Lifetime Data Analysis

Design and Analysis of Individually Randomized Group-Treatment Trials with Time to Event Outcomes

--Manuscript Draft--

Manuscript Number:	LIDA-D-24-00118
Full Title:	Design and Analysis of Individually Randomized Group-Treatment Trials with Time to Event Outcomes
Article Type:	General Manuscript Submission
Keywords:	Design effect; Group treatment; Intracluster correlation coefficient; Log-rank test; Optimal design; Random cluster size; Sample size formula
Corresponding Author:	Sin-Ho Jung, PhD Duke University School of Medicine Durham, UNITED STATES OF AMERICA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Duke University School of Medicine
Corresponding Author's Secondary Institution:	
First Author:	Sin-Ho Jung, PhD
First Author Secondary Information:	
Order of Authors:	Sin-Ho Jung, PhD
Order of Authors Secondary Information:	
Funding Information:	

Noname manuscript No.
(will be inserted by the editor)

Design and Analysis of Individually Randomized Group-Treatment Trials with Time to Event Outcomes

Sin-Ho Jung

Received: date / Accepted: date

Declarations

Funding: The authors did not receive support from any organization for the submitted work.

Conflict of interest/Competing interests: The author declares that he has no conflict of interest.

Availability of data and material: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability We developed our method under Fortran.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Sin-Ho Jung
Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, 27705, USA
E-mail: sinho.jung@duke.edu
ORCID: 0000-0002-1473-7236

Design and Analysis of Individually Randomized Group-Treatment Trials with Time to Event Outcomes

Summary: In a typical individually randomized group-treatment (IRGT) trial, subjects are randomized between a control arm and an experimental arm. While the subjects randomized to the control arm are treated individually, those in the experimental arm are assigned to one of clusters for group treatment. By sharing some common frailties, the outcomes of subjects in the same groups tend to be dependent, whereas those in the control arm are independent. In this paper, we consider IRGT trials with time to event outcomes. We modify the two-sample log-rank test to compare the survival data from TRGT trials, and derive its sample size formula. The proposed sample size formula requires specification of marginal survival distributions, bivariate survival distribution for the experimental arm, cluster size distribution, and accrual period or accrual rate together with additional follow-up period. In a sample size calculation, the cluster sizes are given and the number of clusters may be calculated or the number of clusters is pre-determined at the study open and calculate the required accrual period to determine the cluster sizes. Simulations and real data example show that the proposed test statistic controls the type I error rate and the formula provides accurately powered sample sizes. We also propose optimal designs minimizing the total sample size or the total cost when the cost per subject is different between two arms.

KEY WORDS: Design effect; Group treatment; Intracluster correlation coefficient; Log-rank test; Optimal design; Random cluster size; Sample size formula

1 Introduction

In a traditional randomized trial, patients are randomized between a control and an experimental arms and individually treated. In this case, their outcome data are independent, so that standard test statistics are used to compare the resulting data between two treatment arms. Their sample size methods are well developed for most type of outcome variables.

Often, small groups of individuals, called clusters, are randomized between treatment arms. Typically, clusters are families, classes, communities, surgeons operating patients, and so on. Such trials are called cluster randomization trials (CRTs). The subjects in each cluster share common frailties, so that their outcomes tend to be positively correlated. Two sample testing methods for CRT have been proposed by many statisticians, including Donner and Banting (1988) and Ahn, Jung, Kang (2003) for binary data, and Jung and Jeong (2003) for time to event endpoints. Sample size methods for these statistical tests are proposed by Kang, Ahn, Jung (2003) for binary data and Li and Jung (2020) for time to event outcome. The methods for clustered survival data under cluster randomization are extended to those under subunit randomization by Jeong and Jung (2006) for weighted rank tests and Li and Jung (2022) for their sample size calculation.

Individually randomized group-treatment (IRGT) trials are composite of traditional independent subject randomization and cluster randomization trials. In an IRGT trial, the control arm is to treat patients individually, whereas the experimental arm is to treat patients using a group training, education, or treatment to increase the treatment effect by close interactions with patients. As a result, the outcome data of the control arm are independent as in traditional trials, but those in the experimental arm are correlated within each group, called cluster as in cluster randomization trials. Hence, two arms in IRGT trials have different dependency structures. Unlike stan-

dard cluster randomization trials, clusters of IRGT trials are usually organized after randomization. But statistically, they have identical statistical issues between the two types of trials, i.e. accounting for the dependency within each cluster.

The number of IRGT trials is growing[6], but design and analysis methods for them are still in need. Turner et al. (2017A, B) discuss statistical issues and analysis methods for IRGT trials that have been developed by then. Roberts and Roberts[9] propose a sample size method for continuous outcomes and Moerbeek and Wong (2008) propose sample size methods for IRGT trials with binary and continuous outcome variables, and Wang, Turner, and Li (2024) propose a sample size method for IRGT trials collecting longitudinal data. These methods assume constant cluster sizes. Candel and Van Breukelen (2009) investigate the relative efficiency of test statistics between constant cluster size cases and varying cluster size cases based on a random effect model that can cover continuous outcomes and binary outcomes for both cluster randomization and subunit randomization trials (SRTs), but do not propose a sample size formula for these types of trials.

Usually, the number of clusters is given in advance and cluster sizes grow during accrual period. Even when we assume identical cluster sizes at the time of study design, the attained cluster sizes may vary due to the random allocation to different clusters post-randomization. As shown by Candel and Van Breukelen (2009), the power of the statistical tests depends on the mean and variance of cluster size distribution for trials with clusters, so that it is important to accurately reflect the expected cluster size distribution in sample size calculations.

In this paper, we propose a modified log-rank test and its sample size calculation method for IRGT trials with time to event endpoints by modifying the log-rank test of Jung and Jeong[3] and its sample size method of Li and Jung[5] that are proposed for CRTs. The proposed log-rank test is so robust that it does not require any assumptions

except that survival time and censoring time are independent for each individual. The censoring times of subjects in each cluster of experimental arm may be correlated or even common. Cluster sizes may be either constant or random. The sample size of the log-rank test modified for IRGT trials depends on the marginal survival and censoring distributions of two arms, bivariate survival and censoring distributions and cluster size distribution of experimental arm. The bivariate censoring distribution can be fully specified by the univariate censoring distribution for popular trial designs. The sample size method allows the cases where the number of clusters is given and cluster sizes grow during accrual period, or cluster sizes are given (whether constant or random) and the number of clusters grows over accrual period. We also propose optimal allocation proportions minimizing the total sample size in the presence of intra-cluster dependency or the total cost when two arms have different cost for treatment per subject. Simulations are conducted to evaluate the performance of the proposed testing and sample size methods. The sample size method is demonstrated with an example mimicking a real IRGT trial.

2 Rank Test for IRGT Trials

Suppose that arm 1 is a control arm and arm 2 is an experimental arm, n_k patients are randomized to arm $k = (1, 2)$, and $n = n_1 + n_2$ denotes the total number of patients. Let $p_k = n_k/n$ denote the allocation proportion for arm k. Arm 2 has ν_2 clusters, so that each patient in arm 2 will be assigned to one of the clusters after randomization. Let m_i denote the number of patients assigned to cluster $i(= 1, ..., \nu_2)$, by which we have $n_2 = \sum_{i=1}^{\nu_2} m_i$. For the large sample approximation, we assume that ν_2 is large, i.e. $\nu_2 = O_p(n)$.

$T_{k1}, ..., T_{k,n_k}$ are survival times of patient i in arm k with survival function (SF) $S_k(t) = P(T_{ki} \geq t)$ and cumulative hazard function $\Lambda_k(t) = -\log S_k(t)$. We want to

test $H_0 : \Lambda_1(t) = \Lambda_2(t)$ against $H_1 : \Lambda_1(t) \neq \Lambda_2(t)$.

