherman & Cessie [51] showed that for clustered data, confidence intervals of the parameters built based on the bootstrap method are superior to the normal confidence intervals built upon the sandwich estimator of variances. In the context of stepped-wedge trials, a special case of CRT, Hussey & Hughes [22] showed that a jackknife estimate of the variance helps to control the size of the test when the number of clusters is limited. Define $\hat{\sigma}_2^{2(Jack)}$ to be the Jackknife estimate of σ_2^2,

$$\hat{\sigma}_2^{2(Jack)} = \frac{N-1}{N}\sum_{i=1}^{N}(\hat{\beta}_2^{(-i)} - \hat{\beta}_2)^2.$$

Here $\hat{\beta}_2^{(-i)}$ denotes the estimate of β_2 based on partial data where the ith cluster is excluded. We perform the re-sampling step on the cluster level instead of the patient level, which retains the within-cluster correlation structure [51]. Using the closed-form solution

for $\hat{\beta}_2$ in Equation (3.7), with some algebra, we are able to show that

$$\hat{\sigma}_2^{2(Jack)} = \frac{N-1}{N}\left\{\sum_{i\in\mathcal{C}}\left[\log\left(1-\frac{y_{i.}}{y_C}\right)-\log\left(1-\frac{t_{i.}}{t_C}\right)\right]^2+\sum_{i\in\mathcal{T}}\left[\log\left(1-\frac{y_{i.}}{y_T}\right)-\log\left(1-\frac{t_{i.}}{t_T}\right)\right]^2\right\}, \tag{3.19}$$

where $y_C = \sum_{i\in\mathcal{C}} y_{i.}$ and $y_T = \sum_{i\in\mathcal{T}} y_{i.}$. For the rest of the paper, we denote the approach using Equation (3.8) to estimate σ_2^2 as "GEE-Naive", and the approach using Equation (3.19) as "GEE-Jackknife". The performance of both approaches will be evaluated by simulation study.

3.2. Simulation

In this section, we conduct simulations to assess the performance of the proposed sample size method. Suppose N clusters are randomized with 1:1 ratio ($\bar{r} = 0.5$) to the control and treatment arms. We consider three distributions for the cluster size m_i: (1) A truncated Poisson distribution with a mean parameter $\lambda = 45$ over a range of $[20, 70]$. We denote this distribution as TrunPoisson with corresponding $\eta_m \approx 45$ and $\tau_m^2 \approx 44.8$; (2) A discrete uniform distribution (DU) with a lower bound 34 and an upper bound 56. The corresponding parameters are $\eta_m = 45$ and variance $\tau_m^2 = 44$; (3) A DU distribution with a lower bound 10 and an upper bound 80. Its corresponding mean is $\eta_m = 45$ and variance is $\tau_m^2 = 420$. For the length of follow-up, we assume that the count outcome is supposed to be measured over a follow-up period of length 1. During the trial, however, patients are likely to drop out, and the potential dropout time (d_{ij}) follows an exponential distribution with rate $\lambda_d = 0.356$. Hence the actual follow-up time is $t_{ij} = \min(1, d_{ij})$. Under the above specification, the probability of early dropout is roughly 30% with $\kappa_t \approx 0.893$, $\gamma_t \approx 0.842$, and $\epsilon_t^2 \approx 0.084$. We explore three values for the over-dispersion parameter θ: 2, 2.5, and 3. Similarly, three values of ICC are explored: $\rho = 0.02, 0.04$, and 0.06, which represent small ICCs commonly reported in CRTs [39]. The nominal levels of type I error and power are set at $\alpha = 0.05$ and $1-\gamma$ = 0.8, respectively. We set the

true value of regression parameter $\beta_1 = 0.6$ and exploring three values of β_2 under the alternative hypothesis: $\beta_{20} = -0.35, -0.38, -0.4$. Given a particular combination of design parameters, the simulation scheme is as follows:

1. Plug the design parameters into Equation (3.15) to compute sample size N.
2. For each scenario, we run $L = 5000$ iterations. In the lth iteration,
   (a) Generate a random dataset of cluster size N under the alternative hypothesis ($\beta_2 = \beta_{20}$). For each cluster, we first generate cluster size m_i from the corresponding TrunPoisson or DU distribution. An m_i-length vector of follow-up times $\boldsymbol{t}_i = (t_{i1}, ..., t_{im_i})$ is generated according to the assumed mechanism of dropout. Randomize the N cluster into control and treatment arms, which determines the mean vector $\boldsymbol{\mu}_i$. An m_i-length vector of correlated count response $\boldsymbol{y}_i = (y_{i1}, ..., y_{im_i})$ is generated using the lognormal-Poisson approach [36] given over-dispersion parameter θ and ICC ρ.
   (b) Based on the generated dataset, we obtain $\hat{\beta}_2$ using Equation (3.7), $\hat{\sigma}_2^{2(Naive)}$ using Equation (3.8) and $\hat{\sigma}_2^{2(Jack)}$ using Equation (3.19).
3. Empirical power of the "GEE-Naive" approach is computed as the proportion of iterations where the null hypothesis is rejected, $|\sqrt{N}\frac{\hat{\beta}_2}{\hat{\sigma}_2^{(Naive)}}| > z_{1-0.05/2}$. The empirical power of the "GEE-Jackknife" approach is computed similarly.
4. Empirical type I errors of the two approaches are computed by the same procedure of Steps 2 and 3, except for setting $\beta_2 = 0$ in Step 2(a).

Tables 3.1-3.3 present the number of clusters, empirical type I error, and empirical power for the "GEE-Naive" approach and the "GEE-Jackknife" approach under different combinations of design parameters. Across all the scenarios considered, the numbers of clusters N range from 14 to 56. As expected, the empirical type I error and power of the "GEE-Naive" approach tend to be larger than the nominal levels due to under-estimated

variances. The "GEE-Jackknife" approach performs better in terms of maintaining the empirical type I error and power at the nominal level under small sample sizes (e.g., $N < 20$). The two approaches start to yield comparable results when the sample sizes are relatively large (e.g., $N > 45$). Since in practice researchers frequently encounter relatively small numbers of clusters in CRTs [23], the "GEE-Jackknife" approach can be very useful. In Tables 3.1 and 3.2 we explore two different distributions of varying cluster size, one being truncated Poisson and the other being discrete uniform distribution, but with comparable means and variances. The two tables show similar performance in preserving the type I error and power close to the nominal level. It shows that the proposed sample size method is robust to accommodate randomly varying cluster sizes following different distributions. On the other hand, Tables 3.2 and 3.3 assume cluster sizes to follow the DU distribution with the same mean but different variances, and the resulting sample sizes are quite different.

In Tables 3.4 and 3.5 we present the sample size under the similar configurations as in Tables 3.2 and 3.3, respectively, with certain pragmatic features ignored: N_m (ignoring the variation in cluster size), N_t (ignoring the variation in length of follow-up), and N_{mt} (ignoring the variation in both cluster size and length of follow-up). The results suggest that ignoring such pragmatic features can lead to severe underestimation (up to 26.67%) of sample size under the design configurations considered. We also observe that the differences between N and N_m in Table 3.4 are smaller than that in Table 3.5, because the variation in cluster size is smaller under DU(22,38) than that under DU(10, 80).

3.3. Application

We apply the proposed sample size approach on the CRT presented in Amatya et al. [3], which evaluated the effectiveness of an educational intervention aiming at improving the management of lung disease in adults attending South African primary-care clinics. The planned follow-up time for all patients was 3 months and the primary endpoint was

the number of clinic visits. Forty clinics were included as clusters and randomized into treatment or control groups. For each of the clinics, the goal of enrollment was to recruit 50 patients. Without loss of generality, let us standardize the follow up time, and set $t^* = 1$. The analysis in Amatya et al. [3] reported $\beta_1 = 1.47$, $\beta_2 = -0.18$, and ICC $\rho = 0.32$. Over-dispersion was not evaluated because a marginal poisson regression model was employed.

Suppose we want to design a new CRT. The desired power and two-sided type I error are set at 80% and 5%, respectively. We set $\beta_1 = 1.47$ which corresponds to a mean of $\exp(1.47) \approx 4.35$ clinical visits under control. Suppose researchers consider the new intervention to be clinically meaningful if the mean is reduced by at least 30%, which corresponds to $\beta_2 = -0.36$. We further assume that the new trial adopts a balanced design ($\bar{r} = 0.5$) and the ICC is $\rho = 0.32$. The mean cluster size η_m is expected to be 50. In addition, we assume each patient may withdraw from the study, and the withdraw time follows an exponential distribution with $\lambda_d = 0.356$. Under the above specification, if the variation in cluster size and the over-dispersion parameter are relatively small, say m_i following DU(40, 60) and $\theta = 1.5$, then the required number of clusters equals to 40. If we ignore pragmatic features like varying cluster size and patients' dropout, the calculated number of clusters becomes 34, resulting in a 15% under-estimation of sample size. If the variation in cluster size and the over-dispersion parameter are relatively large, say m_i following DU(20, 80) and $\theta = 3$, then we have $N = 87$. Ignoring the variation in cluster size and patients' early dropout leads to a sample size of 68, which is a 21.84% under-estimation.

Table 3.1: Sample size (empirical type I error of "GEE-Naive",empirical type I error of "GEE-Jackknife", empirical power of "GEE-Naive", empirical power of 'GEE-Jackknife") assuming cluster size m_i follows the TrunPoi(45,20,70) distribution

	β_2	$\rho = 0.02$	$\rho = 0.04$	$\rho = 0.06$
$\theta = 2$	-0.35	18(0.094,0.058,0.856,0.813)	25(0.077,0.056,0.829,0.791)	33(0.068,0.055,0.834,0.805)
	-0.38	15(0.098,0.062,0.850,0.787)	22(0.076,0.057,0.844,0.801)	29(0.077,0.060,0.841,0.802)
	-0.4	14(0.100,0.062,0.852,0.790)	20(0.086,0.063,0.837,0.787)	26(0.068,0.053,0.836,0.800)
$\theta = 2.5$	-0.35	22(0.086,0.057,0.844,0.800)	32(0.070,0.053,0.835,0.802)	41(0.070,0.057,0.823,0.797)
	-0.38	19(0.077,0.055,0.854,0.807)	27(0.077,0.058,0.831,0.796)	36(0.066,0.054,0.829,0.797)
	-0.4	17(0.089,0.059,0.849,0.788)	25(0.069,0.052,0.834,0.791)	33(0.071,0.059,0.842,0.807)
$\theta = 3$	-0.35	26(0.081,0.058,0.836,0.801)	38(0.065,0.052,0.841,0.815)	49(0.070,0.062,0.826,0.805)
	-0.38	23(0.077,0.058,0.856,0.817)	33(0.067,0.055,0.837,0.803)	42(0.067,0.054,0.822,0.799)
	-0.4	21(0.087,0.057,0.855,0.812)	30(0.074,0.056,0.849,0.807)	39(0.072,0.058,0.828,0.801)

Table 3.2: Sample size (empirical type I error of "GEE-Naive",empirical type I error of "GEE-Jackknife", empirical power of "GEE-Naive", empirical power of 'GEE-Jackknife") assuming cluster size m_i follows the DU(34,56) distribution

	β_2	$\rho = 0.02$	$\rho = 0.04$	$\rho = 0.06$
$\theta = 2$	-0.35	18(0.094,0.062,0.851,0.803)	25(0.071,0.053,0.830,0.790)	33(0.075,0.059,0.827,0.797)
	-0.38	15(0.094,0.058,0.852,0.792)	22(0.076,0.056,0.847,0.806)	29(0.073,0.056,0.835,0.798)
	-0.4	14(0.094,0.062,0.857,0.797)	20(0.085,0.060,0.845,0.799)	26(0.079,0.061,0.832,0.791)
$\theta = 2.5$	-0.35	22(0.085,0.056,0.851,0.809)	32(0.069,0.055,0.827,0.800)	41(0.068,0.057,0.828,0.804)
	-0.38	19(0.082,0.057,0.848,0.797)	27(0.080,0.063,0.842,0.803)	36(0.071,0.059,0.844,0.817)
	-0.4	17(0.088,0.060,0.839,0.785)	25(0.075,0.058,0.838,0.801)	33(0.079,0.062,0.831,0.800)
$\theta = 3$	-0.35	26(0.080,0.059,0.846,0.813)	38(0.069,0.057,0.828,0.802)	49(0.069,0.058,0.810,0.789)
	-0.38	23(0.082,0.059,0.847,0.807)	33(0.071,0.058,0.835,0.806)	42(0.062,0.053,0.825,0.799)
	-0.4	21(0.077,0.056,0.848,0.801)	30(0.071,0.056,0.839,0.808)	39(0.069,0.055,0.829,0.804)

