Sample size planning in the design and analysis of cluster randomized trials using the symbolic two-step method

David Zahrieh *, Jennifer Le-Rademacher

Department of Health Sciences Research, Mayo Clinic, Rochester, MN, 55905, USA

ARTICLE INFO
Keywords:
Care delivery research
Cluster randomized trial
Sample size estimation
Symbolic data analysis

ABSTRACT

Introduction: Evidence that can be used to improve clinical practice patterns and processes is frequently generated through standard, parallel-arms cluster randomized trial (CRT) designs that test interventions implemented at the center-level. Although the primary endpoint of these trials is often a center-level outcome, patient-level factors may vary between centers and, consequently, may influence the center-level outcome. Furthermore, there may be important factors that predict the variation in the center-level outcome and this knowledge can help contextualize the trial results and inform practice patterns.

Methods: Our symbolic two-step method that applies symbolic data analysis to account for patient-level factors when estimating and testing a center-level effect on both the average center-level outcome and its variation was developed for such settings. Herein, we sought to extend the method to prospectively size a CRT so that the application of our method in data analysis is consistent with the design.

Results: Our formulaic approach to sample size planning incorporated predictive factors of the within-center variation and accounted for patient-level characteristics. The sample size approximation performed well in many different pragmatic settings.

Conclusions: Our symbolic two-step method provides an alternate approach in the design and analysis of CRTs evaluating novel improvement processes within care delivery research.

1. Introduction

It is well recognized that a hospital’s (or medical center, hereinafter referred to as a center) processes and policies have a direct impact on patient care and subsequently affect patient outcomes. Many of these processes and policies are modifiable. For example, implementation of a medical team training program by the Veterans Health Administration to encourage teamwork and effective communication was shown to improve surgical mortality [1]. In care delivery clinical trials evaluating improvement processes, cluster randomization is often used to assign centers to either a new process (the intervention) or an existing process. The effectiveness of the intervention is determined by center-level outcomes. Because a center-level outcome is often an accumulation of patient-level outcomes and symbolic data analysis (SDA) embraces the idea of aggregating individual-level data (the micro-data) into group-level distributional summaries (the symbols) and then building models for inference directly at the group-level based on these summaries [2], data from care delivery clinical trials fit naturally in the symbolic data framework. SDA was first introduced by Diday [3] and continues to be a growing field of statistics. Unlike a classical observation, which takes a single value, a symbolic observation can take a set of values. In the above care delivery clinical trial setting, each center is a symbolic-valued observation comprising a set of patient-level outcomes.

Despite randomization, arm imbalance in patient-level factors may occur in cluster randomized trials (CRTs) as centers may have different patient-mixes. The difference in patient-mix may potentially confound the effect of the new process. To achieve a meaningful estimate of the effect of the intervention on a particular center-level outcome, it is crucial in these circumstances to account for the imbalance in patient-level factors between centers. Furthermore, because a center-level outcome is an accumulation of patient-level outcomes, there is an inherent variation in center-level outcome, which may be linked to the patient-mix or clinical practice at each center. Oftentimes, characterization of this variation is not the focus in the statistical analysis; however, it is sometimes of interest to identify factors affecting or predicting the within-center variance, which may in turn lead to targeted changes in clinical practice to reduce such variation.

A two-step method using the symbolic data framework (hereinafter...
referred to as the symbolic two-step) was recently proposed [4] that can adjust for patient-level factors while also maintaining the separation between the center and within-center contributions to the total variance to allow estimation of the effect of the intervention on the mean center-level outcome while simultaneously modeling the within-center outcome variability; consequently, the method conveniently facilitates sample size planning for CRTs. We leveraged this observation and extended the proposed methodology to estimate the sample size when planning a prospective CRT so that both the study design and the data analysis accounts for patient-level factors and incorporates our knowledge of important predictors of the variation of the center-level outcome.

2. Materials and methods

2.1. Symbolic data and the symbolic two-step method

In SDA, we can consider the realization \( \psi_i \) of a symbolic-valued random variable \( X_i \) such that \( \psi_i \) can be written as \( \psi_i = (f_1, f_2, \ldots, f_m) \) where \( m \) is the number of values in \( \psi_i \) and \( f_j(\psi_i) \) is the relative frequency that \( \psi_i \) occurs within \( \psi_j \) for \( j = 1, \ldots, m \). Furthermore, we can assume that \( X_i \) takes a distribution of values with a known density function \( f \) and a parameter vector \( \Theta_i \), where \( \Theta_i \) is defined to ensure a one-to-one correspondence between \( X_i \) and \( \Theta_i \); thus, \( \Theta_i \) comprises the smallest set of parameters that uniquely identify \( f \) in the family of Gaussian distributions. In the symbolic data framework, we say that \( f \) and \( \Theta_i \) are the density function and the parameter vector of the random variable \( X_i \) within \( X \). We refer to \( f \) as the internal density of \( X_i \); additionally, we refer to \( \Theta_i \) as the vector of internal parameters of \( X_i \). Because \( \Theta_i \) is the vector of parameters within \( X_i \) and \( X_i \) is a random variable, \( \Theta_i \) is not fixed but varies along with \( X_i \). In other words, we suppose that \( \Theta_i \) is a random vector corresponding to the symbolic-valued random variable \( X_i \), with known probability density function \( g \). Because there exists a one-to-one correspondence between \( X_i \) and \( \Theta_i \), the probability that \( X_i \) takes a set of values \( \psi_i \) equals the probability that \( \Theta_i \) takes value \( \psi_i \).

