Traditional trials with cluster-randomized designs

MedianaDesigner package

1. Introduction

This document provides a description of the statistical methodology used in the module that supports traditional trials with cluster-randomized designs (ClustRand function).

For more information on the MedianaDesigner package, visit the following web pages at

http://www.mediana.us/medianadesigner

http://medianasoft.github.io/MedianaDesigner

2. Cluster-randomized trials

2.1. Trial design

Cluster randomized designs have been developed for settings where trial designs with individual randomization are inappropriate or impractical. In this setting, groups of subjects or patients, known as clusters, are randomly assigned to experimental treatments/interventions or controls. Clusters are commonly defined in terms of clinical sites in clinical trials or different types of communities such as households in community intervention studies. It is important to note that subjects or patients within each cluster are considered exchangeable. For more information on cluster-randomized trials, see Hayes and Moulton (2009) and Ahn, Heo and Zhang (2015).

An evaluation of treatment effects in cluster-randomized trials is most commonly performed using the following two methods:

- Generalized estimating equations (GEEs).
- Generalized linear mixed-effects models (GLMEMs).

In the GLMEM framework random cluster effects and consequently between-cluster variation are explicitly modelled. Likelihood-based methods are then utilized to estimate the parameters of interest such as the overall treatment effect in a trial. GEEs rely on an alternative approach where between-cluster variation is not explicitly modelled and the focus is shifted to within-cluster correlations.

Both frameworks account for the fact that the outcomes for patients within the same cluster are correlated. The strength of this correlation is quantified using the intracluster correlation

coefficient (ICC). The ICC is an estimate of the relative variability within and between clusters, i.e., it measures the degree to which patient outcomes in a given cluster are similar to each other compared with patient outcomes from the other clusters. A lower ICC indicates that the patient outcomes within each cluster are weakly correlated, which ultimately improves the probability of success in the trial.

A traditional trial design without interim analyses is considered in this module. The current version of the module supports parallel-group trials for evaluating the efficacy profile of several dosing regimens of an experimental treatment or several types of interventions compared to a common control, e.g., placebo. The primary efficacy endpoint in this trial could be either continuous (normally distributed) or binary and the efficacy analysis is based on a superiority assessment.

General considerations for cluster-randomized designs are presented in Section 2.2. The design of cluster-randomized trials based on the GEE and GLMEM methods is discussed in Section 2.3. A review of sensitivity assessments supported by this module is provided in Section 2.4. The process of designing cluster-randomized trials is illustrated in Section 3.

2.2. General considerations

To generate the patient outcomes in cluster-randomized trials with a continuous endpoint, the ClustRand function relies on a multivariate normal distribution that incorporates random cluster effects. For trials with a binary endpoint, a beta-binomial method is used to generate sets of clusters with correlated binary outcomes. This was method was introduced in Lee and Dubin (1994) and has been employed in multiple studies.

If multiple experimental treatments or interventions are compared to a control, a multiplicity adjustment is applied to the treatment-control comparisons to preserve the overall Type I error rate in the trial. The ClustRand function supports the most commonly used multiple testing procedures to guarantee error rate control:

- Bonferroni procedure.
- Holm procedure.
- Fixed-sequence procedure.
- Hochberg procedure.

For more information on multiple testing procedures supported by the MedianaDesigner package, see the documentation for the MultAdj function.

A standard definition of power, known as disjunctive power, is used in trials with several treatment-control comparisons. In this case power is defined as the probability that at least one treatment provides an improvement over the control.

2.3. Design of cluster-randomized trials using GEEs and GLMEMs

While the GEE method is commonly used in the design of cluster-randomized trials, it is important to keep in mind that the standard implementation of GEEs may lead to biased inferences. Specifically, the standard implementation relies on the sandwich variance estimator and this estimator is biased when the number of clusters is small (e.g., less than 40), which leads to Type I error rate inflation. In general, the degree of Type I error rate inflation in a cluster-randomized trial depends only on the number of clusters, e.g., with a small cluster count, the Type I error rate is not affected by the cluster size or ICC. To mitigate this issue, the ClustRand function supports alternative approaches to defining the variance estimator for GEEs, including bias-corrected estimator proposed by Kauermann and Carroll (2001) and Mancl and DeRouen (2001). Note that the estimator developed by Fay and Graubard (2001) is generally similar to the Kauermann-Carroll correction (Scott et al., 2017) and is not supported by this function.

In addition, the GLMEM method is also quite popular for modeling correlated patient outcomes within each cluster. This method relies on mixed-effects models with a random term for the cluster. To perform a reliable evaluation of treatment effects with the GLMEM method, it is important to employ small-sample corrections based on the Kenward-Rogers denominator degrees of freedom. This approach reduces Type I error inflation in trials with small cluster counts and low ICC values (Johnson et al., 2015).

2.4. Sensitivity assessments

An important feature of the ClustRand function is that, in addition to standard power calculations, it supports a broad range of sensitivity assessments to address potential issues related to Type I error rate inflation (see Section 2.3) and explicitly quantify the impact of key factors on the probability of success:

- ICC misspecification.
- Cluster size variability.
- Small number of clusters.

The sensitivity assessments help the trial's sponsor design more robust trials and prevent issues that often arise due to inadequate planning such as reduced power due to cluster size variability or inflated Type I error rate.

After the initial power calculation, the ClustRand function enables the user to perform the following sensitivity assessments:

- A range of ICC values could be considered.
- Unequal numbers of clusters could be considered in the two trial arms.
- Unequal cluster sizes could be considered in each trial arm.
- Clusters of random size could be considered.