Survival times are subject to censoring, so that we observe (X_{ki}, δ_{ki}) from patient i in arm k, where X_{ki} is the minimum of survival time T_{ki} and censoring time C_{ki} and δ_{ki} is the event indicator taking 1 if the patient has an event and 0 otherwise. We assume that censoring times are independent of survival times.

For patient i in arm k, let $N_{ki}(t) = \delta_{ki} I(X_{ki} \leq t)$ and $Y_{ki}(t) = I(X_{ki} \geq t)$ denote the event and at-risk processes, respectively, $N_k(t) = \sum_{i=1}^{n_k} N_{ki}(t)$, $N(t) = N_1(t) + N_2(t)$, $Y_k(t) = \sum_{i=1}^{n_k} Y_{ki}(t)$, and $Y(t) = Y_1(t) + Y_2(t)$. The weighted rank statistic for testing $H_0 : \Lambda_1(t) = \Lambda_2(t)$ against $H_1 : \Lambda_1(t) \neq \Lambda_2(t)$ is given as

$$W = \sqrt{n} \int_0^\infty H(t)\{d\hat{\Lambda}_1(t) - d\hat{\Lambda}_2(t)\} \tag{1}$$

where $\hat{\Lambda}_k(t) = \int_0^t Y_k^{-1}(s) dN_k(s)$ is the Nelson-Aalen estimator (Nelson 1969, Aalen 1978) of $\Lambda_k(t)$ and $H(t)$ is a predictable process of bounded variation that is uniformly convergent to a square integrable function $h(t)$. For example, $H(t) = n^{-1} Y_1(t) Y_2(t)/Y(t)$ for the log-rank statistic (Peto and Peto 1972), $H(t) = n^{-2} Y_1(t) Y_2(t)$ for the Gehan-Wilcoxon test (Gehan 1965), and $H(t) = n^{-1} \hat{S}^-(t) Y_1(t) Y_2(t)/Y(t)$ for the Prentice-Wilcoxon test (Prentice 1978), where $Y(t) = Y_1(t) + Y_2(t)$ and $\hat{S}^-(t)$ is the left-continuous version of the Kaplan-Meier (1958) estimate from the pooled sample $\{(X_{kj}, \delta_{kj}), j = 1, ..., n_k, k = 1, 2\}$. When individuals within clusters in arm 2 are independent, the asymptotic distribution of the test statistics is well developed (Gill 1980, Fleming and Harrington 1991).

Since the patients of arm 2 are clustered after randomization, their survival times will be dependent within each cluster, whereas those in arm 1 stay independent. In order to account for this, we change the notations for arm 2 patients slightly. Let (X_{2ij}, δ_{2ij}) denote the survival data for patient $j(= 1, ..., m_i)$ in cluster $i(= 1, ..., \nu_2)$ of arm 2, and $N_{2ij}(t)$ and $Y_{2ij}(t)$ are redefined from $N_{2i'}(t)$ and $Y_{2i'}$ accordingly. By the new arrangement, we have $N_2(t) = \sum_{i=1}^{\nu_2} \sum_{j=1}^{m_i} N_{2ij}(t)$ and $Y_2(t) = \sum_{i=1}^{\nu_2} \sum_{j=1}^{m_i} Y_{2ij}(t)$.

Under H_0, W is expressed as

$$W = \sqrt{n}\{\sum_{i=1}^{n_1}\int_0^\infty H(t)Y_1^{-1}(t)dM_{1i}(t) - \sum_{i=1}^{\nu_2}\int_0^\infty H(t)Y_2^{-1}(t)dM_{2i}(t)\}$$

where $M_{1i}(t) = \int_0^t\{dN_{1j}(t) - Y_{1j}(t)d\Lambda_1(t)\}$, $M_{2i}(t) = \sum_{j=1}^{m_i} M_{2ij}(t)$, and $M_{2ij}(t) = \int_0^t\{dN_{2ij}(t) - Y_{2ij}(t)d\Lambda_2(t)\}$.

Let $G(t) = P(C_{ki} \geq t)$ denote the SF of the censoring distribution which is common between two arms. Then, $n_k^{-1}Y_k(t)$ and $H(t)$ uniformly converge to $y_k(t) = S_k(t)G(t)$ and $h(t)$, respectively. Read Ying and Wei (1994) about the uniform convergence of $n_2^{-1}Y_2(t)$ with clustered data. Hence, under H_0, we have

$$W = \frac{1}{\sqrt{n}}\Big\{\sum_{i=1}^{n_1}\int_0^\infty \frac{h(t)}{p_1y_1(t)}dM_{1i}(t) - \sum_{i=1}^{\nu_2}\int_0^\infty \frac{h(t)}{p_2y_2(t)}dM_{2i}(t)\Big\} + o_p(1)$$

$$= \frac{1}{\sqrt{n}}(\sum_{i=1}^{n_1}\epsilon_{1i} - \sum_{i=1}^{\nu_2}\epsilon_{2i})$$

where

$$\epsilon_{ki} = \int_0^\infty \frac{h(t)}{p_ky_k(t)}dM_{ki}$$

As in Appendix A, under H_0, W is asymptotically normal with mean 0 and variance σ^2 that can be consistently estimated by

$$\hat{\sigma}^2 = n^{-1}(\sum_{i=1}^{n_1}\hat{\epsilon}_{1i}^2 + \sum_{i=1}^{\nu_2}\hat{\epsilon}_{2i}^2)$$

where $\hat{\epsilon}_{ki}$ are obtained from $\hat{\epsilon}_{ki}$ by replacing $h(t)$, $p_ky_k(t)$, and $\Lambda_k(t)$ with $H(t)$, $n^{-1}Y_k(t)$, and $\hat{\Lambda}(t) = \int_0^t Y^{-1}(t)dN(t)$, the Nelson-Aalen (Nelson 1969, Aalen 1978) estimator from the pooled data, respectively. We reject H_0 if $|W|/\hat{\sigma} > z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$ percentile of $N(0,1)$ distribution.

3 Sample size calculation for the log-rank test

We derive a sample size formula for the log-rank test with $H(t) = Y_1(t)Y_2(t)/\{nY(t)\}$, i.e.

$$W = \frac{1}{\sqrt{n}}\int_0^\infty \Big\{\frac{Y_2(t)}{Y(t)}dN_1(t) - \frac{Y_1(t)}{Y(t)}dN_2(t)\Big\}$$

Sample size formula for other weighted rank tests can be similarly derived.

The power of the log-rank test for IRGT trials depends the distribution of cluster sizes m_i in arm 2, so that we have to specify the cluster size distribution for a sample size calculation. Cluster sizes m_i in arm 2 may be either random or constant. Let $\bar{m} = E(m_i)$ and $\bar{\bar{m}} = E(m_i^2)$ denote the first and second moments, respectively, of cluster size distribution. If cluster sizes are constant with $m_i = m$, then we have $\bar{m} = m$ and $\bar{\bar{m}} = m^2$. Note that $n_2/\nu_2 = \nu_2^{-1} \sum_{i=1}^{\nu_2} m_i$ converges to $\bar{m}$.

For a sample size formula of W, we need to derive the distribution of W under $H_1 : \Lambda_1(t) \neq \Lambda_2(t)$. As shown in Appendix A, the asymptotic distribution of W under H_1 depends on the bivariate distributions of survival times and censoring times between two patients in each cluster of arm 2 as well as their marginal distributions for two arms.

