Traditional designs with multiple objectives

MedianaDesigner package

1. Introduction

This document describes the statistical methodology used in the module that supports simulation-based power calculations in clinical trials with multiple objectives (MultAdj 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. Clinical trials with multiple objectives

It is very common to employ multiple objectives, e.g., several dose-control comparisons, in late-stage clinical trials to better describe the efficacy profiles of experimental treatments. The evaluation of multiple objectives gives rise to multiplicity and results in an inflated probability of incorrect decisions (overall Type I error rate). A variety of statistical methods known as multiplicity adjustments or, more formally, multiple testing procedures have been proposed in the literature to address Type I error rate inflation. For a review of commonly used multiplicity adjustments with clinical trial applications, see Dmitrienko and D'Agostino (2013) and Alosh, Bretz and Huque (2014).

2.1. Clinical trials with a single source of multiplicity

Multiplicity problems are often classified using the concept of a source of multiplicity. A source of multiplicity is defined as a factor that contributes to Type I error rate inflation. Examples of Phase III clinical trials with a single source of multiplicity include:

- Trials with several clinical endpoints.
- Trials with several doses or dosing regimens compared to a common control.
- Trials with several pre-defined patient populations, e.g., the overall population of patients with a condition of interest and a subset of patients with a certain baseline characteristic.

In all of these cases, a single analytical strategy such as the evaluation of several clinical endpoints induces multiplicity that needs to be addressed using appropriate statistical methods.

To protect the overall Type I error rate in multiplicity problems with a single source of multiplicity, the following multiple testing procedures can be applied:

- Bonferroni procedure.
- Holm procedure.
- Fixed-sequence procedure.
- Chain or graphical procedures.
- Hochberg procedure.
- Hommel procedure.

These multiple testing procedures are supported by the MedianaDesigner package. The MultAdj function can be used to perform simulation-based power calculations in clinical trials with a single source of multiplicity using any of these multiplicity adjustments.

2.2. Clinical trials with several sources of multiplicity

In addition, it is increasingly more common to employ complex sets of objectives, e.g., multiple clinical endpoints evaluated at several dose levels or in several patient populations, in Phase III clinical trials. Since a combination of several analytical strategies is used in these settings, the resulting multiplicity problems are referred to as problems with several sources of multiplicity.

To protect the overall Type I error rate in these complex multiplicity problems, a class of multiple testing procedures known as gatekeeping procedures has been introduced. Gatekeeping procedures are designed for settings with two or more sources of multiplicity and take into account the information on clinically relevant relationships among the individual objectives. The MultAdj function enables simulation-based power calculations in clinical trials with several sources of multiplicity using efficient gatekeeping procedures derived from the Holm, Hochberg and Hommel procedures.

2.3. Clinical trials with multiple related objectives

Finally, multiple objectives such as multiple endpoints can be evaluated in a clinical trial with the goal of quantifying an overall treatment effect on the objectives. Most commonly, goals of this kind are pursued in clinical trials with biologically related endpoints. In this case, the individual endpoints are treated as components of a single underlying clinical variable and the treatment effect across the endpoints is assessed using global testing procedures, e.g., the O'Brien global procedure based on ordinary least squares. Global testing procedures are most commonly used in Phase II clinical trials. The MultAdj function supports simulation-based power calculations in clinical trials with multiple related endpoints using the O'Brien procedure.

3. Case study

The process of running simulation-based sample size calculations in clinical trials with multiple objectives will be illustrated using a Phase III trial with a single source of multiplicity. This case study is based on a confirmatory trial for the treatment of Type 2 diabetes (Rosenstock et al., 2009). A single primary endpoint is assumed in this case study and multiplicity is induced by the comparison of three doses of an experimental treatment to placebo (the doses will be labelled L, M and H). To control the overall Type I error rate at the nominal level (one-sided 0.025 level) in this trial, a multiplicity adjustment needs to be prospectively defined.

The trial information used in sample size calculations is summarized below:

- A balanced design will be considered in the trial and 30% of the patients are expected to be lost to follow-up prior to completing the treatment period. The common sample size for each trial arm (number of enrolled patients) will range between 120 and 150 patients.
- A number of treatment effect scenarios (sets of effect sizes for the individual dose-placebo comparisons) will be considered to support a robust sample size calculation:
    - Scenario 1 (realistic scenario): The effect sizes at Doses L, M and H are 0.25, 0.3, 0.35, respectively.
    - Scenario 2 (optimistic scenario): The effect sizes at Doses L, M and H are 0.3, 0.35, 0.4, respectively.
    - Scenario 3 (pessimistic scenario): The common effect size across the doses is 0.3.

In general, since a stronger beneficial effect is expected at the highest dose (Dose H), weights will be pre-defined for the three dose-placebo comparisons to quantify their relative importance:

- Dose L versus placebo: Weight is 0.25.
- Dose M versus placebo: Weight is 0.35.
- Dose H versus placebo: Weight is 0.4.

A multiplicity adjustment based on the Hommel procedure will be utilized in this case study. This multiple testing procedure guarantees overall error rate control in this setting due to the fact that the test statistics associated with the three comparisons follow a trivariate normal distribution with and the three pairwise correlation coefficients are positive (Dmitrienko and D'Agostino, 2013). This multiplicity adjustment will be set up to account for the weighting scheme defined above.

The overall probability of success will be defined in a standard way, i.e., as the probability of establishing a statistically significant effect for at least one dose-placebo comparison while controlling the overall Type I error rate at a one-sided 0.025 level (this definition of power is known as disjunctive power). The sample size will be chosen to guarantee 80% power across the scenarios described above.

Power calculations were performed under the total of 12 scenarios to assess the probability of success under the selected treatment effect scenarios over a range of sample sizes. The simulation results are presented in Table 1. As expected, this table shows that disjunctive power increases with the sample size under each scenario. If Scenario 1 was the only treatment effect to be used in this sample size calculation, the common number of enrolled patients per arm would be set to 130 since this sample size results in 80% power. To support more robust calculations, Scenario 3 was included and it is clear that a larger sample size will be required to achieve the goal of powering this trial at 80%. The final sample size (in terms of the enrolled patients) will be set to 140 patients per arm.

References

Alosh, M., Bretz, F., Huque, M. (2014). Advanced multiplicity adjustment methods in clinical trials. Statistics in Medicine. 33, 693-713.

Dmitrienko, A., D'Agostino, R. Sr. (2013). Traditional multiplicity adjustment methods in clinical trials. Statistics in Medicine. 32, 5172-5218.

Rosenstock et al. (2009). Effect of saxagliptin monotherapy in treatment-naive patients with type 2 diabetes. Current Medical Research and Opinion. 25, 2401-2411.

Table 1. Disjunctive power as a function of the sample size under the three treatment effect scenarios

Common sample per trial arm	Scenario 1	Scenario 2	Scenario 3
120 patients	0.765	0.878	0.736
130 patients	0.805	0.904	0.776
140 patients	0.834	0.926	0.809
150 patients	0.865	0.941	0.824

Sample size is defined as the number of enrolled patients.