Table 3.3: Sample size (empirical type I error of "GEE-Naive",empirical type I error of "GEE-Jackknife", empirical power of "GEE-Naive", empirical power of 'GEE-Jackknife") assuming cluster size m_i follows the DU(10,80) distribution

	β_2	$\rho = 0.02$	$\rho = 0.04$	$\rho = 0.06$
$\theta = 1.5$	-0.2	19(0.099,0.058,0.860,0.798)	28(0.080,0.056,0.841,0.794)	37(0.075,0.057,0.819,0.785)
	-0.25	17(0.098,0.062,0.859,0.794)	25(0.081,0.059,0.850,0.798)	32(0.084,0.064,0.827,0.782)
	-0.3	15(0.108,0.064,0.856,0.788)	22(0.098,0.065,0.838,0.797)	30(0.082,0.059,0.830,0.788)
$\theta = 2$	-0.2	24(0.081,0.058,0.844,0.795)	35(0.068,0.050,0.831,0.799)	46(0.067,0.052,0.825,0.791)
	-0.25	21(0.089,0.061,0.857,0.807)	30(0.079,0.059,0.828,0.784)	40(0.080,0.054,0.815,0.784)
	-0.3	19(0.088,0.056,0.867,0.803)	28(0.076,0.054,0.834,0.792)	37(0.080,0.058,0.828,0.788)
$\theta = 2.5$	-0.2	28(0.078,0.057,0.831,0.791)	42(0.073,0.058,0.818,0.783)	56(0.063,0.052,0.828,0.804)
	-0.25	24(0.087,0.053,0.840,0.792)	36(0.077,0.059,0.823,0.783)	48(0.075,0.055,0.816,0.785)
	-0.3	22(0.092,0.063,0.826,0.797)	33(0.080,0.061,0.837,0.794)	44(0.069,0.057,0.830,0.802)

Table 3.4: Sample size (N, N_m, N_t, N_{mt}) assuming cluster size m_i follows either DU(34,56) or DU(10,80) distributions

DU	θ	β_2	$\rho = 0.02$	$\rho = 0.04$	$\rho = 0.06$
DU(34,56)	$\theta = 2$	-0.35	(18,18,15,15)	(25,25,22,21)	(33,33,28,28)
		-0.38	(15,15,13,13)	(22,22,19,19)	(29,28,25,24)
		-0.4	(14,14,12,12)	(20,20,17,17)	(26,26,23,22)
	$\theta = 2.5$	-0.35	(22,22,18,18)	(32,31,27,27)	(41,41,35,35)
		-0.38	(19,19,16,16)	(27,27,23,23)	(36,35,31,30)
		-0.4	(17,17,15,15)	(25,25,21,21)	(33,32,28,28)
	$\theta = 3$	-0.35	(26,26,22,22)	(38,37,32,32)	(49,48,42,42)
		-0.38	(23,22,19,19)	(33,32,28,28)	(42,42,37,36)
		-0.4	(21,21,17,17)	(30,29,26,25)	(39,38,34,33)
DU(10,80)	$\theta = 2$	-0.35	(19,18,16,15)	(28,25,24,21)	(37,33,32,28)
		-0.38	(17,15,14,13)	(25,22,21,19)	(32,28,28,24)
		-0.4	(15,14,13,12)	(22,20,19,17)	(30,26,26,22)
	$\theta = 2.5$	-0.35	(24,22,20,18)	(35,31,30,27)	(46,41,40,35)
		-0.38	(21,19,17,16)	(30,27,26,23)	(40,35,35,30)
		-0.4	(19,17,16,15)	(28,25,24,21)	(37,32,32,28)
	$\theta = 3$	-0.35	(28,26,24,22)	(42,37,36,32)	(56,48,48,42)
		-0.38	(24,22,21,19)	(36,32,31,28)	(48,42,42,36)
		-0.4	(22,21,19,17)	(33,29,29,25)	(44,38,38,33)

CHAPTER 4
SAMPLE SIZE CALCULATION FOR CLUSTER RANDOMIZED TRIALS WITH MULTIPLE BINARY CO-PRIMARY ENDPOINTS

In this chapter, we investigate sample size calculation for CRTs with $K \geq 2$ binary co-primary points.

4.1. Statistical Model and Sample Size Approach

Suppose in a CRT, N clusters are randomized to two arms: control or experimental. For simplicity, we assume a common cluster size (number of subjects in each cluster) and denoted by m. Researchers want to evaluate K primary binary endpoints simultaneously. Let y_{ijk} denote the observation of the kth primary endpoint from the jth subject in the ith cluster. Let $E(y_{ijk}) = \mu_{ijk}$, and μ_{ijk} is modeled by a logit model,

$$\log\left(\frac{\mu_{ijk}}{1-\mu_{ijk}}\right) = \beta_{1k} + \beta_{2k} r_i, \tag{4.1}$$

where $r_i = 0/1$ indicates that the ith cluster is assigned to the control/experimental arm. We define $\bar{r} = E(r_i)$ to be the probability of a cluster receiving the experimental intervention. The parameter β_{1k} represents the log-transformed odds for the kth endpoint under control, while β_{2k} denotes the log-transformed odds ratio between the experimental and control arms, representing the intervention effect on the kth endpoint. What makes the design of CRT with multiple endpoints challenging is that within each cluster, there are multiple sources of dependence among the responses. We define: (1) $\rho_0^k = \text{corr}(y_{ijk}, y_{ij'k})$ for $j \neq j'$, which is the inter-subject correlation within endpoint. It can also be considered as an endpoint-specific ICC; (2) $\rho_1^{kk'} = \text{corr}(y_{ijk}, y_{ij'k'})$ for $j \neq j'$ and $k \neq k'$, which

characterizes the inter-subject correlation across endpoints. (3) $\rho_2^{kk'} = \text{corr}(y_{ijk}, y_{ijk'})$ for $k \neq k'$, which characterizes the intra-subject correlation across endpoints. It is obvious that $\rho_1^{kk'} = \rho_1^{k'k}$ and $\rho_2^{kk'} = \rho_2^{k'k}$. In the following, we define $\rho_1^{kk'}$ and $\rho_2^{kk'}$ in such a way that $k < k'$. Note that, $\rho_1^{kk} = \rho_0^k$ and $\rho_2^{kk} = 1$. Hence we have three sets of correlation coefficients: $\boldsymbol{\rho}_0 = \{\rho_0^k\}$, $\boldsymbol{\rho}_1 = \{\rho_1^{kk'}\}$, and $\boldsymbol{\rho}_2 = \{\rho_2^{kk'}\}$, where $\boldsymbol{\rho}_0$ is of size K and both $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ are of size $K(K-1)/2$. Let $\boldsymbol{Y}_{ik} = (y_{i1k}, ..., y_{imk})'$ be the cluster-specific response vector for the kth endpoint, and $\boldsymbol{R}$ be the correlation matrix of $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}', ..., \boldsymbol{Y}_{iK}')'$, the correlation matrix $\boldsymbol{R}$ can be expressed as

$$\begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} & \cdots & \boldsymbol{R}_{1K} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} & \cdots & \boldsymbol{R}_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{R}_{1K} & \boldsymbol{R}_{2K} & \cdots & \boldsymbol{R}_{KK} \end{pmatrix},$$

where

$$\boldsymbol{R}_{kk} = (1 - \rho_0^k)\boldsymbol{I}_m + \rho_0^k \boldsymbol{J}_m,$$

and

$$\boldsymbol{R}_{kk'} = (\rho_2^{kk'} - \rho_1^{kk'})\boldsymbol{I}_m + \rho_1^{kk'} \boldsymbol{J}_m.$$

Throughout the derivation, $\boldsymbol{I}_u$ is defined as a $u \times u$ identity matrix and $\boldsymbol{J}_u$ is defined as a $u \times u$ matrix with all elements being 1. There are constraints on the values of the correlation parameters $\{\boldsymbol{\rho}_0, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2\}$ so that $\boldsymbol{R}$ is positive definite. To check whether $\boldsymbol{R}$ is positive definite, it is equivalent to check whether all eigenvalues of $\boldsymbol{R}$ are greater than 0. For randomized trials with multiple endpoints, researchers [10, 26] have modeled the dependence among endpoints assuming a common correlation ρ. This parsimonious correlation structure inspires us to consider a parsimonious matrix $\boldsymbol{R}$ for CRTs with multiple endpoints, which is $\rho_0^k = \rho_0$, $\rho_1^{kk'} = \rho_1$, $\rho_2^{kk'} = \rho_2$, $\forall\, k, k'$. For this special case, $\boldsymbol{R}$ can be

succinctly presented as

$$(1-\rho_0+\rho_1-\rho_2)\boldsymbol{I}_{mK}+(\rho_2-\rho_1)\boldsymbol{J}_K\otimes\boldsymbol{I}_m+(\rho_0-\rho_1)\boldsymbol{I}_K\otimes\boldsymbol{J}_m+\rho_1\boldsymbol{J}_{mK}, \tag{4.2}$$

where $\otimes$ is the Kronecker product operator. Li et al. [30] showed that (4.2) has four distinct eigenvalues,

$$\begin{aligned}
\gamma_1 &= 1-\rho_0-\rho_1-\rho_2,\\
\gamma_2 &= 1-\rho_0-(K-1)(\rho_1-\rho_2),\\
\gamma_3 &= 1+(m-1)(\rho_0-\rho_1)-\rho_2,\\
\gamma_4 &= 1+(m-1)\rho_0+(K-1)(m-1)\rho_1+(K-1)\rho_2.
\end{aligned}$$

Therefore, the condition of $\boldsymbol{R}$ being positive definite is $\min\{\gamma_1,\gamma_2,\gamma_3,\gamma_4\}>0$.

Define mean vectors $E(\boldsymbol{Y}_{ik})=\boldsymbol{\mu}_{ik}$ and $E(\boldsymbol{Y}_i)=\boldsymbol{\mu}_i=(\boldsymbol{\mu}_{i1}',...,\boldsymbol{\mu}_{iK}')'$. Under Model (4.1), the parameter μ_{ijk} does not depend on the subscript j, hence we write $\boldsymbol{\mu}_{ik}=\mu_{ik}\mathbf{1}_m$, where $\mathbf{1}_m$ is a vector of length m with all elements being 1. Furthermore, we have $\mu_{ik}=\theta_{0k}=\frac{\exp(\beta_{1k})}{\exp(\beta_{1k})+1}$ under control ($r_i=0$), and $\mu_{ik}=\theta_{1k}=\frac{\exp(\beta_{1k}+\beta_{2k})}{\exp(\beta_{1k}+\beta_{2k})+1}$ under experimental intervention ($r_i=1$). Finally, we define variance matrix

$$\boldsymbol{v}_{ik}=Cov(\boldsymbol{Y}_{ik})=\boldsymbol{V}_{ik}^{0.5}\boldsymbol{R}_{kk}\boldsymbol{V}_{ik}^{0.5} \tag{4.3}$$

and covariance matrix

$$\boldsymbol{v}_{ikk'}=Cov(\boldsymbol{Y}_{ik},\boldsymbol{Y}_{ik'})=\boldsymbol{V}_{ik}^{0.5}\boldsymbol{R}_{kk'}\boldsymbol{V}_{ik'}^{0.5}, \tag{4.4}$$

where $\boldsymbol{V}_{ik} = \text{diag}[\mu_{ik}(1-\mu_{ik})\mathbf{1_m}]$ is an $m \times m$ diagonal matrix.

