Randomization tests in R

Emmanuel Quinaux, Director Biostatistical Services and Nicolas Dubois, Senior Biostatistician,
IDDI, Louvain-La-Neuve (Belgium)

Abstract

Minimization is a dynamic, covariate-adaptive, treatment-allocation procedure that can be used to achieve balance across many factors simultaneously. The flexibility of minimization is often unfortunately overlooked due to concerns regarding the complexity of implementing dynamic treatment allocation. In addition, when minimization is used as the method of randomization, a randomization test for the analysis is required. A simple R script can easily perform such a test.

Minimization

Important prognostic factors are identified before the trial starts, and assignment of a new patient to a treatment group is determined to minimize the differences between the groups in terms of these factors, by balancing allocation over factor margins. Unlike stratified randomization, minimization aims to minimize the total imbalance for all factors simultaneously instead of considering mutually exclusive strata. Pocock and Simon defined a more general procedure, but the most commonly used approaches for implementation are the range and variance methods, introduced with an illustrative example below.

Consider the example of a trial with two treatments, A and B, with age (≤65/>65) and sex (F/M) as important prognostic factors. Center is added as a factor for which balanced treatment allocation is desirable. A 60-year-old woman in center XYZ is ready to be randomized into the trial that has the following status (Table 1):

If the variance method is used, the preferred treatment can be obtained through a simple summation of the numbers under each treatment already allocated to A and B in each factor. Denote these total allocations by T_A and T_B, respectively (note that these sums are not numbers of patients, given the overlap between table rows). In the above example (Table 1), $T_A = 94$ and $T_B = 96$, and since $T_A < T_B$, treatment A is the preferred treatment.

The next step consists in allocating the preferred treatment with probability p (with $0.5 < p \leq 1$) and the other treatment with probability $1-p$. In case there is no preferred arm, allocate A or B at random with probability 0.5.

- If $p = 1$, the allocation is deterministic and always goes to the preferred treatment;
- If $0.5 < p < 1$, a stochastic rather than a deterministic implementation of minimization is obtained.

Table 1. Illustrative example of the implementation of the minimization.

Number of patients allocated to	A	B
Age: < 65	23	22
Gender: Female	55	54
Center: XYZ	16	20
	94 (T_A)	96 (T_B)
	↓	↓
	Allocate A to minimize "total" imbalance	

Pros & cons of minimization

+ Tends to ensure balance in prognostic factors
+ Number of prognostic factors can be large
+ No deterministic allocations
- Reliable information on prognostic factors must be available when randomizing
- Predictability > 0.5 when prognostic factors and treatment allocations of enrolled subjects known
- No pre-specified randomization lists
- A randomization test will be required when minimization is used

Case study

In a study in patients with recurrent epithelial ovarian cancer, patients were randomized between 2 treatments (experimental vs control) in a 1:1 ratio. The IRT used a minimization algorithm using pre-dose CA125 values (250 > or >-250 kU/l), whether the patients are in first or second relapse and center. The objective of the study was to show non-inferiority in terms of progression-free survival (PFS) as assessed by the blinded central review.

Randomization test

The randomization test is based on a large number of simulated trials in which patients are reallocated randomly to the control arm or the experimental arm using the same stochastic minimization, so as to produce an empirical distribution of the test statistic under the null hypothesis. This empirical distribution is used to assign a statistical significance, via an empirically estimated re-randomization p-value, to the observed log-rank test statistic calculated using the original randomization allocation.

A total of 10000 simulations were conducted, in which the seed passed to the stochastic minimization algorithm is randomly generated. Patients were re-randomized in the same order they were originally randomized. As the stochastic minimization proceeds within each site, the next patient to be randomized is compared on the basis of the stratification factors to the patients previously randomized in that site, and the preferred treatment arm is the one that minimizes the imbalance in strata. The stochastic minimization algorithm is designed so that a patient has an 80% chance of being allocated to the preferred treatment arm, therefore subsequent re-randomizations will introduce new allocations.

This process generated a total of 10000 treatment allocations, i.e., 10000 variations of treatments. For each of these treatment allocation variants, the pre-specified statistical analysis (e.g., a log-rank statistics) was carried out, and the subsequent test statistic estimate was recorded. The empirical p-value is the frequency, calculated as the total number of times out of 10000, that a simulated test statistic is strictly larger (i.e., more extreme) than the asymptotically estimated test statistic on the observed data using the original randomization allocation (see Chart 1).

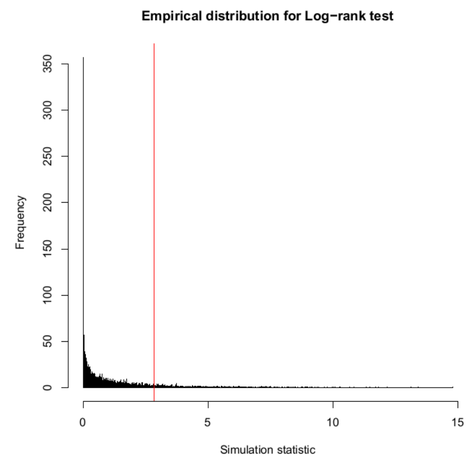

Chart 1. Empirical distribution for the log-rank test

R code

The randomization test can be performed with an R script of 100 lines of code. This script:

- computes the observed log-rank test statistic,
- imports the data related to the subjects to be randomized,
- runs a function "miniAlgo" that assigns the treatment allocation based on the minimization algorithm,
- performs a simulation in order to provide an empirical distribution of the log-rank test statistics under the null hypothesis,
- computes the randomization test's p-value.

Results

The primary efficacy endpoint was PFS defined as the duration of time until progression, based on CT scans evaluated according to RECIST by blinded central review, or death from any cause. PFS was analyzed by the log-rank test stratified according to the stratification factors (CA 125 values (>250 or ≤250 kU/l) and relapse status (first or second)). Comparing the two progression-free survival curves with the stratified log-rank test resulted in a p-value of 0.0919.

The aim of the re-randomization is to find the empirical log-rank test p-value. After re-randomization, the empirical distribution is presented by Chart 1.
The red line gives the observed value in statistical analysis of the log-rank test: 2.840393.

The empirical p-value, being the number of test values superior to the observed value in statistical analysis divided by the number of simulations, was equal to 0.0919. It was exactly the same as the asymptotic log-rank p-value.

Conclusions

- Minimization algorithm and randomization tests are easy to implement in R
- In the case study we got an empirical p-value exactly equal to the asymptotic p-value

Contact:
Emmanuel Quinaux
IDDI
Email: emmanuel.Quinaux@iddi.com
Website: www.iddi.com
Phone: +32 10 61 44 44

Reference: Coart E, Bamps P, Quinaux E, Sturbois G, Saad ED, Burzykowski T, Buyse M. Minimization in randomized clinical trials. **Stat Med.** 42: 5285-5311, 2023