Suppose that, for arm $k(= 1, 2)$, patients have marginal hazard function $\lambda_k(t) = \partial\Lambda_k(t)/\partial t$, and two patients in each cluster of arm 2 have a bivariate SF

$$S_2(t_1, t_2) = P(T_{2ij} \geq t_1, T_{2ij'} \geq t_2)$$

and joint probability density function $f_2(t_1, t_2) = \partial^2 S_2(t_1, t_2)/\partial t_1 \partial t_2$. Further, for arm 2, we define the joint hazard function

$$\lambda_2(t_1, t_2) = f_2(t_1, t_2)/S_2(t_1, t_2)$$

and two conditional hazard functions

$$\lambda_{2(1|2)}(t_1, t_2) = \lim_{\epsilon \to 0} \frac{P(t_1 \leq T_{2ij} < t_1 + \epsilon, T_{2ij'} \geq t_2)}{P(T_{2ij} \geq t_1, T_{2ij'} \geq t_2)} = -\frac{\partial S_2(t_1, t_2)/\partial t_1}{S_2(t_1, t_2)}$$

and

$$\lambda_{2(2|1)}(t_1, t_2) = \lim_{\epsilon \to 0} \frac{P(T_{2ij} \geq t_1, t_2 \leq T_{2ij'} < t_2 + \epsilon)}{P(T_{2ij} \geq t_1, T_{2ij'} \geq t_2)} = -\frac{\partial S_2(t_1, t_2)/\partial t_2}{S_2(t_1, t_2)}$$

for $j \neq j'$. For two subunits $j \neq j'$ in a cluster of arm 2, let $G(t_1, t_2) = P(C_{2ij} \geq t_1, C_{2ij'} \geq t_2)$ denote the bivariate SF of the clustered censoring times.

As $n \to \infty$, $\nu_2^{-1} \sum_{i=1}^{\nu_2} m_i$ and $\nu_2^{-1} \sum_{i=1}^{\nu_2} m_i^2$ converges to $\bar{m}$ and $\bar{\bar{m}}$, respectively. From Appendix A, under H_1, W is asymptotically normal with mean $\omega\sqrt{n}$ and variance

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + (\bar{\bar{m}}/\bar{m} - 1)c_2 = \sigma_1^2 + \sigma_2^2\{1 + (\bar{\bar{m}}/\bar{m} - 1)\rho\}$$

where

$$\omega = p_1 p_2 \int_0^\infty \frac{S_1(t)S_2(t)G(t)}{p_1 S_1(t) + p_2 S_2(t)}\{d\Lambda_1(t) - d\Lambda_2(t)\}$$

$$\sigma_k^2 = p_k p_{k-3}^2 \int_0^\infty \frac{S_k(t)S_{3-k}(t)^2 G(t)}{\{p_1 S_1(t) + p_2 S_2(t)\}^2} d\Lambda_k(t)$$

$$c_2 = p_1^2 p_2 \int_0^\infty \int_0^\infty \frac{S_1(t_1)S_1(t_2)S_2(t_1,t_2)G(t_1,t_2)}{\{p_1 S_1(t_1) + p_2 S_2(t_1)\}\{p_1 S_1(t_2) + p_2 S_2(t_2)\}} dA_2(t_1,t_2)$$

$$dA_2(t_1,t_2) = \{\lambda_2(t_1,t_2) - \lambda_{2(1|2)}(t_1,t_2)\lambda_2(t_2) - \lambda_{2(2|1)}(t_1,t_2)\lambda_2(t_1) + \lambda_2(t_1)\lambda_2(t_2)\}dt_1 dt_2$$

and

$$\rho = \frac{c_2}{\sigma_2^2} = \frac{\int_0^\infty \int_0^\infty \frac{S_1(t_1)S_1(t_2)S_2(t_1,t_2)G(t_1,t_2)}{\{p_1 S_1(t_1)+p_2 S_2(t_1)\}\{p_1 S_1(t_2)+p_2 S_2(t_2)\}} dA_2(t_1,t_2)}{\int_0^\infty \frac{S_1^2(t)S_2(t)G(t)}{\{p_1 S_1(t)+p_2 S_2(t)\}^2} d\Lambda_2(t)}$$

denotes the intracluster correlation coefficient (ICC) of arm 2.

Hence, the power of the log-rank test is given as

$$1 - \beta = P\Big(\frac{|W|}{\hat{\sigma}} \geq z_{1-\alpha/2} | H_1\Big) \approx P\Big(\frac{W - \omega\sqrt{n}}{\sigma} \geq z_{1-\alpha/2} - \frac{\omega\sqrt{n}}{\sigma} | H_1\Big) \tag{2}$$

Since $(W - \omega\sqrt{n})/\sigma$ is $N(0,1)$ under H_1, from (2), we have

$$1 - \beta = \bar{\Phi}\Big(z_{1-\alpha/2} - \frac{\omega\sqrt{n}}{\sigma}\Big) \tag{3}$$

where $\bar{\Phi}(\cdot)$ is the SF of $N(0,1)$ distribution. By solving (3) with respect to n, we obtain the sample size

$$n = \frac{\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\omega^2} \tag{4}$$

From Appendix B, under the nearby alternative hypothesis with where $S_1(t) \approx S_2(t)$, we have $\omega = p_1 p_2 d(\Delta - 1) \approx p_1 p_2 d \log\Delta$ and $\sigma^2 = p_1 p_2 d \times \mathrm{DE}$, so that we have

$$\frac{\sigma^2}{\omega^2} = \frac{\mathrm{DE}}{p_1 p_2 d(\log\Delta)^2}$$

and (4) is simplified to

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{p_1 p_2 d(\log \Delta)^2} \mathrm{DE} \tag{5}$$

where $\Delta = \Lambda_1(t)/\Lambda_2(t)$ is the hazard ratio, $d = p_1 d_1 + p_2 d_2$, $d_k = -\int_0^\infty G(t) dS_k(t)$ is the probability that a patient in arm k experiences an event during the study, and

$$\mathrm{DE} = 1 + p_1 \rho(\bar{\bar{m}}/\bar{m} - 1) \tag{6}$$

is the design effect factor indicating the increase of sample size due to the ICC in arm 2 that is expressed as

$$\rho = d^{-1} \int_0^\infty \int_0^\infty S_2(t_1, t_2) G(t_1, t_2) dA_2(t_1, t_2)$$

If the survival times of arm 2 are independent, then we have $\rho = 0$ and $\mathrm{DE} = 1$, so that $D = n \times d$ in (5) is reduced to the formula for required number of events by Schoenfeld (1983) for the log-rank test with independent survival data. Furthermore, the design effect for CRTs $\mathrm{DE}_c = 1 + \rho(\bar{\bar{m}}/\bar{m} - 1)$ by Li and Jung[5] is larger than that of IRGT trials (6), so that IRGT trials are more powerful than CRTs.

When n patients are recruited, the last cluster of arm 2 may not have been filled up yet. In this case, we propose to randomize patients between two arms until the last cluster of arm 2 is filled in. In (5), the expected number of clusters in arm 2 is $p_2[n/\bar{m}]$, where $[x]$ is the round-up of x. By rounding up the expected number of clusters in arm 2, the total number of patients randomized is likely to be slightly larger than n calculated by (5).

3.1 Distributional Assumptions for Sample Size Calculation

Note that the sample size of the log-rank test modified for IRGT trials depends on the bivariate distributions of censoring and survival times, and cluster size distribution of arm 2. For IRGT trials, however, the bivariate censoring distribution function can be

fully specified by its univariate distribution function as described below. In this section, we consider some survival and censoring distributions for sample size calculations.

3.1.1 Censoring Distribution

In IRGT trials, patients are randomized between two arms, and only those randomized to arm 2 will be clustered for treatment. In this sense, whichever arm they are assigned to, all patients are recruited independently, so that we have $G(t_1, t_2) = G(t_1)G(t_2)$ if there are no loss to follow-up (i.e., administrative censoring only) or dependent loss to follow-up within each cluster. If there exists loss to follow-up, however, the censoring times of patients in arm 2 may be possibly correlated within each cluster. If it is the case, we can model the bivariate censoring distribution for arm 2 accounting for the expected cluster-specific loss to follow-up at the study design stage. In this paper, we consider administrative censoring only, so that $G(t_1, t_2) = G(t_1)G(t_2)$ for arm 2 patients.