The first three assessments are quite self-explanatory and the last assessment assumes that the cluster size can vary across the simulation runs, which is very common in real-life cluster-randomized trials. To perform a sensitivity analysis with clusters of random size, the user specifies the expected size of each cluster and the cluster sizes are generated from a generalized Bernoulli distribution. Note that the expected size does not have to be constant across the clusters and thus the user can assess the impact of unequal cluster sizes on the trial's operating characteristics.

3. Case study

A cluster-randomized clinical trial with a binary endpoint will be used to illustrate power calculations supported by the ClustRand module. A two-arm design with a single dose of an experimental treatment versus placebo will be considered and the GEE method will be considered.

The following assumptions were made in this case study. A common ICC of 0.3 was assumed in the two trial arms. The response rates were set to 30% and 60% in the placebo and treatment arms, respectively (in this setting a higher response rate indicates a beneficial efficacy profile for this endpoint).

An initial power calculation relied on a basic set up with an equal number of clusters in the two arms and equally-sized clusters. Specifically, 16 clusters of 10 patients each were considered in each arm, which results in the sample size of 160 patients per arm. Power calculations for the resulting cluster-randomized design were performed using 10,000 simulation runs with the nominal one-sided Type I error rate set to 0.025. The key operating characteristics of the cluster-randomized design are summarized in Table 1. This table presents the Type I error rate and power computed using the three variance estimators supported by the ClustRand function. The sandwich estimator is the standard variance estimator used in GEEs and the associated power level for the selected design was 81%. However, this estimator is known to underestimate the true variance in small samples and often leads to Type I error rate inflation. Indeed, it is shown in Table 1 that the one-sided error rate for the sandwich estimator is well over the target value of 0.025. This table also presents the error rate and power calculation results for the bias-corrected variance estimators developed by Kauermann and Carroll (2001) and Mancl and DeRouen (2001). In this case, the one-sided Type I error rate of the Mancl-DeRouen estimator was quite close to the target value of 0.025. Note that the associated power level was a few percentage points below 80%. If the target level for power is set to 80%, the trial's sponsor will need to increase the sample size to achieve the desired probability of success in the trial.

In addition to the operating characteristics presented in Table 1, the ClustRand function computes descriptive statistics such as the sample response rates and their quantiles, e.g., the

lower and upper quartiles, across the simulation runs. These quantities help the user assess how closely the simulated values follow the assumed values of the treatment effect parameters.

The basic design considered above relies on a number of assumptions that may be violated in the actual trial, e.g.,

- The cluster size is fixed across the clusters.
- All clusters have the same size.
- An equal number of clusters is assumed in the two trial arms.

As emphasized in Section 2.4, it is advisable to assess the impact of these assumptions on the conclusions by performing sensitivity assessments with respect to key factors that influence the probability of success in cluster-randomized trials. It is helpful to examine the impact of each factor or a combination of the factors on the power values to ultimately arrive at a robust trial design.

As an illustration, a sensitivity analysis with random cluster sizes was considered. Cluster sizes are sampled from a generalized Bernoulli distribution and, to fix the distribution's parameters, the expected cluster sizes need to be specified. For example, as a starting point, the clusters in the trial could be assumed to have the same average size, i.e., 16 clusters in each arm with the expected size of 10. Table 2 summarizes the results of this sensitivity analysis. It presents the probability of success (power) for the GEE method with the Kauermann-Carroll and Mancl-DeRouen estimators. As shown in this table, even if heterogenous cluster sizes are anticipated, the trial's power remains very high. In fact, the power dropped by only a single percentage point compared to Table 1 where a fixed clustering scheme was assumed. As an extension of this sensitivity assessment, the user could consider cases where the expected size varies across the clusters. Settings like this one mimic real-life trials where clusters are quite unlikely to have the same size.

References

Ahn, C., Heo, M., Zhang, S. (2015). Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. Chapman and Hall/CRC.

Fay, M.P., Graubard, B.I. (2001). Small sample adjustments for Wald type tests using sandwich estimators. Biometrics. 57, 1198-1206.

Hayes, R.J., Moulton, L.H. (2009). Cluster Randomised Trials: A Practical Approach. Chapman and Hall/CRC.

Johnson, J.L., Kreidler, S.M., Catellier, D.J., Murray, D.M., Muller, K.E., Glueck, D.H. (2015). Recommendations for choosing an analysis method that controls Type I error for unbalanced cluster sample designs with Gaussian outcomes. Statistics in Medicine. 34, 3531-3545.

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

Lee, E.W., Dubin, N. (1994). Estimation and sample size considerations for clustered binary responses. Statistics in Medicine. 13, 1241-1252.

Mancl, L. A., DeRouen, T.A. (2001). A covariance estimator for GEE with improved small sample properties. Biometrics. 57, 126-134.

Scott, J.M., DeCamp, A., Juraska, M., Fay, M.P., Gilbert, P.B. (2017). Finite-sample corrected generalized estimating equation of population average treatment effects in stepped wedge cluster randomized trials. Statistical Methods in Medical Research 26. 583-597.

Table 1. General characteristics of the cluster-randomized design based on the GEE method (initial power calculations)

Variance estimator	Parameter	Value
Sandwich	One-sided Type I error rate	0.033
	Power	0.81
Kauermann-Carroll	One-sided Type I error rate	0.028
	Power	0.792
Mancl-DeRouen	One-sided Type I error rate	0.024
	Power	0.775

Table 2. Power of the cluster-randomized design based on the GEE method (sensitivity assessment with random cluster sizes)

Variance estimator	Parameter	Value
Kauermann-Carroll	Power	0.783
Mancl-DeRouen	Power	0.762

Clusters are assumed to have the same average size (16 clusters in each arm with the average size of 10).