The symbolic data framework above formed the basis for a two-step method recently proposed [4] that accounts for patient-level factors in estimating and testing center-level effects on both the average center-level outcome and its variation. The first step of the method estimates the effect of patient factors on the outcome, treating individual patients as the units of observation; importantly, the estimated effect is obtained from a stratified model that stratifies by center to avoid bias toward larger centers and to preserve the separation of the within and between variation among centers. The second step evaluates the center-level intervention effect on the center-mean outcome, after adjusting for patient-level factors from step one, treating centers as the units of observation. In parallel, the predictive effects of center-level characteristics on the within-center variation of the patient-adjusted outcomes are separately evaluated in the second step. Because this method considers the distribution of patient-adjusted outcomes from each center as the center-level response, Le-Rademacher [4] applied SDA; a symbolic observation can take on a distribution of values defined by parameters internal to the symbolic-valued random variable \( \psi \). Using the theoretical framework provided by Le-Rademacher and Billard [6], the distribution of patient outcomes is modeled in step two via classical linear regression models.

In the context of CRTs, suppose we measure some continuous outcome variable \( y_{ij} \) and let \( \psi_{ij} \) be the vector of salient patient-level characteristics of patient \( j \) from center \( i \) \((i = 1, \ldots, n)\). Although the number of patients within each center need not be equal, for simplicity we will assume \( m \) patients within each center. Suppose we randomize the \( n \) centers with equal allocation to receive an intervention or control such that the centers are divided into equal groups with \( c \) centers per group. Ordinary least squares (OLS) is used to estimate the effect of patient-level characteristics on the outcome using the stratified model

\[
y_{ij} = \alpha_i + \beta_i^{T}\theta_j + \epsilon_{ij}
\]

where \( \epsilon_{ij} \) are assumed to be independent with mean zero. The superscript \( s \) is used to emphasize that the error terms are obtained from the stratified model. In the stratified model (1), the vector \( \beta_j \) represents the within-center effects of patient-level characteristics on the patient-level outcome and are constrained to be the same across centers. Because these effects are constrained to be the same across centers, the model permits the separation of the variation caused by patient-mix from the variation caused by center-level characteristics. The intercept \( \alpha_i \) in model (1) represents the average outcome of center \( i \) after adjusting for the patient-level characteristics. Then,

\[
r_i = y_{ij} - \beta_i^{T}\theta_j = \alpha_i + \epsilon_{ij}
\]

represents the outcome of patient \( j \) from center \( i \) adjusted for his or her patient-level characteristics and can be estimated by \( \hat{r}_i = y_{ij} - \hat{\beta}_i^{T}\hat{\theta}_j \), where \( \hat{\beta}_j \) is a vector of unbiased estimators of \( \beta_j \).

Now, consider \( r_i \) to be a symbolic-valued random variable corresponding to the set of adjusted outcomes of the \( m \) patients from center \( i \) with internal mean \( \mu_i \) and internal variance \( \sigma_i^2 \), such that \( \mu_i \) and \( \sigma_i^2 \) are random variables. Here, \( r_i \) represents the center-level outcome of interest from center \( i \). Given a center \( i \), suppose further that the outcome of patient \( j \) \((r_{ij})\) follows a Gaussian distribution with random parameter vector \( \theta_j = (\mu_j, \sigma_j^2) \) corresponding to the internal parameters of \( r_i \). Note that, based on this internal distributional assumption, \( \mu_i = \alpha_i \) in models (1) and (2). We observe the realization \((r_1, \sigma_1^2), \ldots, (r_n, \sigma_n^2)\) of \( \bar{\theta} \) where \( r_i = \frac{1}{m} \sum_{j=1}^{m} r_{ij} \) and \( \sigma_i^2 = \frac{1}{m} \sum_{j=1}^{m} (r_{ij} - \bar{r}_i)^2 \) represent the sample mean and sample variance for the \( i \)th center.