If patients are recruited with a constant rate over an accrual period a and followed for additional period b after completion of accrual, then the censoring distribution is $U(b, a + b)$, so that we have

$$G(t) = I(t < a + b) - \frac{t - b}{a} I(b \leq t < a + b)$$

3.1.2 Survival Distribution

For arm 1, we assume exponential distribution with hazard rate λ_1. We may use any multivariate survival distribution for arm 2 for sample size calculation. But, for sample size calculations in this paper, we consider the Clayton and Cuzick's (1985) frailty model with an exponential marginal distribution whose bivariate SF is given as

$$S_2(t_1, t_2) = \{\exp(\lambda_2 t_1/\theta) + \exp(\lambda_2 t_2/\theta) - 1\}^{-\theta}$$

where $\theta = 1/(2\tau) - 1/2$ and $\tau \in [0, 1]$ is the Kendall's tau, 0 meaning no relationship and 1 meaning a perfect relationship between paired survival times. For this joint survival distribution, we have the joint hazard function

$$\lambda_2(t_1, t_2) = \frac{\lambda_2^2(1 + 1/\theta)e^{(t_1+t_2)\lambda_2/\theta}}{(e^{\lambda_2 t_1/\theta} + e^{\lambda_2 t_2/\theta} - 1)^2}$$

and two conditional hazard functions

$$\lambda_{2(1|2)}(t_1, t_2) = \frac{\lambda_2 e^{\lambda_2 t_1/\theta}}{e^{\lambda_2 t_1/\theta} + e^{\lambda_2 t_2/\theta} - 1}$$

$$\lambda_{2(2|1)}(t_1, t_2) = \frac{\lambda_2 e^{\lambda_2 t_2/\theta}}{e^{\lambda_2 t_1/\theta} + e^{\lambda_2 t_2/\theta} - 1}$$

and

$$dA_2(t_1, t_2) = \lambda_2^2\Big\{\frac{(1 + 1/\theta)e^{(t_1+t_2)\lambda_2/\theta}}{(e^{\lambda_2 t_1/\theta} + e^{\lambda_2 t_2/\theta} - 1)^2} - \frac{e^{\lambda_2 t_1/\theta} + e^{\lambda_2 t_2/\theta}}{e^{\lambda_2 t_1/\theta} + e^{\lambda_2 t_2/\theta} - 1} + 1\Big\}dt_1 dt_2 \qquad (7)$$

Hence, for the Clayton and Cuzick's (1985) model with an exponential marginal distribution and independent $U(b, a + b)$ censoring assumption, we have

$$\omega = p_1 p_2(\lambda_1 - \lambda_2)\int_0^{a+b} \frac{e^{-(\lambda_1+\lambda_2)t}(1 - a^{-1}[t - b]_+)}{p_1 e^{-\lambda_1 t} + p_2 e^{-\lambda_2 t}}dt$$

$$\sigma_k^2 = p_k p_{k-3}^2 \lambda_k \int_0^{a+b} \frac{e^{-(\lambda_k+2\lambda_{3-k})t}(1 - a^{-1}[t - b]_+)}{(p_1 e^{-\lambda_1 t} + p_2 e^{-\lambda_2 t})^2}dt$$

and

$$c_2 = p_1^2 p_2 \int_0^{a+b}\int_0^{a+b} \frac{e^{-\lambda_1(t_1+t_2)}\{e^{\lambda_2 t_1/\theta} + e^{\lambda_2 t_2/\theta} - 1\}^{-\theta}}{(p_1 e^{-\lambda_1 t_1} + p_2 e^{-\lambda_2 t_1})(p_1 e^{-\lambda_1 t_2} + p_2 e^{-\lambda_2 t_2})}$$
$$\times(1 - a^{-1}[t_1 - b]_+)(1 - a^{-1}[t_2 - b]_+)dA_2(t_1, t_2)$$

for formula (4), and

$$d_k = 1 - (1 - e^{-a\lambda_k})\frac{e^{-b\lambda_k}}{a\lambda_k}$$

and $\rho = c_2/d$ with

$$c_2 = p_1^2 p_2 \int_0^{a+b}\int_0^{a+b} \frac{(1 - a^{-1}[t_1 - b]_+)(1 - a^{-1}[t_2 - b]_+)}{\{\exp(\lambda_2 t_1/\theta) + \exp(\lambda_2 t_2/\theta) - 1\}^{\theta}}dA_2(t_1, t_2)$$

for formula (5). Here, $[x]_+ = x \times I(x > 0)$ and $dA_2(t_1, t_2)$ in (7) is used for the both sample size formulas.

3.2 When Number of Clusters is Given

So far, we have assumed that clusters and their subunits are recruited over accrual period a. In an IRGT trial, however, all of clusters are recruited before opening the trial, so that number of clusters ν_2 is pre-determined together with their relative sizes. Furthermore, it is easier to estimate the accrual rate r for a new trial based on the number of patients who have visited the clinics participating in the trial recently, rather than to specify an accrual period a. Let γ_i denote the relative size of cluster $i(= 1, ..., \nu_2)$ with $\sum_{i=1}^{\nu_2} \gamma_i = 1$. In this case, $p_k r$ patients will be randomized to arm k during a unit time of accrual period. And, $p_2 r$ patients of arm 2 will be randomly allocated to ν_2 clusters with the proportions of γ_i per unit time period, so that accrual rate of cluster i in arm 2 is $r_i = p_2 r \gamma_i$. So, for an accrual period a, the number of patients in arm 1 will be $n_1 = p_1 ar$ and the number of patients in cluster i of arm 2 will be a function of a given as $m_i(a) = [a \times r_i] = [p_2 ar\gamma_i]$, so that

$$\bar{m}(a) = \nu_2^{-1} \sum_{i=1}^{\nu_2} m_i(a) = \frac{p_2 ar}{\nu_2}$$

and

$$\bar{\bar{m}}(a) = \nu_2^{-1} \sum_{i=1}^{\nu_2} m_i^2(a) = \frac{p_2^2 a^2 r^2 \sum_{i=1}^{\nu_2} \gamma_i^2}{\nu_2}$$

Also, in the left hand size of (5), $n \approx ar$, and $d = d(a)$ and $\rho = \rho(a)$ are functions of unknown parameter a. By plugging these functions of a in (5), we have

$$a \times r = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{p_1 p_2 d(a) (\log \Delta)^2} \mathrm{DE}(a) \tag{8}$$

where $\mathrm{DE}(a) = 1 + p_1 \rho(a) \{\bar{\bar{m}}(a)/\bar{m}(a) - 1\}$. Note that, in $\mathrm{DE}(a)$, $\bar{\bar{m}}(a)/\bar{m}(a) = p_2 ar \sum_{i=1}^{\nu_2} \gamma_i^2$. The required accrual period $a = a^*$ is obtained by solving equation (8), and the total number of patients is $n = a^* \times r$. By rounding up $m_i(a)$ values to $[p_2 ar\gamma_i]$, the final sample size may be slightly increased.

4 Optimal Designs

One may want to find $p = p_1$ minimizing the total sample size n. In (5), sample size n depends on $p = p_1$ through

$$Q_1(p) = \frac{1 + p\rho(\bar{\bar{m}}/\bar{m} - 1)}{p(1-p)}$$

In fact, $d = pd_1 + (1-p)d_2$ in (5) also depends on p, but $d_1 \approx d_2$ under the nearby alternative hypothesis, so that d is approximately free of p. By some simple calculations, we find that $Q_1(p)$ is minimized at

$$p = \frac{-1 + \sqrt{1 + \rho(\bar{\bar{m}}/\bar{m} - 1)}}{\rho(\bar{\bar{m}}/\bar{m} - 1)} \tag{9}$$

As $p = p_1$ changes, however, d and ρ changes a little, so that the p value minimizing n is sightly different from that given in (9). It is easy to numerically find p minimizing n directly without any approximation.