Let $\boldsymbol{\beta}_k = (\beta_{1k}, \beta_{2k})'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_K')'$. Based on the GEE approach, utilizing an independent working correlation structure, the parameters $\boldsymbol{\beta}$ can be estimated by solving:

$$\boldsymbol{S}(\boldsymbol{\beta}) = N^{-1/2} \sum_{i=1}^{N} \boldsymbol{D}_i(\boldsymbol{\beta})^T \boldsymbol{V}_i^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} = \mathbf{0}, \tag{4.5}$$

where

$$\boldsymbol{D}_i(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{D}_{i1}(\boldsymbol{\beta}_1) & \mathbf{0_{m\times 2}} & \cdots & \mathbf{0_{m\times 2}} \\ \mathbf{0_{m\times 2}} & \boldsymbol{D}_{i2}(\boldsymbol{\beta}_2) & \cdots & \mathbf{0_{m\times 2}} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0_{m\times 2}} & \mathbf{0_{m\times 2}} & \cdots & \boldsymbol{D}_{iK}(\boldsymbol{\beta}_K) \end{pmatrix} \tag{4.6}$$

with $\boldsymbol{D}_{ik}(\boldsymbol{\beta}_k) = \frac{\partial \boldsymbol{\mu}_{ik}}{\partial \boldsymbol{\beta}_k} = \mu_{ik}(1-\mu_{ik})\mathbf{1}_m[1, r_i]$ for $k = 1, \cdots, K$, and

$$\boldsymbol{V}_i = \begin{pmatrix} \boldsymbol{V}_{i1} & \mathbf{0_{m\times m}} & \cdots & \mathbf{0_{m\times m}} \\ \mathbf{0_{m\times m}} & \boldsymbol{V}_{i2} & \cdots & \mathbf{0_{m\times m}} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0_{m\times m}} & \mathbf{0_{m\times m}} & \cdots & \boldsymbol{V}_{iK} \end{pmatrix}. \tag{4.7}$$

Equations (4.6) and (4.7) imply that the score function (4.5) can be decomposed into K sub-functions. That is, for each endpoint k, the parameters $\boldsymbol{\beta}_k$ can be separately estimated from

$$\boldsymbol{S}_k(\boldsymbol{\beta}_k) = N^{-1/2} \sum_{i=1}^{N} \boldsymbol{D}_{ik}(\boldsymbol{\beta}_k)^T \boldsymbol{V}_{ik}^{-1} [\boldsymbol{Y}_{ik} - \boldsymbol{\mu}_{ik}(\boldsymbol{\beta}_k)] = \mathbf{0}.$$

With some derivation, the solution is

$$\begin{aligned}
\hat{\beta}_{1k} &= \log\left(\frac{\sum_{i=1}^{N}(1-r_i)\sum_{j=1}^{m}y_{ijk}}{mN_0-\sum_{i=1}^{N}(1-r_i)\sum_{j=1}^{m}y_{ijk}}\right),\\
\hat{\beta}_{2k} &= \log\left(\frac{\sum_{i=1}^{N}r_i\sum_{j=1}^{m}y_{ijk}}{mN_1-\sum_{i=1}^{N}r_i\sum_{j=1}^{m}y_{ijk}}\right)-\log\left(\frac{\sum_{i=1}^{N}(1-r_i)\sum_{j=1}^{m}y_{ijk}}{mN_0-\sum_{i=1}^{N}(1-r_i)\sum_{j=1}^{m}y_{ijk}}\right),
\end{aligned}$$

where $N_1=\sum_{i=1}^{N}r_i$ and $N_0=N-N_1$ are the number of clusters receiving the experimental and control intervention, respectively.

Through Taylor expansion, Liang & Zeger [33] showed that $N^{1/2}(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)$ can be approximated by $\boldsymbol{\Gamma}_k^{-1}\boldsymbol{S}_k(\boldsymbol{\beta}_k)$, where

$$\begin{aligned}
\boldsymbol{\Gamma}_k &= \lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\boldsymbol{D}_{ik}(\boldsymbol{\beta}_k)^T\boldsymbol{V}_{ik}^{-1}\boldsymbol{D}_{ik}(\boldsymbol{\beta}_k) \qquad (4.8)\\
&= m\begin{pmatrix}(1-\bar{r})\xi_{1k}+\bar{r}\xi_{2k} & \bar{r}\xi_{2k}\\ \bar{r}\xi_{2k} & \bar{r}\xi_{2k}\end{pmatrix}.
\end{aligned}$$

Here $\xi_{1k}=\theta_{0k}(1-\theta_{0k})$ and $\xi_{2k}=\theta_{1k}(1-\theta_{1k})$. On the other hand, as a linear combination of $\boldsymbol{Y}_{ik}$, it is straightforward to show that $\boldsymbol{S}_k(\boldsymbol{\beta}_k)$ asymptotically has a zero-mean multivariate normal distribution with variance

$$\begin{aligned}
\boldsymbol{\Omega}_k &= N^{-1}\sum_{i=1}^{N}\boldsymbol{D}_{ik}(\boldsymbol{\beta}_k)^T\boldsymbol{V}_{ik}^{-1}\boldsymbol{v}_{ik}\boldsymbol{V}_{ik}^{-1}\boldsymbol{D}_{ik}(\boldsymbol{\beta}_k)\\
&= [m+m(m-1)\rho_0^k]\begin{pmatrix}(1-\bar{r})\xi_{1k}+\bar{r}\xi_{2k} & \bar{r}\xi_{2k}\\ \bar{r}\xi_{2k} & \bar{r}\xi_{2k}\end{pmatrix}.
\end{aligned}$$

The asymptotic normality of $\boldsymbol{S}_k(\boldsymbol{\beta}_k)$ implies that $N^{1/2}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$ also asymptotically follows a normal distribution with zero-mean and covariance

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Gamma}_k^{-1}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}_k^{-1}. \tag{4.9}$$

Define σ_{2k}^2 to be the $(2,2)$th element of $\boldsymbol{\Sigma}_k$. We can show that

$$\sigma_{2k}^2 = \frac{1+(m-1)\rho_0^k}{m(1-\bar{r})\xi_{1k}} + \frac{1+(m-1)\rho_0^k}{m\bar{r}\xi_{2k}}. \tag{4.10}$$

A Wald-type test statistic for β_{2k} can be constructed as $w_k = \frac{N^{1/2}\hat{\beta}_{2k}}{\hat{\sigma}_{2k}}$, where $\hat{\sigma}_{2k}$ denotes the estimate of σ_{2k}. In practice, $\hat{\sigma}_{2k}^2$ can be obtained from the (2,2)th element of the robust sandwich type estimator $\hat{\boldsymbol{\Sigma}}_k$ which is given by

$$\hat{\boldsymbol{\Sigma}}_k = \hat{\boldsymbol{\Gamma}}_k^{-1}\hat{\boldsymbol{\Omega}}_k\hat{\boldsymbol{\Gamma}}_k^{-1}, \tag{4.11}$$

where

$$\hat{\boldsymbol{\Gamma}}_k = N^{-1}\sum_{i=1}^{N}\boldsymbol{D}_{ik}(\hat{\boldsymbol{\beta}}_k)^T\hat{\boldsymbol{V}}_{ik}^{-1}\boldsymbol{D}_{ik}(\hat{\boldsymbol{\beta}}_k), \tag{4.12}$$

and

$$\hat{\boldsymbol{\Omega}}_k = N^{-1}\sum_{i=1}^{N}\boldsymbol{D}_{ik}(\hat{\boldsymbol{\beta}}_k)^T\hat{\boldsymbol{V}}_{ik}^{-1}\hat{\boldsymbol{\epsilon}}_{ik}\hat{\boldsymbol{\epsilon}}_{ik}^T\hat{\boldsymbol{V}}_{ik}^{-1}\boldsymbol{D}_{ik}(\hat{\boldsymbol{\beta}}_k). \tag{4.13}$$

In (4.13), $\hat{\boldsymbol{\epsilon}}_{ik} = \boldsymbol{Y}_{ik} - \hat{\boldsymbol{\mu}}_{ik}$ is the residual vector of the ith cluster on the kth endpoint.

Note that, the vector of test statistics $\boldsymbol{W} = (w_1, ...w_K)'$ asymptotically follows a multivariate normal distribution with mean vector $\boldsymbol{\eta} = \left(\frac{N^{1/2}\beta_{21}}{\sigma_{21}}, ..., \frac{N^{1/2}\beta_{2K}}{\sigma_{2K}}\right)'$ and we denote the covariance matrix by $\boldsymbol{\Phi}$. Let $\phi_{kk'}$ be the (k,k')th element of $\boldsymbol{\Phi}$. It is obvious that $\phi_{kk} = 1$ for $k = 1, \cdots, K$. To make joint inference on the K co-primary endpoints, we need to learn $\phi_{kk'} = Cov(w_k, w_{k'})$ for $k \neq k'$, i.e., the covariance of test statistics between endpoints. To derive $\phi_{kk'}$, we first derive $\boldsymbol{\Sigma}_{kk'} = Cov[N^{1/2}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k), N^{1/2}(\hat{\boldsymbol{\beta}}_{k'} - \boldsymbol{\beta}_{k'})]$. Note that

$$\begin{aligned}\boldsymbol{\Sigma}_{kk'} &\approx Cov[\boldsymbol{\Gamma}_k^{-1}\boldsymbol{S}_k(\boldsymbol{\beta}_k), \boldsymbol{\Gamma}_{k'}^{-1}\boldsymbol{S}_{k'}(\boldsymbol{\beta}_{k'})] \\ &= \boldsymbol{\Gamma}_k^{-1} Cov[\boldsymbol{S}_k(\boldsymbol{\beta}_k), \boldsymbol{S}_{k'}(\boldsymbol{\beta}_{k'})]\boldsymbol{\Gamma}_{k'}^{-1} \\ &= \boldsymbol{\Gamma}_k^{-1}\boldsymbol{\Omega}_{kk'}\boldsymbol{\Gamma}_{k'}^{-1},\end{aligned}$$

where

$$\boldsymbol{\Omega}_{kk'} = N^{-1}\sum_{i=1}^{N} \boldsymbol{D}_{ik}(\boldsymbol{\beta}_k)^T \boldsymbol{V}_{ik}^{-1}\boldsymbol{v}_{ikk'}\boldsymbol{V}_{ik'}^{-1}\boldsymbol{D}_{ik'}(\boldsymbol{\beta}_{k'}).$$

Recall that $\boldsymbol{v}_{ikk'}$ is defined in (4.4). With some simplification, we have

$$\boldsymbol{\Omega}_{kk'} = [m\rho_2^{kk'} + m(m-1)\rho_1^{kk'}] \begin{pmatrix} (1-\bar{r})\xi_{1kk'} + \bar{r}\xi_{2kk'} & \bar{r}\xi_{2kk'} \\ \bar{r}\xi_{2kk'} & \bar{r}\xi_{2kk'} \end{pmatrix},$$

where $\xi_{1kk'} = \sqrt{\xi_{1k}\xi_{1k'}}$ and $\xi_{2kk'} = \sqrt{\xi_{2k}\xi_{2k'}}$.

Recall that $\boldsymbol{\Gamma}_k$ is presented in (4.12). Then we obtain $\sigma^2_{2kk'} = Cov[N^{1/2}(\hat{\beta}_{2k} - \beta_{2k}), N^{1/2}(\hat{\beta}_{2k'} - \beta_{2k'})]$, which is the $(2,2)$th element of $\boldsymbol{\Sigma}_{kk'}$:

$$\sigma^2_{2kk'} = \frac{[\rho_2^{kk'} + (m-1)\rho_1^{kk'}]\xi_{1kk'}}{m(1-\bar{r})\xi_{1k}\xi_{1k'}} + \frac{[\rho_2^{kk'} + (m-1)\rho_1^{kk'}]\xi_{2kk'}}{m\bar{r}\xi_{2k}\xi_{2k'}}.$$

Then the (k,k')th element of the covariance matrix for the vector of test statistics $\boldsymbol{W}$ has the following expression,

$$\phi_{kk'} = \begin{cases} 1 & k = k', \\ \frac{\sigma^2_{2kk'}}{\sigma_{2k}\sigma_{2k'}} & k \neq k'. \end{cases}$$

With the distribution of $\boldsymbol{W}$ fully characterized, joint inference about the co-primary endpoints can be performed to evaluate arbitrarily constructed hypotheses. For example, the intersection-union (IU) hypothesis has been frequently employed [10, 52, 53]

$$\begin{aligned} H_0 &: \beta_{2k} = 0, \quad \text{for at least one } k \\ H_1 &: \beta_{2k} > 0, \quad \forall\, k. \end{aligned} \tag{4.14}$$

The rationale for the IU hypothesis is that, in order to avoid "cherry picking" or "by chance" findings that may mislead the conclusion, it is appropriate to require statistical significance on all primary endpoints [52]. The cost, however, is that achieving significance simultaneously becomes more difficult as the number of endpoints increases. Given the number of clusers N, allocation ratio $\bar{r}$, cluster size m, true parameters $\{\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_K\}$ and correlation $\{\boldsymbol{\rho}_0, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2\}$, the power to detect the intervention effect can be calculated by

$$\text{power} = \text{Prob}\left(\mathcal{R} = \bigcap_{k=1}^{K}\{w_k > c_k\}\right) = \int_{c_1}^{\infty} \cdots \int_{c_K}^{\infty} f_{\boldsymbol{W}}(w_1, ..., w_K)\, d_{w_1} \cdots d_{w_K}, \tag{4.15}$$

where $\boldsymbol{c} = \{c_1, \cdots, c_K\}$ are endpoint-specific critical values for rejection, $f_{\boldsymbol{W}}(\cdot)$ denotes the density function of a multivariate normal distribution with mean $\boldsymbol{\eta}$ and variance matrix $\boldsymbol{\Phi}$, and $\mathcal{R}$ represents the rejection region corresponding to a particular hypothesis. In (4.14) we assume that the desired treatment effect is represented by $\beta_{2k} > 0$. It can easily accommodate treatment effect in the opposite direction, such as an adverse event, by changing the definition of the outcome. Another example is the union-intersection (UI) hypothesis, which considers the experimental treatment to be effective if any one of the endpoints shows statistical significance [6]. It is expressed as

$$
\begin{aligned}
H_0 &: \beta_{2k} = 0, \quad \forall\, k, \\
H_1 &: \beta_{2k} > 0, \quad \text{for at least one } k.
\end{aligned}
$$

For a UI hypothesis, the rejection region is $\mathcal{R} = \left(\Pi - \bigcap_{k=1}^{K}\{w_k \le c_k\}\right)$ where Π indicates the full space on $\boldsymbol{W}$.