Let \( x_{0i} \) denote an indicator variable, which takes the value one if the \( i \)th center was randomly allocated to receive the intervention and zero otherwise. Furthermore, because we anticipate heterogeneous within-center variances, we allow the variances to vary from center to center as a parametric function of explanatory variables. Suppose that there is a single dichotomous variable \( x_{0i} \) that is driving the within-center variances. Because \( \theta_j \) has a 1-1 relationship with \( r_i \) and \( \theta_j \) captures the total variance of \( r_{ij} \), the effect of the indicator variables \( x_{0i} \) and \( x_{0i} \) on \( r_i \) can be expressed in terms of their effect on \( \theta_j \). With \( \theta_j \) being a vector of single observations, the effect of \( x_{0i} \) and \( x_{0i} \) on \( \theta_j \) can be estimated using OLS. Specifically, the effect of \( x_{0i} \) on \( \mu_j \) and the effect of \( x_{0i} \) on \( \sigma_j^2 \) can be modeled separately using OLS by the following two linear models

\[
\mu_i = \gamma_{\mu_0} + \gamma_{\mu_x} x_{0i} + \epsilon_i
\]

\[
\sigma_i^2 = \gamma_{\sigma_0} + \gamma_{\sigma_x} x_{0i} + \epsilon_i
\]

where \( \epsilon_i \sim N(0, \sigma_0^2) \) and \( \epsilon_j \sim N(0, \sigma_j^2) \). Stated differently, the sample means \((\bar{r}_1, \ldots, \bar{r}_n)\) of the patient-adjusted outcome are regressed on the variable \( x_{0i} \) and the sample log within-center variations \((\log(\sigma_1^2) - \log(\sigma_i^2), \ldots, \log(\sigma_n^2) - \log(\sigma_i^2))\) of the patient adjusted outcome are separately regressed on the variable \( x_{0i} \). The coefficient \( \gamma_{\mu_0} \) represents the intervention effect on the center-mean outcome and is the primary effect of interest in a CRT design. The coefficient \( \gamma_{\mu_x} \) represents the effect of the dichotomous predictor variable on the log within-center outcome variability. The standard errors (SEs) for the estimated coefficients \( \gamma_{\mu_0} \) and \( \gamma_{\mu_x} \) are

\[
\text{SE}(\gamma_{\mu_0}) = \frac{\sigma_0}{\sqrt{\sum_{i=1}^{n} (\bar{r}_i - \gamma_{\mu_0} - \gamma_{\mu_x} x_{0i} - \epsilon_i)^2}},
\]

\[
\text{SE}(\gamma_{\mu_x}) = \frac{\sigma_0}{\sqrt{\sum_{i=1}^{n} (\bar{r}_i - \gamma_{\mu_0} - \gamma_{\mu_x} x_{0i} - \epsilon_i)^2}},
\]
\( \hat{\sigma} \) equals the square root of the sum of squared errors divided by \( n - 2 \) and confidence intervals (CIs) are immediately obtained in closed form from classical linear regression theory [7].

2.2. Sample size determination

Suppose we randomize \( n \) centers with equal allocation to receive the intervention or control such that the centers are divided into equal groups with \( c \) centers per group (i.e., \( n = 2c \)). Further, we will continue to assume \( m \) patients within each center. Consider a normally distributed quantitative endpoint, where the objective is to compare the mean of the sample mean within the intervention arm, say, with \( \mu_{r_0} + \sigma_{r_0} \), and confidence intervals (CIs) are immediately obtained in closed form from classical linear regression theory [7].

The expected value of a patient-adjusted outcome \( r_j \) is

\[
E(r_j) = E[E(r_j | r_i)] = E(\mu_i) = \mu_{r_0} + \mu_{r_0} x_i
\]

such that \( \mu_{r_0} \) is the expected value for a randomly selected patient in the absence of the intervention, while \( \mu_{r_0} + \mu_{r_0} x_i \) is the expected value in the presence of the intervention. The expected value of \( r_A = r_B \) is then

\[
E(r_A - r_B) = (\mu_{r_0} + \mu_{r_0} x_i) - (\mu_{r_0} + \mu_{r_0} x_i) = \mu_{r_0}
\]

The variance of \( r_j \) will be a sum of two components, namely, the internal and external variation, where the internal variation is the mean of a log normal distribution. That is, for the patient-adjusted outcome \( r_j \), the variance is

\[
V(r_j) = E[V(r_j | r_i)] + E[V(\mu_i | r_i)] = E(\sigma^2) + E(\mu^2)
\]

\[
= E(\exp(\xi_j)) + \sigma^2
\]

\[
= \exp(\gamma_{m_0} x_i + \sigma_{\gamma_0}^2) + \sigma^2
\]

Based on this expression, the variance of the observed mean \( r_i \) for the \( i \)th center with \( m \) observations is given by

\[
V(r_i) = \frac{\exp(\gamma_{m_0} x_i + \sigma_{\gamma_0}^2)}{m} + \sigma^2
\]

Now, suppose that the dichotomous variable \( x_m \) is a stratification factor at the time of random assignment such that proportion \( \pi \) of the \( c \) centers within each arm will correspond to the level of the factor \( x_m = 1 \) and proportion \( 1 - \pi \) of the \( c \) centers within each arm will correspond to the level of the factor \( x_m = 0 \). We can express the internal variation as a weighted average, where \( \pi \) and \( 1 - \pi \) are the weights, such that the variance of the sample mean within the intervention arm, say, with \( c \) centers is

\[
V(r_i) = \frac{(1 - \pi) \times \exp(\gamma_{m_0} + \sigma_{\gamma_0}^2) + \pi \times \exp(\gamma_{m_0} + \gamma_{m_0} + \sigma_{\gamma_0}^2)}{cm} \]