On the other hand, the cost per patient may be different between two arms. In this case, we may want to find the allocation proportion $p = p_1$ minimizing the total cost for the study. Let η_k denote the cost per patient for arm k and $\eta = \eta_2/\eta_1$ the relative cost between two arms. For a given $p = p_1$, the total cost is

$$C(p) = \eta_1 n_1 + \eta_2 n_2 = \eta_1 n\{p + \eta(1-p)\}$$

By plugging n from (5) in this equation and removing the factors free of p, we can find the optimal p minimizing

$$Q_2(p) = \frac{\{p + (1-p)\eta\}\{1 + p\rho(\bar{\bar{m}}/\bar{m} - 1)\}}{p(1-p)\{pd_1 + (1-p)d_2)\}}$$

From $Q_2(p)$, we find that optimal allocation proportion $p = p_1$ depends on c, cluster size distribution, and bivariate survival distribution through ICC ρ of arm 2. We can use a numerical method to identify the p value minimizing $Q_2(p)$.

5 Numerical Studies

5.1 Simulations

Our sample size formula is derived using an asymptotic theory. We conduct extensive simulations to evaluate the finite sample properties of our proposed test statistic and sample size formulas.

Survival times for arm 1 $t_{11}, ..., t_{1,n_1}$ are generated from independent exponential distribution with hazard rate λ_1, i.e. for independent and identically distributed (IID) $U(0,1)$ random variables $u_{11}, ..., n_{1,n_1}$, $t_{1i} = -\lambda_1^{-1} \log u_{1i}$.

We assume that clustered survival times for arm 2 have the Clayton and Cuzick's model with an exponential marginal distribution whose joint SF for a cluster with size m is given as

$$S_2(t_1, ..., t_m) = \{\sum_{j=1}^{m} \exp(\lambda_2 t_j/\theta) - (m-1)\}^{-\theta}$$

For this distribution, the conditional SF of t_j given $(t_1, ..., t_{j-1})$ is

$$S_2(t_j|t_1, ..., t_{j-1}) = \frac{\partial^{j-1} S_2(t_1, ..., t_j)}{\partial t_1 \cdots \partial t_{j-1}} \Big/ \frac{\partial^{j-1} S_2(t_1, ..., t_{j-1})}{\partial t_1 \cdots \partial t_{j-1}}$$

$$= \left(\frac{\sum_{j'=1}^{j} e^{\lambda_2 t_{j'}/\theta} - j + 1}{\sum_{j'=1}^{j-1} e^{\lambda_2 t_{j'}/\theta} - j + 2} \right)^{-\theta - j + 1}$$

For IID $U(0,1)$ random variables $u_1, .., u_m$, survival times $(t_1, ..., t_m)$ for a cluster in arm 2 are sequentially generated by solving the equations $u_1 = S_2(t_1) = \exp(-\lambda_2 t_1)$ and $u_j = S_2(t_j|t_1, ..., t_{j-1})$ for $j \geq 2$ with respect to $t_1, ..., t_m$. That is, starting from $e_0 = 0$ and $t_1 = -\lambda_2^{-1} \log u_1$, calculate e_j and then t_{j+1} using algorithm $e_j = e_{j-1} + \exp(\lambda_2 t_j/\theta)$ and

$$t_{j+1} = \frac{\theta}{\lambda_2} \log\{j - e_j + (e_j - j + 1) u_{j+1}^{-1/(\theta+j)}\}$$

for $1 \leq j \leq m-1$.

In this simulation study, we assume that cluster sizes in arm 2 are fixed at $m_i = 10$ or $m_i = 15$, or random with a discrete uniform distribution between 8 and 12, denoted

as $U(8,12)$ or a discrete uniform distribution between 13 and 17, denoted as $U(13,17)$. Note that $m_i = 10$ and $U(8,12)$ have the same mean $\bar{m} = 10$, but different variances 0 and 2, respectively, and $m_i = 15$ and $U(8,12)$ have the same mean $\bar{m} = 15$, but different variances 0 and 2, respectively. We set $\alpha = 0.05$; $1-\beta = 0.8$, 0.85, or 0.9; $p_1 = p_2 = 1/2$ for balanced allocations; $\lambda_1 = 0.5$, and $\lambda_2 = 0.5$ under H_0 and $\lambda_2 = 0.3$ or 0.35 under H_1. Kendall's $\tau = 0.1$, 0.2, or 0.3; $(a,b) = (3,2)$.

At first, we calculate the required sample size n using (5) under each design setting of $(\alpha, 1-\beta, \lambda_1, \lambda_2, \tau, p_1, a, b)$ and cluster size distribution. Once a sample size is calculated, we generate a sample of IRGT data under H_1 and apply the log-rank test (1) with $H(t) = Y_1(t)Y_2(t)/\{nY(t)\}$. A total of 5,000 samples are generated under the same design setting and the empirical power $1-\hat{\beta}$ is estimated by the proportion of simulation samples rejecting H_0 out of the 5,000 samples. If the empirical power is close to the nominal level, $1-\beta$, then we will conclude that the calculated power is accurately powered. In order to evaluate the finite sample performance of the log-rank statistic, we also generate 5,000 samples of size n under H_0 and estimate the empirical type I error rate $\hat{\alpha}$.

(Table 1 may be placed around here.)

Table 1 summarizes the simulation results. As expected, sample size increases in $1-\beta$, τ, mean cluster size $\bar{m}$ (e.g. $m_i = 10$ vs. $m_i = 15$), and variance of cluster size distribution (e.g. $m_i = 10$ vs. $m_i \sim U(8,12)$), and decreases in hazard ratio $\Delta = \lambda_1/\lambda_2$. The sample sizes maintain the nominal power well overall. If the calculated sample size is small and is slightly larger than a multiple of the cluster size, e.g. $n = 251$ for $(1-\beta, m_i, \lambda_2, \tau) = (0.8, 10, 0.3, 0.1)$, the empirical power can be higher than the nominal level. This is because the trial will continue randomizing patients until the last cluster of arm 2 is completely filled and its impact will be large

when the total sample size is small. Under H_0, the log-rank test tends to be slightly anti-conservative when sample size is small. In summary, the calculated sample sizes are appropriately powered and the log-rank test controls type I error rate accurately overall.

5.2 Optimal Sample Sizes

We identify optimal allocation proportion p_1 and the corresponding sample size n under the design settings considered for simulations of Table 1. To identify the proportion minimizing total cost, we consider relative cost of $\eta = \eta_2/\eta_1$ of 5/4, 3/2, 7/4, and 2. Table 2 lists optimal sample size n (together with the optimizing p_1 value in the parentheses) under the settings. For comparison, Table 2 also lists n for $p_1 = p_2 = 1/2$ that are listed in Table 1. Since the optimal allocation proportion does not depend on the nominal power, only $1 - \beta = 0.8$ is considered here. From Table 2, we find that p_1 value minimizing the total sample size n decreases in τ. In other words, if the ICC of arm 2 is high, then the optimal allocation randomizes much less patients to arm 1. On the other hand, as the cost of arm 2 relative to that of arm 1 increases, the allocation proportion p_1 minimizing the total cost tends to increase from the p_1 value minimizing n. That is, the p_1 value just minimizing n assigns too many patients to arm 2, so that the p_1 value minimizing the total cost moves toward the balanced allocation with $p_1 = 1/2$ if the relative cost for arm 2 is high.