To account for the uncertainty in estimating the asymptotic variance of $\hat{\beta}_{2k}$, we may alternatively assume w_k to follows a t-distribution. Then the power function (4.15) can be modified as

$$
\text{power} = \text{Prob}\left(\mathcal{R} = \bigcap_{k=1}^{K}\{w_k > c_k\}\right) = \int_{c_1}^{\infty} \cdots \int_{c_K}^{\infty} f^*_{\boldsymbol{W}}(w_1, ..., w_K)\, d_{w_1} \cdots d_{w_K}, \qquad (4.16)
$$

where $f^*_{\boldsymbol{W}}(\cdot)$ denotes the density function of a multivariate t-distribution with location parameter $\boldsymbol{\eta}$, shape matrix $\boldsymbol{\Phi}$, and $N - 2K$ degrees of freedom.

Equation (4.10) implies that as $m \to \infty$, σ^2_{2k} converges to $\frac{\rho_0^k}{(1-\bar{r})\xi_{1k}} + \frac{\rho_0^k}{\bar{r}\xi_{2k}}$ instead of shrinking to 0. That is, when the number of clusters (N) is fixed, there is an upper limit of power increase that can be achieved by enlarging the cluster size (m) to infinity. This fact reflects an important point for the design of CRTs in practice. For CRTs, the limiting factor for power is usually the number of clusters N instead of the cluster size m. Therefore, an under-powered trial because of insufficient recruitment of clusters can not always be compensated by increasing the cluster size m. Motivated by this point, in the rest of the study, we discuss sample size calculation in terms of determining the number of clusters N given cluster size m and other design factors.

In practice, one convenient approach to specifying the critical values is to set $c_1 = \cdots = c_K = z_\alpha$, where z_α is the $(1-\alpha)$th quantile of the standard normal distribution [52, 53]. Such an approach controls the type I error rate strictly below α over the null

space ($H_0 : \beta_{2k} = 0 \text{ for at least one } k$). The upper limit of type I error (α) is reached under the special scenario where one endpoint shows no treatment effect and the other $K-1$ endpoints show large effects. Researchers [10, 26] have proposed other methods to obtain less conservative critical values $\boldsymbol{c}$. For example, rather than controlling the type I error rate strictly below α over the null space, [10] proposed a method to calculate a common critical value c among all endpoints which controls the "average type I error rate" to be less than or equal to the nominal level α over the null space. In this study we assume that $\boldsymbol{c}$ has been specified through a certain existing approach, the implementation of which is straightforward with $\boldsymbol{\eta}$ and $\boldsymbol{\Phi}$ fully characterized. Given a particular sample size N, evaluating the power only involves integration of a multivariate normal or t distribution through (4.15) and (4.16). By calculating power for an increasing series of N, we can identify the smallest sample size to achieve the desired power.

It is noteworthy that there is a monotone relationship between correlation parameters and sample size requirement, described by the following theorem.

Theorem 4.1 *With all other design parameters fixed, a larger $\rho_0^k \ \forall k$ is always associated with a smaller power (larger sample size); on the other hand, a larger $\rho_1^{kk'}$ or $\rho_2^{kk'} \ \forall k \neq k'$ is always associated with a larger power (smaller sample size).*

Note that Theorem 4.1 holds for power calculated based on either multivariate normal or t distribution. The proof is presented in Appendix C.1.

In practice, the ICCs (ρ_0^k) are usually positive, and Theorem 1 suggests that ignoring the ICCs (i.e., setting $\rho_0^k = 0$) will lead to an under-estimated sample size and an under-powered trial. On the other hand, the consequence of ignoring the correlations between endpoints (i.e. setting $\boldsymbol{\rho}_1 = \boldsymbol{\rho}_2 = \mathbf{0}$) is uncertain depending on the direction and magnitude of the true correlation.

When the number of clusters N is small, the uncorrected robust sandwich estimator of σ_{2k}^2 (4.11) is known to be biased downwards, leading to an inflated type I error [32]. Many researchers [24, 38] have proposed bias-corrected variance estimators to address

this problem. To introduce these bias-corrected variance estimators, we use a general expression of $\hat{\boldsymbol{\Omega}}_k$:

$$\hat{\boldsymbol{\Omega}}_k = N^{-1} \sum_{i=1}^{N} \boldsymbol{D}_{ik}(\hat{\boldsymbol{\beta}}_k)^T \hat{\boldsymbol{V}}_{ik}^{-1} \boldsymbol{B}_{ik} \hat{\boldsymbol{\epsilon}}_{ik} \hat{\boldsymbol{\epsilon}}_{ik}^T \boldsymbol{B}_{ik}^T \hat{\boldsymbol{V}}_{ik}^{-1} \boldsymbol{D}_{ik}(\hat{\boldsymbol{\beta}}_k). \tag{4.17}$$

Note that with $\boldsymbol{B}_{ik} = \boldsymbol{I}_m$, Equation (4.17) corresponds to the uncorrected sandwich estimator (4.11). We denote this uncorrected robust sandwich estimator as BC_{naive}. By setting $\boldsymbol{B}_{ik} = (\boldsymbol{I}_m - \boldsymbol{H}_{ik})^{-1}$, where $\boldsymbol{H}_{ik} = \boldsymbol{D}_{ik}(\hat{\boldsymbol{\beta}}_k)\hat{\boldsymbol{\Gamma}}_k^{-1}\boldsymbol{D}_{ik}(\hat{\boldsymbol{\beta}}_k)^T\hat{\boldsymbol{V}}_{ik}^{-1}$, we have the bias-corrected variance estimator proposed by Mancl & DeRouen (2001) [38] which we denote as BC_{MD}. Finally, with $\boldsymbol{B}_{ik} = (\boldsymbol{I}_m - \boldsymbol{H}_{ik})^{-1/2}$, (4.17) becomes the bias-corrected variance estimator proposed by Kauermann & Carroll [24] which we denote as BC_{KC}. As was shown in Preisser et al. [47], we have $BC_{naive} < BC_{KC} < BC_{MD}$. In the simulation study, we will explore the performance of these three approaches in preserving the desired power and type I error under small sample sizes.

4.2. Simulation

We conduct simulations to evaluate the performance of the proposed sample size method. In clinical trials that evaluate multiple co-primary endpoints, most studies consider less than or equal to 3 endpoints [41]. Hence we explore two scenarios: $K = 2$ and 3. We assume balanced randomization, i.e. $\bar{r} = 0.5$. In practice, even for $K = 3$, it is difficult to specify every elements of $\{\boldsymbol{\rho}_0, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2\}$. In this simulation we consider the simplified case where $\rho_0^k = \rho_0$, $\rho_1^{kk'} = \rho_1$, $\rho_2^{kk'} = \rho_2 \; \forall \; k, k'$. For ρ_0, we explore the values of $0.01, 0.05$ which reflect small ICCs commonly reported in CRTs [39]. We further assume ρ_1 to be smaller than ρ_0, and set $\rho_1 = \rho_0/2$. For ρ_2, we consider values of 0.2 and 0.5 which represent moderate within-subject correlation. For the case of $K = 2$, we set $\boldsymbol{b}_1 = (\beta_{11}, \beta_{12})' = (0, 0.3)'$. We explore two sets of treatment effects $\boldsymbol{b}_2 = (\beta_{21}, \beta_{22})' = (0.6, 0.6)'$ or $(0.6, 0.7)'$. For the case of $K = 3$, we similarly set $\boldsymbol{b}_1 = (\beta_{11}, \beta_{12}, \beta_{13})' = (0, 0.1, 0.3)'$,

and explore treatment effects $\boldsymbol{b}_2 = (\beta_{21}, \beta_{22}, \beta_{23})' = (0.6, 0.6, 0.6)'$ or $(0.5, 0.6, 0.7)'$. We set the critical values $c_1 = \cdots = c_k = c$ at the 95th percentile of the standard normal or t distribution. For the cluster size m, we explore values of 60 and 80. For each combination of design parameters, the intersection-union hypothesis (4.14) will be evaluated. The simulation algorithm for a particular combination of design parameters is described as follows:

1. Numerically search for N, which is the smallest even number of clusters that achieves power≥ 0.8. The powers calculated based on (4.15) or (4.16) are recorded as the theoretical powers.

2. We run 5000 iterations, and for each iteration:

   (a) Generate a dataset of N clusters with cluster size m and K binary endpoints. Given the set of design parameters, every cluster has a vector of correlated binary outcomes $\boldsymbol{Y}_i$, generated using the R package "MultiOrd" [4].

   (b) For each endpoint, we calculate the test statistics w_k.

3. The empirical power is computed as the proportion of iterations with $I\left(\bigcap_{k=1}^{K} w_k > c\right) = 1$, where $c = z_{0.95}$ or $t_{0.95,N-2K}$.

4. The empirical type I error is evaluated by setting the last element of $\boldsymbol{b}_2$ to 0 (e.g., if $\boldsymbol{b}_2 = (0.6, 0.6, 0.6)'$, we set it to $\boldsymbol{b}_2 = (0.6, 0.6, 0)'$) and repeat Steps 2-3. Note that, setting one element of $\boldsymbol{b}_2$ to 0 will lead to a theoretical type I error close to the nominal level α. The theoretical type I error is calculated using the same settings based on (4.15) or (4.16).

Tables 4.1-4.4 summarize the simulation results for all combinations of design parameters and the two power functions (Equations (4.15) and (4.16)). In terms of power, the empirical power under the BC_{naive} adjustment approach is consistently larger than the theoretical power across all scenarios, while the empirical power under the BC_{MD} approach is consistently smaller than the theoretical power. Overall the empirical power

under the BC_{KC} approach is most close to the theoretical level. In terms of type I error, the empirical type I errors under BC_{naive} is consistently larger than the theoretical level. This is especially true under the normal distribution. The inflation of the type I error is less severe under the t-distribution. Under the BC_{MD} approach the empirical type I error is consistently smaller than the theoretical level across all scenarios. For the BC_{KC} estimator, the empirical type I error under the normal distribution seems to be slightly larger than the theoretical level, while under t-distribution the empirical type I error generally agrees with the theoretical level. The simulation results suggest that overall using the t-distribution with the BC_{KC} adjustment approach has the best performance in terms of maintaining both power and type I error at the nominal levels.

4.3. Application

This section illustrates the application of the proposed method to a CRT study [29] where the research goal is to evaluate whether a cancer Screening Office System (cancer SOS) intervention can improve participation in cancer screening tests. This study includes three binary endpoints, each indicating participation in one of three targeted cancer screening tests: Papanicolaou (Pap) smears (denoted by $k = 1$), mammograms ($k = 2$), and fecal occult blood tests ($k = 3$). All tests are recommended to be performed annually for women of age 50 or older. Eight clinics participated in this study, and the average number of participants in each clinic is 150. It is reported that, during a 12-months follow-up, the estimated probabilities of taking the three screening tests are $(0.484, 0.709, 0.123)$ under control, and $(0.624, 0.758, 0.397)$ under cancer SOS. The corresponding regression parameters are $\hat{\boldsymbol{\beta}}_1 = (-0.064, 0.571)$, $\hat{\boldsymbol{\beta}}_2 = (0.891, 0.251)$, and $\hat{\boldsymbol{\beta}}_3 = (-1.964, 1.546)$, respectively. The estimated intraclass correlations $(\rho_0^1, \rho_0^2, \rho_0^3)$ are $(0.069, 0.003, 0.16)$. However, the estimations of correlation parameters $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ are not reported.