The variance of the difference in means between these two independent samples will then be

\[
Var(r_A - r_B) = Var(r_1) + Var(r_2) = \frac{2}{c} \left( \frac{1 - \pi) \times \exp(\gamma_{m_0} + \sigma_{\gamma_0}^2) + \pi \times \exp(\gamma_{m_0} + \gamma_{m_0} + \sigma_{\gamma_0}^2)}{m} \right)
\]

The approximate power to detect a difference of \( \Delta = r_{10} \) for conducting a two-tailed test of size \( \alpha \) is then

\[
\Phi \left( \frac{\Delta}{\sqrt{\text{Var}(r_A - r_B)}} - Z_{1 - \frac{\alpha}{2}} \right)
\]

where \( \Phi \) is the cumulative standard Normal distribution and \( Z_{1 - \frac{\alpha}{2}} \) is the \( \frac{1}{2} \) quantile of the standard Normal distribution function. With \( (1 - \beta) \times 100\% \) power, this means that the approximate number of centers required per arm can be computed as

\[
\left( Z_{1 - \frac{\alpha}{2}} + 1 - \beta \right)^2 \times 2 \times \left( \frac{1 - \pi) \times \exp(\gamma_{m_0} + \sigma_{\gamma_0}^2) + \pi \times \exp(\gamma_{m_0} + \gamma_{m_0} + \sigma_{\gamma_0}^2)}{m} \right)
\]

2.3. Simulation

To assess the performance of the approximate sample size formula (5) under different scenarios, we conducted a simulation study. The goal for carrying out these simulation studies was to understand the settings when the formula would be deemed reliable for estimating the approximate sample size for a standard, parallel-arms CRT and, therefore, make sample size determination by way of information-based simulation methods unnecessary.

2.4. Data generating model

Data were generated from a multilevel (or hierarchical) linear model [8]. We considered grouped data - patients within centers - where some information was available on patients and some information was at the center-level. With this nested structure, we had observations \( j = 1, \ldots, m \) clustered in centers \( i = 1, \ldots, n \). Within each center \( i \), we assumed that we had three continuous individual-level predictors \( z_{i1}, z_{i2}, z_{i3} \), and one center-level predictor labeled as \( x_{im} \) such that the center-level predictor was an indicator variable (intervention or control) and was the primary variable of interest.

The multilevel linear model can be expressed as a linking of local regressions in each center with a random intercept term \( \mu_i \) indexed by center. The local individual-level predictors were generated from a \( N(0,1) \).

Within each center \( i \),

\[
y_j \sim N(\mu_i + \beta_1 z_{i1} + \beta_2 z_{i2} + \beta_3 z_{i3}, \sigma_{\gamma_0}^2) \quad \text{for} \quad j = 1, \ldots, m
\]

We supposed that there was a single dichotomous variable \( x_m \) that was driving the within-center variances, such that when \( x_m = 1 \) the \( i \)th center was associated with decreased within-center variation and zero otherwise. Then, the variances on the log scale were allowed to vary as a parametric function of the indicator variable \( x_m \), such that

\[
\xi_i = \log(\sigma^2) \sim N(\gamma_{m_0} + \gamma_{m_0} x_m, \sigma^2) \quad \text{for} \quad i = 1, \ldots, n
\]

The \( n \) within-center variances \( \sigma^2 \) were generated first from a log normal distribution with location parameter \( \gamma_{m_0} + \gamma_{m_0} x_m \) (i.e. the mean of the log of the distribution expressed as a function of the center-level predictor variable \( x_{im} \) and shape parameter \( \sigma^2 \) (i.e. the standard deviation of the log of the distribution).

The center-level predictor was included as a predictor in the next level of the model:

\[
\mu_i \sim N(\gamma_0 + x_m, \sigma^2)
\]

The errors in this model (with mean zero and standard deviation \( \sigma_{\gamma_0} \) represented variation among centers that was not explained by the single center-level and three local individual-level predictors. Furthermore, we assumed \( \xi_i \) and \( \mu_i \) were independent.
2.5. Model parameterizations