5.3 A Real Example

Harper et.al. (2015) conducted a CRT to study interventions for reducing unintended pregnancy for young women. They performed the study in 40 clinics across USA, of which 20 clinics were randomly assigned to receive a specialized training on intrauterine devices or progestin implants (arm 2) and the other 20 clinics to receive a standard care (arm 1). If the standard care does not need a group training, we can consider

an IRGT design for this trial. Suppose that only the experimental arm has $\nu_2 = 20$ clusters while the control arm of standard care has independent individuals, with a balanced allocation, i.e. $p_1 = p_2 = 1/2$. The 20 training groups have similar recruiting capacity, so that we assume $\gamma_i = 1/\nu_2 = 1/20$ for $i = 1, ..., \nu_2$. The 1-year pregnancy-free probability for the control arm is known to be 80%, and we want to calculate the required accrual period and the sample size to detect a hazard ratio of $\Delta = \lambda_1/\lambda_2 = 2$ with a power of $1 - \beta = 90\%$. The expected annual accrual rate for the whole study is $r = 200$ and all participating women will be followed until an unexpected pregnancy or for additional $b = 1$ year after completion of accrual. Assuming exponential survival distributions, we have $\lambda_1 = -\log(0.8) = 0.223$ and, under H_1, $\lambda_2 = \lambda_1/\Delta = 0.112$. Assuming $\tau = 0.05$, corresponding to $\theta = 9.5$ and $\text{corr}(T_{2ij}, T_{2ij'}) = 0.1$, we need an expected accrual period of $a = 1.76$ years and a total sample size of $n = 353$ from (8) for the log-rank test with 2-sided $\alpha = 0.05$. From 5,000 simulations we observe an empirical type I error rate of 0.051 and power of 0.892. By rounding up m_i values using $m_i(a) = [p_2 a r \gamma_i]$, the final sample size is slightly increased to $n = 361$ by a balanced randomization. The allocation proportion minimizing the total sample size is $p_1 = 0.58$ with $n = 345$. If the cost for arm 2 is 50% higher than that of arm 1, i.e. $\eta = 1.5$, then the allocation proportion minimizing the total cost is $p_1 = 0.64$ with $n = 353$. Note that this numerical study is slightly different from those for Tables 1 and 2. Accrual rate r is given and the number of clusters for arm 2 is fixed here, while accrual period a and cluster sizes are given and the required number of clusters for arm 2 is calculated in the simulations for Tables 1 and 2.

6 Discussion

The proposed testing method and its sample size formula are very flexible to reflect the diverse design aspects of IRGT trials. The methods developed using large sample

approximations are shown to have good finite sample performance through simulations.

As shown through numerical studies, the sample size depends on many design parameters including the survival and censoring distributions, and ICC and cluster size distribution of arm 2. The sample size calculation of an IRGT trial should be based on the best estimate of these design parameters reflecting the basic design of the new trial.

In designing a new IRGT trial, we propose to use a sample size robust to possible mis-specification of these parameters. As a common example, when all participating clusters are expected to have similar recruiting capacity, we may assume constant cluster sizes at the sample size calculation. When the trial is conducted, however, the attained cluster sizes may vary slightly due to randomization. In this case, the calculated sample size will be slightly under-powered by the increased variance of cluster size distribution. Accounting for possible loss to follow-up snd any possible discrepancy between the design parameters specified for sample size calculation and the true or attained ones, we propose to increase the calculated sample size by 5% to 10%.

The Fortran program developed for numerical studies is available from the author.

Appendix A: Limiting distribution of the log-rank statistic under H_1

We derive the distribution of W under $H_1 : \Lambda_1(t) \neq \Lambda_2(t)$. By definition of the log-rank test is

$$W = \frac{1}{\sqrt{n}} \int_0^\infty \Big\{ \frac{Y_2(t)}{Y(t)} dN_1(t) - \frac{Y_1(t)}{Y(t)} dN_2(t) \Big\}$$

so that we have

$$W = \frac{1}{\sqrt{n}} \int_0^\infty \frac{Y_2(t)}{Y(t)} \{dN_1(t) - Y_1(t) d\Lambda_1(t)\} - \frac{1}{\sqrt{n}} \int_0^\infty \frac{Y_1(t)}{Y(t)} \{dN_2(t) - Y_2(t) d\Lambda_2(t)\}$$

$$+ \frac{1}{\sqrt{n}} \int_0^\infty \frac{Y_1(t) Y_2(t)}{Y(t)} \{d\Lambda_1(t) - d\Lambda_2(t)\} \qquad (A.1)$$

Since $n_k^{-1} Y_k(t)$ uniformly converge to $y_k(t) = S_k(t) G(t)$, from (A.1), we have

$$W = \frac{p_2}{\sqrt{n}} \sum_{i=1}^{n_1} \int_0^\infty \frac{y_2(t)}{y(t)} dM_{1i}(t) - \frac{p_1}{\sqrt{n}} \sum_{i=1}^{\nu_2} \int_0^\infty \frac{y_1(t)}{y(t)} dM_{2i}(t) + \omega\sqrt{n} + o_p(1) \qquad (A.2)$$

where $y(t) = p_1 y_1(t) + p_2 y_2(t)$ and

$$\omega = p_1 p_2 \int_0^\infty \frac{y_1(t) y_2(t)}{y(t)} \{d\Lambda_1(t) - d\Lambda_2(t)\}$$

$$= p_1 p_2 \int_0^\infty \frac{S_1(t) S_2(t) G(t)}{p_1 S_1(t) + p_2 S_2(t)} \{d\Lambda_1(t) - d\Lambda_2(t)\} \qquad (A.3)$$

Since the patients in the two arms are independent, the first two terms of (A.2) are independent. Further, $M_{11}, ..., M_{1,n_1}$ are independent 0-mean martingales, so that the variance of the first term of (A.2) is given as

$$\sigma_1^2 = p_1 p_2^2 \int_0^\infty \frac{y_1(t) y_2(t)^2}{y(t)^2} d\Lambda_1(t)$$

$$= p_1 p_2^2 \int_0^\infty \frac{S_1(t) S_2(t)^2 G(t)}{\{p_1 S_1(t) + p_2 S_2(t)\}^2} d\Lambda_1(t) \qquad (A.4)$$

By Ying and Wei (1994), $n^{-1/2} \sum_{i=1}^{\nu_2} M_{2i}(t)$ converges weakly to a Gaussian process $U_2(t)$. As in Lee, Wei and Ying (1993), we can show that the second term of (A.2) converge a.s. to $\int_0^\infty y_1(t)/y(t)\} dU_2(t)$. Hence, by the standard central limit theorem

for sum of independent random variables, the second term of (A.2) converges to a normal distribution with mean 0 and variance

$$\mathrm{var}\Big\{\int_0^\infty \frac{y_1(t)}{y(t)} dM_{2i}(t)\Big\} = \int_0^\infty \frac{y_1^2(t)}{y^2(t)} Y_{2i} d\Lambda_2(t)$$

$$+2\int_0^\infty \int_0^\infty \frac{y_1(t_1)y_1(t_2)}{y(t_1)y(t_2)} \sum_{1\le j<j'\le m_i} Y_{i(j,j')}(t_1,t_2) dA_2(t_1,t_2)$$

where $Y_{2i}(t) = \sum_{j=1}^{m_i} Y_{2ij}(t)$, $Y_{i(j,j')}(t_1,t_2) = I(X_{2ij} \ge t_1, X_{2ij'} \ge t_2)$ for $j \ne j'$ and

$$dA_2(t_1,t_2) = \{\lambda_2(t_1,t_2) - \lambda_{2(1|2)}(t_1,t_2)\lambda_2(t_2) - \lambda_{2(2|1)}(t_1,t_2)\lambda_2(t_1) + \lambda_2(t_1)\lambda_2(t_2)\}dt_1 dt_2$$

by similar arguments of Li and Jung[5].