Suppose researchers want to design a new CRT to evaluate the effect of a new intervention on encouraging patients to take the three cancer screening tests. To test the IU hypothesis $H_0 : \beta_{2k} = 0 \text{ for at least one } k$ vs $H_1 : \beta_{2k} > 0 \ \forall\, k$ with a 0.8 power at the 0.05 significance level. Based on the preliminary data obtained from the aforementioned trial, we use the following design parameters to calculate the required sample size. We assume a balanced design ($\bar{r} = 0.5$) with cluster size $m = 150$. The regression parameters are specified as $\boldsymbol{\beta}_1 = (-0.064, 0.683)$, $\boldsymbol{\beta}_2 = (0.891, 1.307)$, and $\boldsymbol{\beta}_3 = (-1.964, 0.865)$. We consider a parsimonious correlation structure, that is $\rho_0^k = \rho_0$, $\rho_1^{kk'} = \rho_1$, $\rho_2^{kk'} = \rho_2$ for $\forall\, k, k'$. We set $\rho_0 = \frac{0.069+0.003+0.16}{3} = 0.077$ and assume $\rho_1 = 0.055$ and $\rho_2 = 0.5$. Using the power function (4.16), the required number of clusters is $N = 26$. Under this setting, if the correlations among the endpoints are ignored (i.e. setting $\rho_1 = \rho_2 = 0$), the calculated sample size becomes $N = 30$, which leads to a 15.38% inflation in sample size.

Table 4.1: Simulation results. In each cell, the first row shows the number of clusters and (theoretical power, empirical powers under BC_{naive}, BC_{MD}, BC_{KC}). The second row shows (theoretical type I error, empirical type I errors of BC_{naive}, BC_{MD}, BC_{KC}) based on multivariate normal distribution with $K = 2$.

(β_{21}, β_{22})	m	(ρ_0, ρ_1)	$\rho_2 = 0.2$	$\rho_2 = 0.5$
$(0.6, 0.6)$	60	$(0.01, 0.005)$	12(0.860,0.880,0.804,0.847)	12(0.867,0.896,0.818,0.859)
			(0.049,0.081,0.051,0.066)	(0.050,0.080,0.052,0.064)
		$(0.05, 0.025)$	26(0.810,0.820,0.781,0.802)	26(0.815,0.834,0.797,0.817)
			(0.050,0.056,0.043,0.049)	(0.050,0.067,0.053,0.061)
	80	$(0.01, 0.005)$	10(0.855,0.880,0.775,0.836)	10(0.863,0.890,0.789,0.848)
			(0.049,0.086,0.045,0.063)	(0.050,0.091,0.051,0.069)
		$(0.05, 0.025)$	24(0.803,0.821,0.779,0.802)	24(0.807,0.822,0.782,0.805)
			(0.050,0.065,0.050,0.057)	(0.050,0.064,0.048,0.055)
$(0.6, 0.7)$	60	$(0.01, 0.005)$	10(0.839,0.870,0.762,0.820)	10(0.848,0.885,0.787,0.848)
			(0.049,0.087,0.050,0.067)	(0.050,0.095,0.052,0.075)
		$(0.05, 0.025)$	24(0.831,0.841,0.800,0.822)	24(0.835,0.845,0.808,0.823)
			(0.049,0.063,0.048,0.056)	(0.050,0.067,0.053,0.061)
	80	$(0.01, 0.005)$	8(0.818,0.864,0.723,0.806)	8(0.827,0.861,0.724,0.806)
			(0.049,0.103,0.049,0.074)	(0.050,0.107,0.055,0.081)
		$(0.05, 0.025)$	22(0.821,0.851,0.806,0.832)	22(0.825,0.841,0.797,0.823)
			(0.049,0.061,0.048,0.055)	(0.050,0.072,0.052,0.062)

Table 4.2: Simulation results. In each cell, the first row shows the number of clusters and (theoretical power, empirical powers under BC_{naive}, BC_{MD}, BC_{KC}). The second row shows (theoretical type I error, empirical type I errors of BC_{naive}, BC_{MD}, BC_{KC}) based on multivariate t-distribution with $K = 2$.

(β_{21}, β_{22})	m	(ρ_0, ρ_1)	$\rho_2 = 0.2$	$\rho_2 = 0.5$
$(0.6, 0.6)$	60	$(0.01, 0.005)$	14(0.866,0.902,0.837,0.872)	12(0.801,0.850,0.748,0.807)
			(0.050,0.059,0.035,0.046)	(0.050,0.059,0.035,0.045)
		$(0.05, 0.025)$	28(0.819,0.842,0.803,0.821)	28(0.823,0.855,0.820,0.840)
			(0.050,0.057,0.042,0.049)	(0.050,0.064,0.051,0.056)
	80	$(0.01, 0.005)$	12(0.859,0.898,0.807,0.859)	12(0.866,0.895,0.810,0.860)
			(0.050,0.056,0.031,0.042)	(0.050,0.056,0.030,0.042)
		$(0.05, 0.025)$	26(0.812,0.829,0.781,0.808)	26(0.816,0.841,0.801,0.820)
			(0.050,0.057,0.042,0.049)	(0.050,0.057,0.043,0.050)
$(0.6, 0.7)$	60	$(0.01, 0.005)$	12(0.843,0.882,0.787,0.842)	12(0.851,0.879,0.782,0.839)
			(0.049,0.060,0.031,0.046)	(0.050,0.062,0.033,0.047)
		$(0.05, 0.025)$	24(0.804,0.821,0.777,0.797)	24(0.809,0.835,0.792,0.813)
			(0.049,0.049,0.035,0.042)	(0.050,0.054,0.041,0.047)
	80	$(0.01, 0.005)$	10(0.814,0.863,0.733,0.809)	10(0.822,0.875,0.745,0.815)
			(0.049,0.059,0.028,0.044)	(0.050,0.048,0.037,0.043)
		$(0.05, 0.025)$	24(0.830,0.857,0.818,0.837)	24(0.834,0.849,0.809,0.832)
			(0.050,0.050,0.037,0.042)	(0.050,0.055,0.041,0.046)

Table 4.3: Simulation results. In each cell, the first row shows the number of clusters and (theoretical power, empirical powers under BC_{naive}, BC_{MD}, BC_{KC}). The second row shows (theoretical type I error, empirical type I errors of BC_{naive}, BC_{MD}, BC_{KC}) based on multivariate normal distribution with $K = 3$.

$(\beta_{21}, \beta_{22}, \beta_{23})$	m	(ρ_0, ρ_1)	$\rho_2 = 0.2$	$\rho_2 = 0.5$
$(0.6, 0.6, 0.6)$	60	$(0.01, 0.005)$	12(0.810,0.853,0.753,0.809)	12(0.827,0.845,0.750,0.802)
			(0.049,0.077,0.044,0.058)	(0.050,0.079,0.048,0.063)
		$(0.05, 0.025)$	30(0.823,0.837,0.799,0.818)	30(0.830,0.841,0.808,0.824)
			(0.049,0.070,0.058,0.065)	(0.050,0.062,0.052,0.056)
	80	$(0.01, 0.005)$	10(0.805,0.844,0.713,0.785)	10(0.821,0.855,0.733,0.805)
			(0.049,0.085,0.047,0.063)	(0.050,0.083,0.044,0.061)
		$(0.05, 0.025)$	28(0.821,0.828,0.796,0.814)	28(0.827,0.841,0.805,0.823)
			(0.050,0.063,0.051,0.057)	(0.050,0.056,0.047,0.050)
$(0.5, 0.6, 0.7)$	60	$(0.01, 0.005)$	14(0.840,0.863,0.790,0.830)	14(0.850,0.871,0.806,0.840)
			(0.048,0.071,0.048,0.059)	(0.050,0.082,0.054,0.067)
		$(0.05, 0.025)$	32(0.811,0.821,0.790,0.807)	32(0.816,0.832,0.801,0.818)
			(0.049,0.059,0.047,0.052)	(0.049,0.066,0.053,0.060)
	80	$(0.01, 0.005)$	12(0.847,0.873,0.796,0.839)	12(0.856,0.883,0.788,0.844)
			(0.049,0.082,0.046,0.060)	(0.050,0.085,0.047,0.065)
		$(0.05, 0.025)$	30(0.811,0.828,0.793,0.811)	30(0.815,0.827,0.794,0.810)
			(0.049,0.062,0.052,0.058)	(0.049,0.067,0.054,0.062)

Table 4.4: Simulation results. In each cell, the first row shows the number of clusters and (theoretical power, empirical powers under BC_{naive}, BC_{MD}, BC_{KC}). The second row shows (theoretical type I error, empirical type I errors of BC_{naive}, BC_{MD}, BC_{KC}) based on multivariate t-distribution with $K = 3$.

$(\beta_{21}, \beta_{22}, \beta_{23})$	m	(ρ_0, ρ_1)	$\rho_2 = 0.2$	$\rho_2 = 0.5$
$(0.6, 0.6, 0.6)$	60	$(0.01, 0.005)$	14(0.806,0.849,0.757,0.807)	14(0.823,0.923,0.789,0.834)
			(0.049,0.056,0.034,0.045)	(0.050,0.054,0.031,0.043)
		$(0.05, 0.025)$	32(0.830,0.867,0.816,0.832)	30(0.807,0.871,0.776,0.796)
			(0.050,0.059,0.045,0.050)	(0.050,0.052,0.041,0.047)
	80	$(0.01, 0.005)$	14(0.878,0.873,0.845,0.892)	12(0.803,0.840,0.724,0.783)
			(0.050,0.058,0.034,0.046)	(0.050,0.054,0.029,0.040)
		$(0.05, 0.025)$	30(0.829,0.879,0.808,0.824)	28(0.802,0.879,0.775,0.799)
			(0.050,0.053,0.041,0.049)	(0.050,0.063,0.049,0.056)
$(0.5, 0.6, 0.7)$	60	$(0.01, 0.005)$	16(0.839,0.844,0.797,0.834)	16(0.849,0.861,0.814,0.841)
			(0.049,0.058,0.038,0.047)	(0.050,0.054,0.034,0.045)
		$(0.05, 0.025)$	34(0.818,0.836,0.807,0.823)	34(0.823,0.842,0.805,0.818)
			(0.049,0.050,0.041,0.045)	(0.050,0.055,0.047,0.050)
	80	$(0.01, 0.005)$	14(0.840,0.862,0.796,0.839)	14(0.849,0.862,0.806,0.847)
			(0.049,0.053,0.031,0.041)	(0.050,0.050,0.030,0.039)
		$(0.05, 0.025)$	32(0.818,0.834,0.807,0.824)	32(0.822,0.834,0.800,0.817)
			(0.049,0.054,0.042,0.048)	(0.050,0.048,0.040,0.043)

CHAPTER 5

CONCLUSION

In this study, we investigate sample size calculation for three kinds of clinical trials: (1). Randomized controlled trials with longitudinal count outcomes; (2). Cluster randomized trials with count outcomes; (3). Cluster randomized trials with multiple binary co-primary endpoints.

For randomized controlled trials with longitudinal count outcomes, we have derived closed-from sample size formulas for both scenarios of comparison of TAR and slopes. Our approach is developed based on the negative binomial distribution. Compared with the traditional Poisson-based approaches, the proposed sample size method offers greater flexibility to accommodate over-dispersion in count variables, which is frequently encountered in practice. In addition to simple computation offered by the closed form formulas, this approach is advantageous in its capability of allowing for arbitrary missing data patterns, correlation structures, and randomization ratios. By including the offset terms, the proposed sample size method provides additional flexibility to measure the count outcome over unequal measurement intervals. We would like to point out that the closed-form sample size formulas are derived based on the MCAR (missing completely at random) assumption. Under the MAR (missing at random) assumption, however, an additional model is needed to account for the missing data mechanism which is assumed to depend on observed data. Because each particular study might have a unique missing data mechanism, it is difficult to derive a general sample size formula under MAR. Hence when there is strong evidence that the missing data is MAR, researchers need to build a model that describes the missing data mechanism in their research setting, and conduct simulation studies to assess its impact on sample size. The proposed closed-form sample size

formula can still be useful by providing the starting point for numerical search or serving as a benchmark to understand the impact of different missing data mechanisms. It is noteworthy that when over-dispersion is caused by excessive prevalence of 0's, assuming the outcome to marginally follow a NB distribution is inappropriate. In such cases, assuming a zero-inflated Poisson or a zero-inflated NB model might be more appropriate.