For each simulated data set, we applied the symbolic two-step method to estimation and using the t-statistic for the regression coefficient, with the standard error for the estimated coefficient, tested at the 0.05 significance level the null hypothesis \( \gamma_{pi} = 0 \) vs the two-sided alternative hypothesis \( \gamma_{pi} \neq 0 \) when the true value \( \gamma_{pi} = 0.65 \) is model (3). Given the power of the symbolic two-step method obtained via simulations \( c \in \{8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48\} \), we graphically compared the number of centers \( c \) with the number of centers \( c \) obtained from the formula. We considered the setting where centers were randomized to the intervention or control arm in a ratio of 1:1 and \( \pi \in (0.75, 0.5, 0.25) \) such that the proportion \( \pi \) of the centers within each arm that corresponded to \( x_{pi} = 1 \) in model (4) was either 0.75, 0.5, or 0.25. Further, for each \( \pi \), we considered four scenarios corresponding to four model parameterizations of model (4), and varied \( \sigma_i \in (0.5, 1.0, 1.5) \) to explore different effect sizes associated with the dichotomous predictor \( x_{pi} \). The four model parameterizations applied to address our goals in this simulation study were:

1. \( \gamma_0 = -1.5, \gamma_1 = 0.0 \)
2. \( \gamma_0 = -1.5, \gamma_1 = 0.68 \)
3. \( \gamma_0 = -1.5, \gamma_1 = -1.37 \)
4. \( \gamma_0 = -1.5, \gamma_1 = -2.985 \).

The first model parameterization corresponded to the setting of homogenous mean log variance in the presence of varying noise as defined by \( \sigma_i \). While the second, third, and fourth scenarios, respectively, corresponded to increasing effect sizes of the dichotomous predictor \( x_{pi} \) such that the fourth scenario represented a substantial difference in mean log within-center variance when \( x_{pi} = 1 \) vs when \( x_{pi} = 0 \) and in the presence of different degrees of variation as defined by \( \sigma_i \). For instance, when \( \sigma_i = 0.5, \gamma_0 = -1.5, \gamma_1 = 2.985 \), the \( E[\sigma_i^2] \) for the \( r \)-th center was 0.25 and 5 for \( x_{pi} = 0 \) and \( x_{pi} = 1 \), respectively, or a 20 fold increase when \( x_{pi} = 1 \). An illustrative sample of the scenarios studied is shown in Table 1.

Testing that the coefficient \( \gamma_{pi} = 0 \) in model (8) against the two-sided alternative that \( \gamma_{pi} \neq 0 \) is equivalent to applying a two-sample t-test to test the null hypothesis that the population means within each arm are equal vs the two-sided alternative that the population means are not equal. Because there are \( c \) centers in each of \( c \) arms, the dataset comprises \( n = 2c \) numbers consisting of the sample means for each center. When the number of centers, and hence the degrees of freedom for the t-test \( (2c - 2 = n - 2 \) degrees of freedom), are small, the percentage points of the t distribution are appreciably larger than those of the standard normal distribution; to correct for this in our simulation study, we added one center to \( c \) in the formula because the calculation makes use of the normal approximation [9]. We assumed a targeted effect (intervention vs. control) of \( \gamma_{pi} = 0.65, \sigma_i \in (0.75, 1.0) \), and settings where the number of observations \( m \) within each center was 25, 50, or 100.

### 3. Results

#### 3.1. Simulation

For the sample of four scenarios illustrated in Table 1, corresponding results are shown in Figs. 1 and 2 with \( \sigma_i \in (0.75, 1.0), m \in (25, 50, 100) \), and \( \pi = 0.75 \). The scenarios were selected for presentation because they encapsulated the performance of the myriad scenarios studied. The figures show the power of the test associated with increasing numbers of centers \( c \) based on simulation as well as the number of centers \( c \) determined from the formula given the power of the test obtained from simulation.

The formula performed well across a wide variety of scenarios one may encounter in practice (not all shown); however, in the presence of substantial differences in mean log within-center variation (parameterization 4), particularly when the proportion \( \pi \) of centers with increased mean log within-center variation was high (\( \pi = 0.75 \)), large error variance (\( \sigma_i = 1.5 \)), the robustness of the formula with 25 observations within each center was lacking for smaller numbers of centers \( c \). In these particular settings, the formula consistently indicated that more centers \( c \) were needed to achieve the desired power. Practical limits of the formula as it pertains to the model parameterization of the internal variation are discussed in the supplementary materials.

#### 3.2. Application

In the context of a recently performed data analysis, we consider an application of sample size planning of a hypothetical standard, parallel arms CRT with a continuous outcome.

#### 3.2.1. Data analysis

Our symbolic two-step method was previously applied to a retrospective evaluation of bone marrow transplant center practices on post-transplant survival [4]. The retrospective data analysis was based on a subset of data obtained from a study conducted by Majhail et al. [10,11]. The subset comprised 3320 patients from 67 transplant centers across the United States (approximately 50 patients per center, on average). The continuous outcome of interest \( y_{ij} \) was the probability of survival at 1-year based on pseudo-observations which represented the estimated value of the survival function at 1-year for each patient \( j \) in center \( i \) [12] and calculated using the R package pseudo. Pseudo-observations are based on the idea of the jackknife estimator and involve transforming a set of survival-time observations into a set of pseudo-observations that can be modeled directly using regression analysis. Pseudo-observations are not necessarily restricted to the range 0–1, as occurred in our retrospective data analysis; however, they provided an alternative, albeit informative approach to illuminate the relationship between transplant center characteristics and survival by allowing the analysis of censored survival data by linear regression.