So, the variance of the second term of (A.5) is

$$v_2 = \frac{p_1^2}{n}\int_0^\infty \frac{y_1^2(t)}{y^2(t)} Y_2(t) d\Lambda_2(t)$$

$$+\frac{p_1^2}{n}\int_0^\infty \int_0^\infty \frac{y_1(t_1)y_1(t_2)}{y(t_1)y(t_2)} \sum_{i=1}^{\nu_2} \sum_{1\le j<j'\le m_i} Y_{i(j,j')}(t_1,t_2) dA_2(t_1,t_2)$$

We can show that

$$\frac{\sum_{i=1}^{\nu_2}\sum_{1\le j<j'\le m_i} Y_{i(j,j')}(t_1,t_2)}{\sum_{i=1}^{\nu_2} m_i(m_i-1)}$$

uniformly converge to $y_2(t_1,t_2) = S_2(t_1,t_2)G(t_1,t_2)$ as in Lin and Ying (1993), and we have $\nu_2^{-1}\sum_{i=1}^{\nu_2} m_i(m_i-1) \to \bar{\bar{m}} - \bar{m}$ and $\nu_2/n = (\nu_2/n_2)(n_2/n) \to p_2/\bar{m}$.

Hence,

$$v_2 = \sigma_2^2 + (\bar{\bar{m}}/\bar{m} - 1)c_2 \qquad (A.5)$$

where

$$\sigma_2^2 = p_1^2 p_2 \int_0^\infty \frac{y_1^2(t)y_2(t)}{y^2(t)} d\Lambda_2(t) = p_1^2 p_2 \int_0^\infty \frac{S_1^2(t)S_2(t)G(t)}{\{p_1 S_1(t) + p_2 S_2(t)\}^2} d\Lambda_2(t) \qquad (A.6)$$

and

$$c_2 = p_1^2 p_2 \int_0^\infty \int_0^\infty \frac{y_1(t_1)y_1(t_2)}{y(t_1)y(t_2)} y_2(t_1,t_2) dA_2(t_1,t_2)$$

$$= p_1^2 p_2 \int_0^\infty \int_0^\infty \frac{S_1(t_1)S_1(t_2)S_2(t_1,t_2)G(t_1,t_2)}{\{p_1 S_1(t_1) + p_2 S_2(t_1)\}\{p_1 S_1(t_2) + p_2 S_2(t_2)\}} dA_2(t_1,t_2) \qquad (A.7)$$

Note that the ICC in arm 2 is expressed as

$$\rho = \frac{c_2}{\sigma_2^2} = \frac{\int_0^\infty \int_0^\infty \frac{S_1(t_1)S_1(t_2)S_2(t_1,t_2)G(t_1,t_2)}{\{p_1S_1(t_1)+p_2S_2(t_1)\}\{p_1S_1(t_2)+p_2S_2(t_2)\}} dA_2(t_1,t_2)}{\int_0^\infty \frac{S_1^2(t)S_2(t)G(t)}{\{p_1S_1(t)+p_2S_2(t)\}^2} d\Lambda_2(t)} \tag{A.8}$$

From (A.4) and (A.5), under H_1, $\mathrm{var}(W) = \sigma^2$ is given as

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + (\bar{\bar{m}}/\bar{m} - 1)c_2 = \sigma_1^2 + \sigma_2^2\{1 + (\bar{\bar{m}}/\bar{m} - 1)\rho\}$$

Appendix B: Limiting distribution of the log-rank statistic under the nearby alternative hypothesis

We simplify the asymptotic mean ω and variance σ^2 of W under the nearby alternative hypothesis, where $S_1(t) \approx S_2(t)(= S(t))$. Under the nearby alternative hypothesis, ω in (A.3) is simplified to

$$\omega = p_1p_2(\Delta - 1)\int_0^\infty S(t)G(t)d\Lambda(t) = p_1p_2(\Delta - 1)$$

where $\Delta = \lambda_1/\lambda_2$, $\Lambda(t) = -\log S(t)$, and $d = \int_0^\infty G(t)dS(t)$ is the probability that a patients experiences an event during study. We propose to calculate d by $p_1d_1 + p_2d_2$ with $d_k = \int_0^\infty G(t)dS_k(t)$. Hence, we have

$$\omega = p_1p_2(\Delta - 1)d$$

By the Taylor expansion, we have $\Delta - 1 \approx \log\Delta$, so that we have

$$\omega = p_1p_2d\log\Delta$$

From (A.4) and (A.6), we have

$$\sigma_k^2 = p_kp_{3-k}^2\int_0^\infty S(t)G(t)d\Lambda(t) = -p_kp_{3-k}^2\int_0^\infty G(t)dS(t) = p_kp_{3-k}^2d$$

under the nearby alternative hypothesis.

From (A.7),

$$c_2 = p_1^2 p_2 \int_0^\infty \int_0^\infty S_2(t_1, t_2) G(t_1, t_2) dA_2(t_1, t_2)$$

Hence, we have

$$\sigma^2 = p_1 p_2^2 d + p_1^2 p_2 d\{1 + (\bar{\bar{m}}/\bar{m} - 1)\rho\} = p_1 p_2 d\{1 + p_1 \rho(\bar{\bar{m}}/\bar{m} - 1)\}$$

where the ICC in (A.8) is reduced

$$\rho = \frac{c_2}{\sigma_2^2} = d^{-1} \int_0^\infty \int_0^\infty S_2(t_1, t_2) G(t_1, t_2) dA_2(t_1, t_2)$$

REFERENCES

Aalen O (1978) Nonparametric inference for a family of counting processes. Annals of Statistics 6(4):701-726.

Ahn C, Jung SH, Kang SH (2003) An evaluation of weighted chi-square statistics for site-specific binary data. Drug Information Journal 37:91-99.

Candel MJ, Van Breukelen GJ (2009) Varying cluster sizes in trials with clusters in one treatment arm: sample size adjustments when testing treatment effects with linear mixed models. Statistics in Medicine 28(18):2307-2324.

Clayton D, Cuzick J (1985) Multivariate generalizations of the proportional Hazards model. Journal of Royal Statistical Society, Ser A (General) 148(2):82-108.

Donner A, Banting D (1988) Analysis of site-specific data in dental studies. Journal of Dental Research 67:1392-1395.

Fleming TR, Harrington DD (1991) Counting Processes and Survival Analysis. Wiley: New York.

Gehan EA (1965) A generalized Wilcoxon test for comparing arbitrarily single censored samples. Biometrika 52:203–223.

Gill RD (1980) Censoring and Stochastic Integrals, Mathematical Centre Tracts 124, Mathematisch Centrum: Amsterdam.

Harper CC, Rocca CH, Thompson KM, et al (2015) Reductions in pregnancy rates in the USA with long-acting reversible contraception: a cluster randomised trial. Lancet 386(9993):562-568.

Jeong JH, Jung SH (2006) Rank tests for clustered survival data when dependent subunits are randomized, Statistics in Medicine 25:361-373.

Jung SH, Jeong JH (2003) Rank tests for clustered survival data. Lifetime Data Analysis 9:21-33.

Kang SH, Ahn C, Jung SH (2003) Sample size calculation for dichotomous outcomes in cluster randomization trials with varying cluster size. Drug Information Journal 37:109-114.

Kaplan EL, Meier P (1958) Nonparametric estimator from incomplete observations. Journal of the American Statistical Association 53:457-481.

Lee EW, Wei LJ, Ying Z (1993) Linear Regression Analysis for Highly Stratified Failure Time Data. Journal of the American Statistical Association 88(422);557-565.

Li J, Jung SH (2020) Sample size calculation for cluster randomization trials with a time-to-event endpoint. Statistics in Medicine 39(25):3608-3623.

Li J, Jung SH (2022) Sample size calculation for clustered survival data under subunit randomization. Lifetime Data Analysis 28(1):40–67.

Lin DY, Ying Z (1993) A Simple Nonparametric Estimator of the Bivariate Survival function under univariate censoring. Biometrika 80:573-581.

Moerbeek M, Wong WK (2008) Sample size formulae for trials comparing group and individual treatments in a multilevel model. Statistics in Medicine 27:2850-2864.