For cluster randomized trials with count outcomes, we have proposed a sample size approach, which, compared to existing methods [3, 60], is advantageous in its flexibility to incorporate pragmatic features, including over-dispersion, varying cluster size, varying length of follow-up, and arbitrary randomization ratio. Furthermore, the sample size formula has a closed form, facilitating its implementation by practitioners. We theoretically demonstrate that ignoring the pragmatic features will lead to under-estimated sample size. To accommodate the pragmatic features, the proposed method requires the specification of additional design parameters, including the first two moments of cluster size and length of follow up and the mean of square root for length of follow up. In practice, these parameters can be conveniently estimated based on data from previous studies. To address the concern of under-estimated variance by the GEE sandwich estimator under relatively small number of clusters, which is frequently encountered in CRTs, we propose a closed-form variance estimator based on the Jackknife approach. Extensive simulation studies have been conducted to evaluate the performance of the proposed sample size formula and the Jackknife inference procedure. The "GEE-Jackknife" approach can maintain the empirical power and type I error at their nominal levels over a wide range of design configurations.

Finally, in this dissertation, we also investigate sample size calculation for CRTs with mutilple ($K \geq 2$) binary co-primary endpoints. Within each cluster three types of correlations are considered: inter-subject correlation within each endpoint, intra-subject correlation across endpoints, and inter-subject correlation across endpoints. Based on the GEE approach, we have derived a closed-form joint distribution for the K test statistics, which can be used to evaluate power and type I error for arbitrarily constructed hypotheses. To

the best of our knowledge, this is the first attempt to systematically investigate sample size calculation for CTRs with multiple binary co-primary endpoints. We further present a theorem that characterizes the relationship between the three types of correlation and testing power. Furthermore, a number of approaches to adjust for the underestimation bias of the GEE variance estimator under small sample size have been compared. The simulation results suggest that the combination of the t-distribution with the adjustment approach proposed by Kauermann & Carroll [24] achieves the best performance in maintaining power and type I error. One frequently employed testing strategy in clinical trials with multiple endpoints is the gatekeeping procedure [7, 13]. It arranges multiple hypotheses into a hierarchical order (for example, primary and secondary), and the secondary hypothesis is evaluated only when the primary hypothesis is rejected. The proposed method is different from the gatekeeping procedure in that the multiple endpoints are considered equally important and they are evaluated simultaneously. The proposed method is developed for scenarios where all co-primary endpoints are binary. In future research, we will work on its extension to other types of endpoints such as continuous, count, and event times, as well as mixed types.

APPENDIX A

APPENDIX OF CHAPTER 2

A.1. Derivation of Equation (2.5)

The GEE estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)'$ is solved from equation

$$S_n(\boldsymbol{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{W}_i^{-1} [\boldsymbol{y}_i - \boldsymbol{\mu_i}(\boldsymbol{\beta})] = 0,$$

where $\boldsymbol{D}_i$ is a $m \times 2$ gradient matrix defined as $\boldsymbol{D}_i = \frac{\partial \boldsymbol{\mu_i}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ and $\boldsymbol{W}_i = \text{diag}[\boldsymbol{\mu}_i(\boldsymbol{\beta}) + v\boldsymbol{\mu}_i(\boldsymbol{\beta})^2]$, which is a $m \times m$ diagonal matrix. We can show that

$$S_n(\boldsymbol{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{1 + v\mu_{ij}(\boldsymbol{\beta})} [y_{ij} - \mu_{ij}(\boldsymbol{\beta})] \boldsymbol{Z}_i = 0.$$

With missing data, it becomes

$$S_n(\boldsymbol{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m} \Delta_{ij} \frac{1}{1 + v\mu_{ij}(\boldsymbol{\beta})} [y_{ij} - \mu_{ij}(\boldsymbol{\beta})] \boldsymbol{Z}_i = 0.$$

Without loss of generality, we assume that patients $(1, \cdots, I_0)$ are assigned to the control group, and patients $(I_0 + 1, \cdots, n)$ are assigned to the treatment group.

For the special case $T_1 = ... = T_m = T$, from Equation (1) we have

$$S_n(\boldsymbol{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m} \Delta_{ij} \frac{\mu_{ij}(\boldsymbol{\beta})}{\mu_{ij}(\boldsymbol{\beta}) + v\mu_{ij}(\boldsymbol{\beta})^2} (y_{ij} - \mu_{ij}(\boldsymbol{\beta})) \boldsymbol{Z}_i$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{I_0} \sum_{j=1}^{m} \Delta_{ij} \frac{T\exp(\beta_1)}{T\exp(\beta_1) + vT^2\exp(\beta_1)^2} (y_{ij} - T\exp(\beta_1)) \begin{pmatrix} 1 \\ \\ 0 \end{pmatrix} +$$

$$n^{-\frac{1}{2}} \sum_{i=I_0+1}^{n} \sum_{j=1}^{m} \Delta_{ij} \frac{T\exp(\beta_1+\beta_2)}{T\exp(\beta_1+\beta_2) + vT^2\exp(\beta_1+\beta_2)^2} (y_{ij} - T\exp(\beta_1+\beta_2)) \begin{pmatrix} 1 \\ \\ 1 \end{pmatrix} = \mathbf{0}$$

With some algebra, we have

$$\sum_{i=1}^{I_0} \sum_{j=1}^{m} \Delta_{ij} \frac{T\exp(\beta_1)}{T\exp(\beta_1) + vT^2\exp(\beta_1)^2} (y_{ij} - T\exp(\beta_1)) + \tag{A.1}$$
$$\sum_{i=I_0+1}^{n} \sum_{j=1}^{m} \Delta_{ij} \frac{T\exp(\beta_1+\beta_2)}{T\exp(\beta_1+\beta_2) + vT^2\exp(\beta_1+\beta_2)^2} (y_{ij} - T\exp(\beta_1+\beta_2)) = 0,$$

and

$$\sum_{i=I_0+1}^{n} \sum_{j=1}^{m} \Delta_{ij} \frac{T\exp(\beta_1+\beta_2)}{T\exp(\beta_1+\beta_2) + vT^2\exp(\beta_1+\beta_2)^2} (y_{ij} - T\exp(\beta_1+\beta_2)) = 0. \tag{A.2}$$

Equations (A.1)-(A.2) lead to

$$\sum_{i=1}^{I_0} \sum_{j=1}^{m} \Delta_{ij} (y_{ij} - T\exp(\beta_1)) = 0,$$

which implies that

$$\hat{\beta}_1 = \log \left(\frac{\sum_{i=1}^{I_0} \sum_{j=1}^{m} \Delta_{ij} y_{ij}}{T \sum_{i=1}^{I_0} \sum_{j=1}^{m} \Delta_{ij}} \right),$$

and

$$\hat{\beta}_2 = \log\left(\frac{\sum_{i=I_0+1}^{n}\sum_{j=1}^{m}\Delta_{ij}y_{ij}}{T\sum_{i=I_0+1}^{n}\sum_{j=1}^{m}\Delta_{ij}}\right) - \log\left(\frac{\sum_{i=1}^{I_0}\sum_{j=1}^{m}\Delta_{ij}y_{ij}}{T\sum_{i=1}^{I_0}\sum_{j=1}^{m}\Delta_{ij}}\right).$$

It follows immediately that

$$\begin{aligned}
\hat{\beta}_1 &= \log\left(\frac{\sum_{i=1}^{n}(1-r_i)\sum_{j=1}^{m}\Delta_{ij}y_{ij}}{T\sum_{i=1}^{n}(1-r_i)\sum_{j=1}^{m}\Delta_{ij}}\right), \\
\hat{\beta}_2 &= \log\left(\frac{\sum_{i=n}^{n}r_i\sum_{j=1}^{m}\Delta_{ij}y_{ij}}{\sum_{i=n1}^{n}r_i\sum_{j=1}^{m}\Delta_{ij}}\right) - \log\left(\frac{\sum_{i=1}^{n}(1-r_i)\sum_{j=1}^{m}\Delta_{ij}y_{ij}}{\sum_{i=1}^{n}(1-r_i)\sum_{j=1}^{m}\Delta_{ij}}\right).
\end{aligned}$$

A.2. Derivation of Equation (2.13)

To facilitate the derivation of Equation (2.13), we first reparameterize the model as

$$\log(\mu_{ij}) = \log(T_j) + b_1 + b_2 r_i + b_3(t_{ij} - \bar{t}_i) + b_4 r_i (t_{ij} - \bar{t}_i).$$

It is clear that $b_4 = \beta_4$. Therefore, $H_0 : \beta_4 = 0$ vs $H_1 : \beta_4 \neq 0$ is equivalent to $H_0 : b_4 = 0$ vs $H_1 : b_4 \neq 0$.

Under the reparameterization, we can show that

$$\begin{aligned}
\boldsymbol{A} &= E\left[\sum_{j=1}^{m} \Delta_{ij} \frac{\mu_{ij}(\boldsymbol{b})}{1 + v\mu_{ij}(\boldsymbol{b})} \begin{pmatrix} 1 & r_i & t_j - \bar{t}_i & r_i(t_j - \bar{t}_i) \\ r_i & r_i^2 & r_i(t_j - \bar{t}_i) & r_i^2(t_j - \bar{t}_i) \\ t_j - \bar{t}_i & r_i(t_j - \bar{t}_i) & (t_j - \bar{t}_i)^2 & r_i(t_j - \bar{t}_i)^2 \\ r_i(t_j - \bar{t}_i) & r_i^2(t_j - \bar{t}_i) & r_i(t_j - \bar{t}_i)^2 & r_i^2(t_j - \bar{t}_i)^2 \end{pmatrix}\right] \\
&= (1 - \bar{r}) \sum_{j=1}^{m} \delta_j \eta_{1j}^2 \begin{pmatrix} 1 & 0 & t_j - \bar{t}_1 & 0 \\ 0 & 0 & 0 & 0 \\ t_j - \bar{t}_1 & 0 & (t_j - \bar{t}_1)^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&+ \bar{r} \sum_{j=1}^{m} \delta_j \eta_{2j}^2 \begin{pmatrix} 1 & 1 & t_j - \bar{t}_2 & t_j - \bar{t}_2 \\ 1 & 1 & t_j - \bar{t}_2 & t_j - \bar{t}_2 \\ t_j - \bar{t}_2 & t_j - \bar{t}_2 & (t_j - \bar{t}_2)^2 & (t_j - \bar{t}_2)^2 \\ t_j - \bar{t}_2 & t_j - \bar{t}_2 & (t_j - \bar{t}_2)^2 & (t_j - \bar{t}_2)^2 \end{pmatrix}.
\end{aligned}$$

Given $\bar{t}_1 = \frac{\sum_{j=1}^m \delta_j \eta_{1j}^2 t_j}{\sum_{j=1}^m \delta_j \eta_{1j}^2}$, $\bar{t}_2 = \frac{\sum_{j=1}^m \delta_j \eta_{2j}^2 t_j}{\sum_{j=1}^m \delta_j \eta_{2j}^2}$, we have

$$\boldsymbol{A} = \begin{pmatrix} a_1 & a_2 & 0 & 0 \\ a_2 & a_2 & 0 & 0 \\ 0 & 0 & a_3 & a_4 \\ 0 & 0 & a_4 & a_4 \end{pmatrix},$$

where $a_1 = (1-\bar{r}) \sum_{j=1}^m \delta_j \eta_{1j}^2 + \bar{r} \sum_{j=1}^m \delta_j \eta_{2j}^2$, $a_2 = \bar{r} \sum_{j=1}^m \delta_j \eta_{2j}^2$, $a_3 = (1-\bar{r}) \sum_{j=1}^m \delta_j \eta_{1j}^2 (t_j - \bar{t}_1)^2 + \bar{r} \sum_{j=1}^m \delta_j \eta_{2j}^2 (t_j - \bar{t}_2)^2$, and $a_4 = \bar{r} \sum_{j=1}^m \delta_j \eta_{2j}^2 (t_j - \bar{t}_2)^2$.