After adjusting for the patient-level characteristics in \( z_{ij} \), namely, Karnofsky performance score, disease status, prior autologous transplant, time from diagnosis to transplant, donor-recipient HLA match, unrelated donor age, and Sorror comorbidity score index, the adjusted

### Table 1

| Parameterization | Mean Within-Center Variance | Mean Within-Center Variance |
|------------------|-----------------------------|----------------------------|
| No. Parameters   | \( E[\sigma_i^2|x_{pi} - 1] \) | \( E[\sigma_i^2|x_{pi} = 0] \) | Weight Average |
| 2                | 0.50                        | 0.25                       | 0.44          |
| 4                | 5.00                        | 0.25                       | 3.81          |
| 2                | 1.36                        | 0.69                       | 1.19          |
| 4                | 13.60                       | 0.69                       | 10.37         |
outcomes of patients from the same center were combined into a distribution of outcome representing center. The distribution of outcome within each center was assumed to be a symbolic random variable ($r_i$ for the $i$th center) corresponding to the set of adjusted outcomes within that center ($f_{ij}$ for $j \in \{1,...,m_i\}$) such that each distribution of outcome was characterized by an internal mean ($\mu_i$) and variance ($\sigma^2_i$). The center-level characteristics considered in the data analysis were allogeneic transplant volume ($\leq 40$ vs. $> 40$), whether the transplant center participated in at least one clinical trial in the past 12 months (yes vs. no), and the number of ventilation units (2 units vs. 1 unit).

Approximately 40% of the centers had an allogeneic transplant volume $\leq 40$ and 92% had participated in a clinical trial within the past 12 months. And 22% of the centers had a single ventilation unit in the transplant center. The results from applying our symbolic two-step method, which are shown in Table 2, indicated that centers with lower allogeneic transplant volume were associated with decreased

---

**Fig. 1.** Each figure shows the power of the test associated with increasing numbers of centers $c$ based on Monte Carlo simulation as well as the number of centers $c$ determined from the approximate sample size formula given the power of the test obtained from Monte Carlo simulations. The targeted effect (intervention vs. control) of $r_{0.65}$ was assumed.

Parameterizations 2 and 4 (Table 1). $\pi = 0.75$, $\alpha_i = 0.55 \in \{8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48\}$.
center-mean survival at 1 year. In addition, the data suggested that increasing the number of ventilation units may result in a clinically important improvement in center-mean 1-year survival. Participation in clinical trials in the past 12 months was associated with lower log within-center variability in 1-year survival; furthermore, these data suggested that centers with lower allogeneic transplant volume may be associated with increased log within-center variability in 1-year survival. Increasing the number of ventilation units seemingly was not a predictor of log within-center variability in 1-year survival in the presence of the other two center-level factors ($P = 0.345$).

### 3.2.2. Study design

Transplant facilities that may not be properly ventilated, designed, or controlled can lead to the spread of airborne pathogens throughout the facility. Consequently, transplant patients with a compromised immune system will be susceptible to infection, leading to premature

**Fig. 2.** Each figure shows the power of the test associated with increasing numbers of centers $c$ based on Monte Carlo simulation as well as the number of centers $c$ determined from the approximate sample size formula given the power of the test obtained from Monte Carlo simulations. The targeted effect (intervention vs. control) of $r_{H_1} = 0.65$ was assumed.

Parameterizations 2 and 4 (Table 1). $\pi = 0.75$, $\sigma_z = 1.5c \in \{8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48\}$. 
Applying the symbolic two-step method. Effect of center-level characteristics on 1-year survival (pseudo-observations) after adjusting for patient-level characteristics.

| Outcome | Center Characteristic | Parameter | Estimate (SE) | 95% CI |
|---------|-----------------------|-----------|---------------|--------|
| Mean \( \mu \) | | Intercept \( \gamma_0 \) | 0.817 (0.077) | (0.667, 0.968) |
| | | Allogeneic transplant volume \( (\leq 40 \text{ vs } >40) \) \( \gamma_2 \) | -0.076 (0.030) | (-0.135, -0.018) |
| | Recent participation in clinical trials (yes vs no) \( \gamma_{30} \) | -0.027 (0.053) | (-0.130, 0.076) |
| | Number of ventilation units (2 vs 1) \( \gamma_{31} \) | 0.036 (0.031) | (0.008, 0.065) |
| Log variance \( \beta \) | | Intercept \( \gamma_0 \) | -1.433 (0.106) | (-1.640, -1.225) |
| | | Allogeneic transplant volume \( (\leq 40 \text{ vs } >40) \) \( \gamma_2 \) | 0.070 (0.041) | (0.001, 0.150) |
| | Recent participation in clinical trials (yes vs no) \( \gamma_{30} \) | -0.170 (0.072) | (-0.312, -0.028) |

Note: Pseudo-observations at 1-year represented the continuous outcome and were obtained from the R package pseudo. Pseudo-observations are not necessarily restricted to the range 0–1, as occurred in the current analysis; however, they provided an alternative, albeit informative approach to illuminate the relationship between transplant center characteristics and both the center-mean and log within-center variability in 1-year survival by allowing the analysis of censored survival data by linear regression. SE = (model-based) standard error. The data set included 3320 patients from 67 transplant centers. After obtaining the patient-adjusted outcomes from Step 1 in the symbolic two-step method, the effect of the center-level characteristics on the center-mean survival at 1 year and log within center variability in 1-year survival were modeled separately using ordinary least squares estimation in Step 2.