Nelson W (1969) Hazard plotting for incomplete failure data. Journal of Quality Technology 1(1):27-52.

Pals SL, Murray DM, Alfano CM, Shadish WR, Hannan PJ, Baker WL (2008) Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. American Journal of Public Health 98(8):1418–1424.

Peto R, Peto J (1972) Asymptotically efficient rank invariant test procedures (with discussion. Journal of the Royal Statistical Society, Ser. A 135:185-206.

Prentice RL. Linear rank tests with right censored data. Biometrika 65:167-179.

Roberts C, Roberts SA (2005) Design and analysis of clinical trials with clustering effects due to treatment. Clinical Trials 2(2):152-162.

Schoenfeld DA (1983) Sample-size formula for the proportional-hazards regression model. Biometrics 39(2):499-503.

Turner EL, Li F, Gallis JA, Prague M, Murray DM (2017A) Review of recent methodological developments in group-randomized trials: Part 1—design. American Journal of Public Health 107(6):907-915.

Turner EL, Prague M, Gallis JA, Li F, Murray DM (2017B) Review of recent methodological developments in group-randomized trials: part 2—analysis. American Journal of Public Health 107(7):1078-1086.

Wang X, Turner EL, Li F (2024) Designing individually randomized group treatment trials with repeated outcome measurements using generalized estimating equations. Statistics in Medicine 43:358–378.

Ying Z, Wei LJ (1994) The Kaplan-Meier estimate for dependent failure time observations. Journal of Multivariate Analysis 50:17-29.

Table 1. Sample size n (and simulation results of empirical type error rate $\hat{\alpha}$ and power $1-\hat{\beta}$) in parenthesis) under various design settings. Cluster sizes are constant, $m_i = 10$ or 15, or random with discrete uniform distribution, $U(8,12)$ or $U(13,17)$, and the hazard rates are $\lambda_1 = \lambda_2 = 0.5$ under H_0 and $\lambda_2 = 0.3$ or 0.35 under H_1.

$1-\beta$	m_i	λ_2	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$
0.8	10	0.3	251(0.060, 0.821)	335(0.069, 0.780)	418(0.067, 0.790)
		0.35	518(0.054, 0.797)	704(0.059, 0.806)	886(0.064, 0.801)
	$U(8,12)$	0.3	252(0.067, 0.842)	337(0.067, 0.800)	421(0.070, 0.801)
		0.35	520(0.057, 0.803)	708(0.059, 0.808)	892(0.055, 0.805)
	15	0.3	298(0.069, 0.793)	428(0.077, 0.808)	558(0.073, 0.784)
		0.35	622(0.055, 0.802)	912(0.061, 0.800)	1195(0.061, 0.794)
	$U(13,17)$	0.3	298(0.069, 0.833)	430(0.070, 0.810)	560(0.070, 0.799)
		0.35	624(0.069, 0.799)	915(0.068, 0.792)	1199(0.057, 0.787)
0.85	10	0.3	287(0.067, 0.853)	383(0.068, 0.854)	478(0.070, 0.834)
		0.35	592(0.057, 0.853)	805(0.059, 0.850)	1013(0.056, 0.839)
	$U(8,12)$	0.3	288(0.067, 0.858)	385(0.067, 0.827)	481(0.066, 0.838)
		0.35	594(0.056, 0.843)	810(0.058, 0.842)	1020(0.059, 0.832)
	15	0.3	340(0.065, 0.853)	490(0.068, 0.859)	638(0.065, 0.847)
		0.35	712(0.057, 0.841)	1044(0.057, 0.843)	1367(0.058, 0.846)
	$U(13,17)$	0.3	341(0.071, 0.860)	491(0.075, 0.834)	640(0.067, 0.846)
		0.35	713(0.062, 0.849)	1047(0.053, 0.837)	1372(0.060, 0.841)
0.9	10	0.3	336(0.065, 0.897)	448(0.062, 0.892)	560(0.064, 0.897)
		0.35	693(0.058, 0.908)	942(0.056, 0.894)	1186(0.059, 0.900)
	$U(8,12)$	0.3	337(0.062, 0.911)	451(0.067, 0.901)	563(0.066, 0.896)
		0.35	696(0.055, 0.900)	948(0.059, 0.905)	1194(0.054, 0.892)
	15	0.3	398(0.070, 0.900)	573(0.072, 0.907)	746(0.066, 0.886)
		0.35	833(0.056, 0.896)	1221(0.061, 0.895)	1600(0.049, 0.901)
	$U(13,17)$	0.3	399(0.070, 0.911)	575(0.063, 0.896)	749(0.066, 0.893)
		0.35	835(0.057, 0.908)	1225(0.061, 0.901)	1605(0.055, 0.889)

Table 2. Sample size n (and p_1) for (0.5) with $p_1 = 0.5$, for p minimizing n, and for those minimizing the total cost with a relative cost of $\eta = \eta_2/\eta_1 = 5/4$, 3/2, 7/4, or 2, for $1 - \beta = 0.8$

					Minimizing the total cost with $\eta = \eta_2/\eta_1$			
m_i	λ_2	τ	(0.5)	$\min n$	5/4	3/2	7/5	2
10	0.3	0.1	251	249(0.43)	250(0.48)	252(0.51)	254(0.53)	256(0.54)
		0.2	285	325(0.39)	326(0.44)	328(0.46)	331(0.48)	335(0.50)
		0.3	351	397(0.36)	398(0.41)	401(0.43)	404(0.45)	409(0.47)
	0.35	0.1	518	508(0.43)	510(0.46)	513(0.48)	518(0.50)	524(0.52)
		0.2	629	670(0.37)	672(0.41)	678(0.44)	685(0.46)	689(0.47)
		0.3	786	819(0.34)	822(0.38)	827(0.40)	834(0.42)	843(0.44)
$U(8, 12)$	0.3	0.1	253	251(0.42)	251(0.48)	253(0.50)	256(0.53)	258(0.54)
		0.2	288	328(0.39)	330(0.44)	332(0.46)	335(0.48)	339(0.50)
		0.3	356	402(0.35)	402(0.40)	406(0.43)	410(0.45)	414(0.47)
	0.35	0.1	522	512(0.42)	514(0.46)	517(0.48)	522(0.50)	528(0.52)
		0.2	637	677(0.37)	679(0.41)	683(0.43)	689(0.45)	697(0.47)
		0.3	797	829(0.34)	832(0.38)	837(0.40)	845(0.42)	855(0.44)
15	0.3	0.1	298	292(0.40)	292(0.45)	295(0.48)	298(0.50)	301(0.52)
		0.2	360	406(0.35)	407(0.40)	410(0.43)	414(0.45)	419(0.47)
		0.3	463	513(0.33)	515(0.37)	520(0.40)	525(0.42)	531(0.44)
	0.35	0.1	622	600(0.40)	602(0.43)	606(0.45)	611(0.47)	618(0.49)
		0.2	809	841(0.33)	842(0.37)	850(0.40)	858(0.42)	868(0.44)
		0.3	1053	1063(0.30)	1066(0.34)	1072(0.36)	1081(0.38)	1093(0.40)
$U(13, 17)$	0.3	0.1	299	293(0.40)	294(0.45)	296(0.48)	299(0.50)	303(0.52)
		0.2	362	408(0.35)	409(0.40)	413(0.43)	416(0.45)	421(0.47)
		0.3	466	516(0.33)	518(0.37)	521(0.39)	528(0.42)	534(0.44)
	0.35	0.1	625	603(0.39)	605(0.43)	608(0.45)	614(0.47)	621(0.49)
		0.2	814	845(0.34)	847(0.37)	854(0.40)	862(0.42)	873(0.44)
		0.3	1060	1069(0.30)	1072(0.34)	1079(0.36)	1088(0.38)	1100(0.40)