Similarly, we have

$$\boldsymbol{V}_1 = \sum_{j=1}^m \sum_{j'=1}^m \delta_{jj'} \rho_{jj'} \eta_{1j} \eta_{1j'} \begin{pmatrix} 1 & 0 & t_{j'} - \bar{t}_1 & 0 \\ 0 & 0 & 0 & 0 \\ t_j - \bar{t}_1 & 0 & (t_j - \bar{t}_1)(t_{j'} - \bar{t}_1) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and

$$\boldsymbol{V}_2 = \sum_{j=1}^m \sum_{j'=1}^m \delta_{jj'} \rho_{jj'} \eta_{2j} \eta_{2j'} \begin{pmatrix} 1 & 1 & t_{j'} - \bar{t}_2 & t_{j'} - \bar{t}_2 \\ 1 & 1 & t_{j'} - \bar{t}_2 & t_{j'} - \bar{t}_2 \\ t_j - \bar{t}_2 & t_j - \bar{t}_2 & (t_j - \bar{t}_2)(t_{j'} - \bar{t}_2) & (t_j - \bar{t}_2)(t_{j'} - \bar{t}_2) \\ t_j - \bar{t}_2 & t_j - \bar{t}_2 & (t_j - \bar{t}_2)(t_{j'} - \bar{t}_2) & (t_j - \bar{t}_2)(t_{j'} - \bar{t}_2) \end{pmatrix}.$$

With some matrix algebra, we can show that

$$\sigma_4^2 = \frac{\sum_{j=1}^m \sum_{j'=1}^m \delta_{jj'}\rho_{jj'}\eta_{1j}\eta_{1j'}(t_j - \bar{t}_1)(t_{j'} - \bar{t}_1)}{(1-\bar{r})(\sum_{j=1}^m \delta_j \eta_{1j}^2 (t_j - \bar{t}_1)^2)^2} + \frac{\sum_{j=1}^m \sum_{j'=1}^m \delta_{jj'}\rho_{jj'}\eta_{2j}\eta_{2j'}(t_j - \bar{t}_2)(t_{j'} - \bar{t}_2)}{\bar{r}(\sum_{j=1}^m \delta_j \eta_{2j}^2 (t_j - \bar{t}_2)^2)^2}.$$

APPENDIX B

APPENDIX OF CHAPTER 3

B.1. Derivation of Equations (3.12) and (3.13)

We can write $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ as

$$\boldsymbol{V}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \sum_{m \in \mathcal{M}} g(m) E \left[\sum_{j=1}^{m} \sum_{j'=1}^{m} (y_{ij} - \mu_{ij})(y_{ij'} - \mu_{ij'}) \right],$$

and

$$\boldsymbol{V}_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \sum_{m \in \mathcal{M}} g(m) E \left[\sum_{j=1}^{m} \sum_{j'=1}^{m} (y_{ij} - \mu_{ij})(y_{ij'} - \mu_{ij'}) \right].$$

Note that, for $E\left[\sum_{j=1}^{m}\sum_{j'=1}^{m}(y_{ij}-\mu_{ij})(y_{ij'}-\mu_{ij'})\right]$, we have

$$E\left[\sum_{j=1}^{m}\sum_{j'=1}^{m}(y_{ij}-\mu_{ij})(y_{ij'}-\mu_{ij'})\right] = \sum_{j=1}^{m} E(y_{ij}-\mu_{ij})^2 + 2\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m} E\left[(y_{ij}-\mu_{ij})(y_{ij'}-\mu_{ij'})\right].$$

Based on the law of total covariance [58], we have

$$\sum_{j=1}^{m} E(y_{ij} - \mu_{ij})^2 = \sum_{j=1}^{m} E\left[Var(y_{ij}|t_{ij})\right] + Var\left[E(y_{ij}|t_{ij})\right]. \tag{B.1}$$

For Equation (B.1), if $i \in \mathcal{C}$,

$$\begin{aligned}\sum_{j=1}^{m} E\left[Var(y_{ij}|t_{ij})\right] + Var\left[E(y_{ij}|t_{ij})\right] &= \sum_{i=1}^{m} E(\theta\mu_1 t_{ij}) + Var(\mu_1 t_{ij}) \\ &= m(\theta\mu_1\gamma_t + \mu_1^2\epsilon_t^2).\end{aligned}$$

Similarly, if $i \in \mathcal{T}$,

$$\sum_{j=1}^{m} E\left[Var(y_{ij}|t_{ij})\right] + Var\left[E(y_{ij}|t_{ij})\right] = m(\theta\mu_2\gamma_t + \mu_2^2\epsilon_t^2).$$

Furthermore, for $2\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m} E\left[(y_{ij} - \mu_{ij})(y_{ij'} - \mu_{ij'})\right]$, we have

$$\begin{aligned}2\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m} E\left[(y_{ij} - \mu_{ij})(y_{ij'} - \mu_{ij'})\right] = 2\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m} & E\{Cov\left[(y_{ij} - \mu_{ij}), (y_{ij'} - \mu_{ij'})|t_{ij}, t_{ij'}\right]\} \\ & + Cov\left[E(y_{ij} - \mu_{ij}|t_{ij}), E(y_{ij'} - \mu_{ij'}|t_{ij'})\right]. \end{aligned} \tag{B.2}$$

Clearly, for Equation (B.2), if $i \in \mathcal{C}$, we have

$$\begin{aligned}2\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m} E\left[(y_{ij} - \mu_{ij})(y_{ij'} - \mu_{ij'})\right] &= 2\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m} \rho\theta\mu_1 E(\sqrt{t_{ij}})E(\sqrt{t_{ij'}}) \\ &= m(m-1)\rho\theta\mu_1\kappa_t^2.\end{aligned}$$

For $i \in \mathcal{T}$, we have

$$2\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m} E\left[(y_{ij}-\mu_{ij})(y_{ij'}-\mu_{ij'})\right] = m(m-1)\rho\theta\mu_2\kappa_t^2.$$

Putting the pieces together, we have if $i \in \mathcal{C}$,

$$E\left[\sum_{j=1}^{m}\sum_{j'=1}^{m}(y_{ij}-\mu_{ij})(y_{ij'}-\mu_{ij'})\right] = m(\theta\mu_1\gamma_t+\mu_1^2\epsilon_t^2)+m(m-1)\rho\theta\mu_1\kappa_t^2,$$

and for $i \in \mathcal{T}$,

$$E\left[\sum_{j=1}^{m}\sum_{j'=1}^{m}(y_{ij}-\mu_{ij})(y_{ij'}-\mu_{ij'})\right] = m(\theta\mu_2\gamma_t+\mu_2^2\epsilon_t^2)+m(m-1)\rho\theta\mu_2\kappa_t^2.$$

B.2. Proof of $N_t < N$

We can reparameterize t_{ij} as t_{ij}/t^* and multiply μ_1 and μ_2 by t^* to standardize patient's follow-up time to be between $(0,1]$. Therefore, without loss of generality, we assume that $t^*=1$.

To show $N_t < N$, we first re-write N_t as

$$N_t = \frac{\eta_m\theta[(1-\bar{r})\mu_1+\bar{r}\mu_2]+(\tau_m^2+\eta_m^2-\eta_m)\rho\theta[(1-\bar{r})\mu_1+\bar{r}\mu_2]}{(1-\bar{r})\bar{r}\mu_1\mu_2\eta_m^2 t^*}.$$

Now it is equivalent to show that

$$\begin{aligned}\frac{\eta_m\theta\gamma_t[(1-\bar{r})\mu_1+\bar{r}\mu_2]+\eta_m\epsilon_t^2\mu_1\mu_2+(\tau_m^2+\eta_m^2-\eta_m)\rho\theta\kappa_t^2[(1-\bar{r})\mu_1+\bar{r}\mu_2]}{(1-\bar{r})\bar{r}\mu_1\mu_2\eta_m^2\gamma_t^2}\\ > \frac{\eta_m\theta[(1-\bar{r})\mu_1+\bar{r}\mu_2]+(\tau_m^2+\eta_m^2-\eta_m)\rho\theta[(1-\bar{r})\mu_1+\bar{r}\mu_2]}{(1-\bar{r})\bar{r}\mu_1\mu_2\eta_m^2 t^*}. \quad \text{(B.3)}\end{aligned}$$

Since t^* is the upper-bound of the patient's follow-up time, given the possibility of drop-out, we must have $\gamma_t < t^*$. Based on the structure of Equation (B.3), to show Equation (B.3), it is equivalent to show that

$$\frac{(\tau_m^2 + \eta_m^2 - \eta_m)\rho\theta\kappa_t^2[(1-\bar{r})\mu_1 + \bar{r}\mu_2]}{(1-\bar{r})\bar{r}\mu_1\mu_2\eta_m^2\gamma_t^2} > \frac{(\tau_m^2 + \eta_m^2 - \eta_m)\rho\theta[(1-\bar{r})\mu_1 + \bar{r}\mu_2]}{(1-\bar{r})\bar{r}\mu_1\mu_2\eta_m^2 t^*},$$

which is equivalent to show that $\frac{\kappa_t^2}{\gamma_t^2} > \frac{1}{t^*}$. Since $0 < t_{ij} \leq 1$, we must have $\kappa_t > \gamma_t$, which implies that $\frac{\kappa_t^2}{\gamma_t^2} > 1 = \frac{1}{t^*}$.

APPENDIX C

APPENDIX OF CHAPTER 4

C.1. Proof of Theorem 4.1

The proof is based on an important theorem (Theorem 4.3.6) by Tong [57].

Theorem C.1 *Let Σ denotes a $L \times L$ positive definite matrix, and assume $\boldsymbol{X}$ to have a density function $f(\boldsymbol{x})$ of the form*

$$f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-1/2} g(\boldsymbol{x}^{'} \boldsymbol{\Sigma}^{-1} \boldsymbol{x}), \tag{C.1}$$

where the function $g(\cdot)$ satisfies

$$\int_0^{\infty} r^{L-1} g(r^2)\, dr < \infty.$$

Let $\boldsymbol{P} = (p_{ij})$ and $\boldsymbol{T} = (t_{ij})$ be two $L \times L$ positive definite matrices. If $p_{ij} \geq t_{ij}$ holds for all i and j, then

$$P_{\boldsymbol{\Sigma}=\boldsymbol{P}}[\cap_{i=1}^{L}\{X_i \leq a_i\}] \geq P_{\boldsymbol{\Sigma}=\boldsymbol{T}}[\cap_{i=1}^{L}\{X_i \leq a_i\}] \tag{C.2}$$

holds for every $\boldsymbol{a} = (a_1, ..., a_L)'$. Furthermore, the inequality is strict if $p_{ij} > t_{ij}$ holds for some i and j and if the support of f is unbounded.

First note that, among many other distributions, the multivariate normal and multivariate t-distributions are of the form (C.1).

Without loss of generality, we give the proof for ρ_0^k. The conclusions for $\rho_1^{kk'}$ and $\rho_2^{kk'}$ can be shown using similar arguments.

Let $k = k^*$. Equation (4.10) implies that $\sigma_{2k^*}^2$ is an increasing function of $\rho_0^{k^*}$, and $\mathbf{\Phi}(k^*, k') = \frac{\sigma_{2k^*k'}^2}{\sigma_{2k^*}\sigma_{2k'}}$ is a decreasing function of $\rho_0^{k^*}$. Let $\boldsymbol{\eta}(k^*) = \frac{N^{1/2}\beta_{2k^*}}{\sigma_{2k^*}}$ denote the k^*th element of $\boldsymbol{\eta}$, which can be shown to be a decreasing function of $\rho_0^{k^*}$. Let τ_1 and τ_2 be two permissible values for $\rho_0^{k^*}$, with $\tau_1 < \tau_2$. Let $\boldsymbol{P} = (p_{kk'})$ be the positive definite matrix $\mathbf{\Phi}$ with $\rho_0^{k^*} = \tau_1$ and $\boldsymbol{T} = (t_{kk'})$ be the positive definite matrix $\mathbf{\Phi}$ with $\rho_0^{k^*} = \tau_2$, with all other parameters fixed. Then we have $p_{k^*k'} > t_{k^*k'}$ for all $k' \neq k^*$ and $p_{kk'} = t_{kk'}$ otherwise. Also, we have $\boldsymbol{\eta}^{\tau_1}(k^*) > \boldsymbol{\eta}^{\tau_2}(k^*)$ and $\boldsymbol{\eta}^{\tau_1}(k) = \boldsymbol{\eta}^{\tau_2}(k)$ for $k \neq k^*$.

Equation (C.2) also implies that

$$P_{\mathbf{\Sigma}=\boldsymbol{P}}[\cap_{i=1}^{L}\{X_i \geq a_i\}] \geq P_{\mathbf{\Sigma}=\boldsymbol{T}}[\cap_{i=1}^{L}\{X_i \geq a_i\}], \tag{C.3}$$

which can be simply shown by replacing $\boldsymbol{X}$ and $\boldsymbol{a}$ by $-\boldsymbol{X}$ and $-\boldsymbol{a}$, respectively.