4. Discussion

The symbolic two-step method that applies SDA accounts for patient-level factors when estimating and testing a center-level effect on both the average center-level outcome and its variability. Because statistical inference is conducted at the center-level, rather than at the individual-level, our methodologic approach parallels cluster-randomized designs. Further, because our method preserves the separation of the within- and between-center variation, and permits the within-center variation to depend on explanatory variables in a manner that can be estimated from the data, it conveniently facilitates sample size planning of CRTs. Our approach benefits PLCs, the usual formula for determining the approximate number of clusters to randomize in a CRT can be applied. In the event that there are two categorical predictive factors, extending the number of stratification factors and applying the approximate sample size formula is straightforward as we demonstrated with our design application. Our approach to sample size determination performed well and was found to be robust for most pragmatic settings studied, and the formulaic approach obviates the need to estimate the sample size via extensive simulation studies. Because obtaining good estimates of the components needed for characterizing the within-center variation may be challenging, particularly in settings with small numbers of centers within centers, a range of plausible values of those components should be considered in order to assess the impact on sample size estimation.

There are limitations to our research. To determine the number of centers to randomize with our method in a prospective CRT design, we assumed a fixed center size; in other words, we did not explore the implications on the study’s power in the event of variable sample size per center. Further, equal weight is given to each center when modeling the average center-level outcome in step 2 of the symbolic two-step method, and weighted alternatives, while conceivable, were not considered. Additionally, the symbolic two-step method is limited to the setting of continuous response outcomes where maintaining the separation between the center and within-center contributions to the total estimates (95% CI) of the coefficients \( \gamma_{30}, \gamma_{31}, \gamma_{32} \), and \( \gamma_{33} \) of \(-1.43 \text{ (1.64, -1.23), 0.07 (0.01, 0.15), and -0.17 (-0.31, -0.03), respectively.} \)

Our estimate of \( \sigma^2 \) was 0.02. Based on the estimated proportion of 67 transplant centers within each level of the crossed two-level stratification factors \( \{\gamma_1, \gamma_2, \gamma_3, \gamma_4\} = (0.02, 0.58, 0.06, 0.34) \), where \( \gamma_1, \gamma_2, \gamma_3, \) and \( \gamma_4 \) correspond to \( (x_{001}, x_{002}) = (0, 0), (x_{010}, x_{002}) = (0, 1), (x_{001}, x_{012}) = (1, 0), \) and \( (x_{001}, x_{002}) = (1, 1) \), with the constraint that \( \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 1 \). Additionally, we will assume the possibility that \( (\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (0.15, 0.30, 0.25, 0.30) \). Further, we consider the following plausible values of \( \gamma_{30}, \gamma_{31}, \) and \( \gamma_{32} \) based on inspection of the corresponding 95% CIs: 1.23, 0.15, and -0.31. Table 3 displays the number of transplant centers \( c \) (per arm) needed to detect a hypothesized difference of 0.05 (intervention vs. usual care ventilation strategies) in center-mean, patient-adjusted 1-year survival probabilities with 80% power and assuming a type I error rate of 5% under our assumptions, as well as the power obtained empirically via 10,000 replicates. Given the value \( c \), the empirical power was consistently close to 80%, i.e. the power applied in the formula.
Table 3
Design application of the symbolic two-step method in sample size planning. Number of transplant centers \( c \) (per arm) needed to detect a hypothesized difference of 0.05 in center-mean, patient-adjusted 1-year survival probabilities based on pseudo-observations with 80% power and a two-sided \( \alpha \) of 5%.

| \( \tau_m \) | \( \tau_i \) | \( \tau_{c0} \) | \( \sigma_i^2 \) | \( m \) | \( \sigma_m^2 \) | \( \sigma_E^2 \) | \( \sigma_{E0}^2 \) | \( \sigma_{E1}^2 \) | \( \sigma_{E2}^2 \) | \( \sigma_{E3}^2 \) | Empirical Power |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| 1.43     | 0.07     | -0.17    | 0.02     | 50       | 0.01     | (0.02, 0.58, 0.06, 0.34) | 0.05      | 0.24      | 0.29      | 0.26      | 0.22      | 90        |
|          |         |          |          | 100      |          |                        |           |           |           |           |           | 77        |
|          |         |          |          | 50       |          | (0.15, 0.30, 0.25, 0.30) |           |           |           |           |           | 92        |
|          |         |          |          | 100      |          |                        |           |           |           |           |           | 78        |
| -1.23    | 0.15     | -0.31    | 0.02     | 50       | 0.01     | (0.02, 0.58, 0.06, 0.34) | 0.05      | 0.30      | 0.22      | 0.34      | 0.25      | 93        |
|          |         |          |          | 100      |          |                        |           |           |           |           |           | 79        |
|          |         |          |          | 50       |          | (0.15, 0.30, 0.25, 0.30) |           |           |           |           |           | 97        |

Note: \( \tau \) = sample size based on the formula. Given \( \tau \), the simulated power obtained empirically via 10,000 replicates is shown at the far right. When \( x_{m0} = 0 \), the allogeneic transplant volume at center \( i \) is > 40, while \( x_{m1} = 1 \) the allogeneic transplant volume at the \( m \)th center is ≤ 40. When \( x_{m1} = 0 \), transplant center \( i \) did not recently participate in a clinical trial, while \( x_{m1} = 1 \) indicates that the \( m \)th center recently participated in a clinical trial.

4.1. Conclusion

Knowing the predictive factors of the variation in the center-level outcome of interest is valuable in both the design and data analysis of standard, parallel arms CRTs and our proposed formulaic approach to sample size planning incorporated this knowledge as well as simultaneously accounted for patient-level characteristics. Our symbolic two-step method provides an alternate approach to both the design and analysis of standard, parallel arms CRTs evaluating novel improvement processes within care delivery research.

Funding

This work was supported in part by CTSA Grant Number UL1 TR002377 from the National Center for Advancing Translational Science (NCATS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. This work was also partially supported by the National Institutes of Health Grants: U10CA180882 (Alliance for Clinical Trials in Oncology Statistics and Data Management Grant), and P30CA15083 (Mayo Clinic Comprehensive Cancer Center grant).

Authors’ contributions

Both authors contributed equally to the manuscript.

Acknowledgements

Not Applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2020.100609.

References

[1] J. Neely, P.D. Mills, Y. Young-Xu, R.T. Carney, P. West, D.H. Berger, L.M. Mazzia, D. E. Pauli, J.P. Bagian, Association between implementation of a medical team training program and surgical mortality, J. Am. Med. Assoc. 304 (15) (2010 Oct 20) 1693–1706.
[2] B. Beranger, H. Lin, S.A. Sisson, New Models for Symbolic Data Analysis. arXiv: 180903659v1, 2018.
[3] E. Diday, The symbolic approach in clustering and related methods of Data Analysis, in: H. Rock (Ed.), Classification and Related Methods of Data Analysis, North-Holland, Aachen, Germany, 1987.
[4] J. Le-Rademacher, A-center-level approach to estimating the effect of center characteristics on center outcomes, in: Advances in the Mathematical Sciences: Research from the 2015 Association for Women in Mathematics Symposium, 2016, pp. 301–321.
[5] L. Billard, E. Diday, Symbolic Data Analysis Conceptual Statistics and Data Mining, Wiley, 2006.
[6] J. Le-Rademacher, L. Billard, Likelihood functions and some maximum likelihood estimators for symbolic data, J. Stat. Plann. Inference 141 (2011) 1593–1602.
[7] D.G. Kleinbaum, L.L. Kupper, A. Nizam, K.E. Muller, in: Duxbury (Ed.), Applied Regression Analysis and Other Multivariable Methods, fourth ed., Books/Cole Cengage Learning, Belmont, CA, 2008.
[8] A. Gelman, J. Hill, Data Analysis Using Regression and Multilevel/hierarchical Models, Cambridge University Press, 2007.
[9] R.J. Hayes, L.H. Moulton, Cluster Randomized Trials, second ed. ed, Chapman & Hall- CRC, 2017.
[10] N.S. Majhail, T. Payton, L.W. Mau, Provider and center characteristics of us transplant centers and their association with survival after allogeneic hematopoietic cell transplantation (htc) in adults: results from a national survey conducted by the center for international blood and marrow transplant research (cibmtr), Blood 122 (2013) 1687.
[11] N.S. Majhail, L.W. Mau, P. Chithakdithai, E.M. Denzen, S. Joffe, S.J. Lee, C. F. LeMaistre, F. Loberiza, S.K. Parsons, R. Repaczki-Jones, P. Robinett, J. Douglas Rizzo, E. Murphy, B. Logan, J. Le-Rademacher, Transplant center characteristics and survival after allogeneic hematopoietic cell transplantation in adults, Bone Marrow Transplant. 55 (5) (2020) 906–917, https://doi.org/10.1038/s41409-019-0748-1 [Epub ahead of print].
[12] P.K. Andersen, J.P. Klein, S. Rosthj, Generalized linear models for correlated pseudo-observations with applications to multi-state models, Biometrika 90 (2003) 15–27.