When $\rho_0^{k^*} = \tau_1$, we have

$$\begin{aligned} Power(\mathbf{\Phi} = \boldsymbol{P}) &= Prob_{\mathbf{\Phi}=P}(\bigcap_{k=1}^{K}\{w_k > c\}) \\ &= Prob_{\mathbf{\Phi}=P}(\bigcap_{k=1}^{K}\{z_k > c - \boldsymbol{\eta}^{\tau_1}(k)\}), \end{aligned}$$

where $\boldsymbol{Z} = (z_1, ..., z_K)'$ has zero mean(location) and correlation(shape) matrix $\boldsymbol{P}$. Equation (C.3) implies that

$$
\begin{aligned}
Prob_{\boldsymbol{\Phi}=\boldsymbol{P}}(\bigcap_{k=1}^{K}\{z_k > c - \boldsymbol{\eta}^{\tau_1}(k)\}) &> Prob_{\boldsymbol{\Phi}=\boldsymbol{T}}(\bigcap_{k=1}^{K}\{z_k > c - \boldsymbol{\eta}^{\tau_1}(k)\}) \\
&> Prob_{\boldsymbol{\Phi}=\boldsymbol{T}}(\bigcap_{k=1}^{K}\{z_k > c - \boldsymbol{\eta}^{\tau_2}(k)\}) \\
&= Power(\boldsymbol{\Phi} = \boldsymbol{T}).
\end{aligned}
$$

Bibliography

[1] Ahn, C., Heo, M. and Zhang, S. [2014], *Sample size calculations for clustered and longitudinal outcomes in clinical research*, Chapman and Hall/CRC.

[2] Albert, P. S. [1999], 'Longitudinal data analysis (repeated measures) in clinical trials', *Statistics in Medicine* **18**(13), 1707–1732.

[3] Amatya, A., Bhaumik, D. and Gibbons, R. D. [2013], 'Sample size determination for clustered count data', *Statistics in medicine* **32**(24), 4162–4179.

[4] Amatya, A. and Demirtas, H. [2015], 'Multiord: An R package for generating correlated ordinal data', *Communications in Statistics-Simulation and Computation* **44**(7), 1683–1691.

[5] Asendorf, T., Henderson, R., Schmidli, H. and Friede, T. [2017], 'Modelling and sample size reestimation for longitudinal count data with incomplete follow up', *Statistical Methods in Medical Research* p. 0962280217715664.

[6] Bauer, P. [1991], 'Multiple testing in clinical trials', *Statistics in medicine* **10**(6), 871–890.

[7] Bauer, P., Röhmel, J., Maurer, W. and Hothorn, L. [1998], 'Testing strategies in multi-dose experiments including active control', *Statistics in Medicine* **17**(18), 2133–2146.

[8] Bloch, D. A., Lai, T. L., Su, Z. and Tubert-Bitter, P. [2007], 'A combined superiority and non-inferiority approach to multiple endpoints in clinical trials', *Statistics in Medicine* **26**(6), 1193–1207.

[9] Calverley, P. M., Rabe, K. F., Goehring, U.-M., Kristiansen, S., Fabbri, L. M., Martinez, F. J. et al. [2009], 'Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomised clinical trials', *The Lancet* **374**(9691), 685–694.

[10] Chuang-Stein, C., Stryszak, P., Dmitrienko, A. and Offen, W. [2007], 'Challenge of multiple co-primary endpoints: a new approach', *Statistics in Medicine* **26**(6), 1181–1192.

[11] Cox, D. R. [1983], 'Some remarks on overdispersion', *Biometrika* **70**(1), 269–274.

[12] Dean, C. B. [1992], 'Testing for overdispersion in Poisson and binomial regression models', *Journal of the American Statistical Association* **87**(418), 451–457.

[13] Dmitrienko, A. and Tamhane, A. C. [2007], 'Gatekeeping procedures with clinical trialapplications', *Pharmaceutical Statistics: The Journal of Applied Statistics the Pharmaceutical Industry* **6**(3), 171–180.

[14] Efron, B. and Tibshirani, R. J. [1994], *An introduction to the bootstrap*, CRC press.

[15] Eldridge, S. M., Ukoumunne, O. C. and Carlin, J. B. [2009], 'The intra-cluster correlation coefficient in cluster randomized trials: a review of definitions', *International Statistical Review* **77**(3), 378–394.

[16] Erhardt, V. and Czado, C. [2010], 'A method for approximately sampling high-dimensional count variables with pre-specified Pearson correlation', *INFORMS Journal on Computing* .

[17] Frome, E. L. [1983], 'The analysis of rates using poisson regression models', *Biometrics* pp. 665–674.

[18] Gao, F., Earnest, A., Matchar, D. B., Campbell, M. J. and Machin, D. [2015], 'Sample size calculations for the design of cluster randomized trials: a summary of methodology', *Contemporary clinical trials* **42**, 41–50.

[19] Garcia, H. H., Pretell, E. J., Gilman, R. H., Martinez, S. M., Moulton, L. H., Del Brutto, O. H., Herrera, G., Evans, C. A. and Gonzalez, A. E. [2004], 'A trial of antiparasitic treatment to reduce the rate of seizures due to cerebral cysticercosis', *New England Journal of Medicine* **350**(3), 249–258.

[20] Guittet, L., Ravaud, P. and Giraudeau, B. [2006], 'Planning a cluster randomized trial with unequal cluster sizes: practical issues involving continuous outcomes', *BMC medical research methodology* **6**(1), 17.

[21] Hilbe, J. M. [2011], *Negative binomial regression*, Cambridge University Press.

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

[23] Ivers, N., Taljaard, M., Dixon, S., Bennett, C., McRae, A., Taleban, J., Skea, Z., Brehaut, J., Boruch, R., Eccles, M. et al. [2011], 'Impact of consort extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8', *Bmj* **343**, d5886.

[24] 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.

[25] Kong, M., Xu, S., Levy, S. M. and Datta, S. [2015], 'Gee type inference for clustered zero-inflated negative binomial regression with application to dental caries', *Computational statistics & data analysis* **85**, 54–66.

[26] Kordzakhia, G., Siddiqui, O. and Huque, M. F. [2010], 'Method of balanced adjustment in testing co-primary endpoints', *Statistics in Medicine* **29**(19), 2055–2066.

[27] Lafaye de Micheaux, P., Liquet, B., Marque, S. and Riou, J. [2014], 'Power and sample size determination in clinical trials with multiple primary continuous correlated endpoints', *Journal of Biopharmaceutical Statistics* **24**(2), 378–397.

[28] Lake, S., Kammann, E., Klar, N. and Betensky, R. [2002], 'Sample size re-estimation in cluster randomization trials', *Statistics in medicine* **21**(10), 1337–1350.

[29] Lee, J.-H., Schell, M. J. and Roetzheim, R. [2009], 'Analysis of group randomized trials with multiple binary endpoints and small number of groups', *PLOS ONE* **4**(10), e7265.

[30] Li, F., Turner, E. L. and Preisser, J. S. [2018], 'Sample size determination for GEE analyses of stepped wedge cluster randomized trials', *Biometrics* .

[31] Li, H., Wang, L., Wei, L. and Quan, H. [2015], 'Sample size calculation for count data in comparative clinical trials with nonuniform patient accrual and early dropout', *Journal of biopharmaceutical statistics* **25**(1), 1–15.

[32] Li, P. and Redden, D. T. [2015], 'Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes', *Statistics in medicine* **34**(2), 281–296.

[33] Liang, K.-Y. and Zeger, S. L. [1986], 'Longitudinal data analysis using generalized linear models', *Biometrika* **73**(1), 13–22.

[34] Liu, H. and Wu, T. [2005], 'Sample size calculation and power analysis of time-averaged difference', *Journal of Modern Applied Statistical Methods* **4**(2), 9.

[35] 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.

[36] Madsen, L. and Dalthorp, D. [2007], 'Simulating correlated count data', *Environmental and Ecological Statistics* **14**(2), 129–148.

[37] Manatunga, A. K., Hudgens, M. G. and Chen, S. [2001], 'Sample size estimation in cluster randomized studies with varying cluster size', *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **43**(1), 75–86.

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

[39] Murray, D. M. [1998], *Design and analysis of group-randomized trials*, Vol. 29, Monographs in Epidemiology & B.

[40] Murray, D. M., Varnell, S. P. and Blitstein, J. L. [2004], 'Design and analysis of group-randomized trials: a review of recent methodological developments', *American journal of public health* **94**(3), 423–432.

[41] Offen, W., Chuang-Stein, C., Dmitrienko, A., Littman, G., Maca, J., Meyerson, L., Muirhead, R., Stryszak, P., Baddy, A., Chen, K. et al. [2007], 'Multiple co-primary endpoints: medical and statistical solutions: a report from the multiple endpoints expert team of the pharmaceutical research and manufacturers of america', *Drug Information Journal* **41**(1), 31–46.

[42] Ogungbenro, K. and Aarons, L. [2010], 'Sample size/power calculations for population pharmacodynamic experiments involving repeated-count measurements', *Journal of Biopharmaceutical Statistics* **20**(5), 1026–1042.

[43] Pan, W. [2001], 'Akaike's information criterion in generalized estimating equations', *Biometrics* **57**(1), 120–125.

[44] Patel, H. I. and Rowe, E. [1999], 'Sample size for comparing linear growth curves', *Journal of Biopharmaceutical Statistics* **9**(2), 339–350.

[45] Patsopoulos, N. A. [2011], 'A pragmatic view on pragmatic trials', *Dialogues in clinical neuroscience* **13**(2), 217.

[46] Piegorsch, W. W. [1990], 'Maximum likelihood estimation for the negative binomial dispersion parameter', *Biometrics* pp. 863–867.

[47] Preisser, J. S., Lu, B. and Qaqish, B. F. [2008], 'Finite sample adjustments in estimating equations and covariance estimators for intracluster correlations', *Statistics in Medicine* **27**(27), 5764–5785.

[48] Ranstam, J. and Cook, J. [2016], 'Statistical models: an overview', *British Journal of Surgery* **103**(8).

[49] Raudenbush, S. W. [1997], 'Statistical analysis and optimal design for cluster randomized trials.', *Psychological Methods* **2**(2), 173.

[50] Roberts, C. and Roberts, S. A. [2005], 'Design and analysis of clinical trials with clustering effects due to treatment', *Clinical Trials* **2**(2), 152–162.

[51] Sherman, M. and Cessie, S. l. [1997], 'A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models', *Communications in Statistics-Simulation and Computation* **26**(3), 901–925.

[52] Sozu, T., Sugimoto, T. and Hamasaki, T. [2010], 'Sample size determination in clinical trials with multiple co-primary binary endpoints', *Statistics in Medicine* **29**(21), 2169–2179.

[53] Sozu, T., Sugimoto, T. and Hamasaki, T. [2012], 'Sample size determination in clinical trials with multiple co-primary endpoints including mixed continuous and binary variables', *Biometrical Journal* **54**(5), 716–729.

[54] Tang, Y. [2017], 'Sample size for comparing negative binomial rates in noninferiority and equivalence trials with unequal follow-up times', *Journal of Biopharmaceutical Statistics* pp. 1–17.

[55] Thall, P. F. and Vail, S. C. [1990], 'Some covariance models for longitudinal count data with overdispersion', *Biometrics* pp. 657–671.

[56] Thanarajasingam, G., Atherton, P. J., Novotny, P. J., Loprinzi, C. L., Sloan, J. A. and Grothey, A. [2016], 'Longitudinal adverse event assessment in oncology clinical trials: the Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254', *The Lancet Oncology* **17**(5), 663–670.

[57] Tong, Y. L. [2014], *Probability inequalities in multivariate distributions*, Academic Press.

[58] Walker, H. M. and Lev, J. [1953], *Statistical inference*, Oxford And IBH Publishing Co.; Calcutta.

[59] Wang, J., Zhang, S. and Ahn, C. [2017], 'Power analysis for stratified cluster randomisation trials with cluster size being the stratifying factor', *Statistical Theory and Related Fields* **1**(1), 121–127.

[60] Wang, J., Zhang, S. and Ahn, C. [2018], 'Sample size calculation for count outcomes in cluster randomization trials with varying cluster sizes', *Communications in Statistics-Theory and Methods* pp. 1–9.

[61] Zeileis, A., Kleiber, C. and Jackman, S. [2008], 'Regression models for count data in R', *Journal of Statistical Software* **27**(8), 1–25.

[62] Zhang, S. and Ahn, C. [2012], 'Sample size calculation for time-averaged differences in the presence of missing data', *Contemporary Clinical Trials* **33**(3), 550–556.

[63] Zhu, H. and Lakkis, H. [2014], 'Sample size calculation for comparing two negative binomial rates', *Statistics in Medicine* **33**(3), 376–387.