A Practical Introduction to Simulating Complex Trial Designs

Dr Pavel Mozgunov, Dr Dominique-Laurent Couturier
MRC Biostatistics Unit

Practical 1: Introduction to simulation in R and simulations of multi-arm trials

You are asked to design a multi-arm trial in a confirmatory setting of a Phase 3 trial called TREADON. TREADON is a four-arm two-stage trial being conducted in plantar heel pain. The aim is to show whether exercise and/or foot orthoses (shoe insoles) provide more pain relief for adults with plantar heel pain than a self-management advice booklet alone. Up to 680 participants will be recruited to detect a small-to-moderate standardised between-group effect-size of 0.3 in at least one intervention arm against the control arm (SMA booklet) with 90% probability, while controlling the probability of the overall type I error at 5% (one-sided). The design uses a group-sequential stopping rule allowing treatments to be dropped for futility or the trial to be stopped if efficacy has been shown after half of the patients have been recruited.

To start off the exercise, we will assume that 170 patients are recruited per treatment arm and there are have only two arms (control and one experimental), single-stage and the responses are normally distributed

$$Y_{0i} \sim N(\mu_0, \sigma^2), \quad Y_{1i} \sim N(\mu_1, \sigma^2),$$

where $\sigma = 1$ is assumed to be known.

Task 1

Write code to simulate *one* simulated trial of n patients on each treatment arm (either under null or alternative hypothesis) and then find whether the corresponding null hypothesis is rejected at level α. *Note: use* `rnorm` *to simulate the observations then compute the corresponding Z-statistic, rejecting the null hypothesis when $Z > \Phi^{-1}(1-\alpha)$*

Task 2

Using your code for one simulated trial, write a `function` that for the given number of patients, μ_1, μ_0 and σ generates the patients' responses and outputs the estimated treatment effect.
Using this function, generate $m = 10000$ realisations of your trial storing the estimate of the treatment effect in a vector. From this find the vector of corresponding Z-statistics $z = (z_1, \ldots, z_{10000})$. You

may now estimate the probability of rejecting the null hypothesis using Monte-Carlo integration:

$$P(Z > \Phi^{-1}(1-\alpha)) \approx \frac{1}{m}\sum_{i=1}^{m} I[z_i > \Phi^{-1}(1-\alpha)].$$

You can perform it under the null and under the alternative configurations. What are the estimated type I error and power under the considered sample size? How can you illustrate these using the distribution of the Z-statistics?
Note: use `for` *loop and insert your code from Task 1 (with necessary changes) inside of this loop*

Task 3

Above, we used $m = 10000$ realisations of the trial to estimate the probability of rejecting the null hypothesis. Is this sufficient?
Suppose we have m realisations of the trial and let p be the probability of rejecting the null hypothesis. The number of hypotheses rejected follows binomial distribution $\text{Bin}(m, p)$. Thus, if $\hat{p}$ is the estimated probability then

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{m}.$$

What would be a reasonable variance to target? How many realisations m would you need? Does the answer depend on whether we are estimating type I error or power?

Task 4

From the results above, you may have noticed that the estimates follow a normal distribution, given by

$$\hat{\mu}_1 - \hat{\mu}_0 = \hat{\theta} \sim N\left(\theta, \frac{2\sigma^2}{n}\right).$$

As an alternative simulations approach, now simulate estimates (!) of each trial directly (rather than the patients response). Compare simulating summary statistics against the patients' responses in terms of the estimates and computational time.
Note: you might want to create a function that runs simulations with the patient response (i.e. using the code above) to streamline the code and comparison.

Task 5

Now, we add the second experimental treatment in our simulation study, $Y_{2i} \sim \mathcal{N}(\mu_2, \sigma^2)$. Start with an approach to generate patients' outcomes directly. Find the joint distribution of the Z-statistics for treatment 1 and treatment 2 (e.g. under the null). Are they correlated? Why? What is the distribution of the Dunnett test statistic? What would be the type I error if the critical value for two-arm trial is used? If it is not an appropriate critical values, what would be the correct one to use? How one can find it?
Note: for the last part you might need to consider a multivariate normal distribution with an appropriate correltion structure. The following function can be used for this:

```
require(mvtnorm)
qmvnorm((1-xxx), tail="lower.tail", mean=c(xxx,xxx), corr=xxx)$quantile
```

Task 6

Now, approach Task 5 using the simulations of the treatment effect estimates directly. What would be the advantages and disadvantages of this approach compared to the one taken in Task 5? Are there any different to the ones discussed above?

Task 7

Consider now alternative hypotheses. What is the power under the LFC (consider the uninteresting treatment effect of 0.1)? What is the power under the global alternative hypothesis? Which should be considered to power the study? For the selected configuration, find the sample size needed to achieve 90% power (if not achieved already).

Task 8

In the actual trial, the outcome is also measured at baseline, and the baseline and final outcomes are expected to be highly correlated. Instead of the simple normal model, write a code for the simulation study that adjusts the analysis for the baseline measurement.

(a) In the first instance, assume that the correlation of the baseline measurement and the outcome is 0.00 and confirm the findings of the previous simulation study under the null hypothesis and under the sample size of 170 patients per arm.

(b) Confirm the findings of the previous simulation study under the LFC and the sample size of 170. Find, what the power would be for this sample size if the true correlation is 0.5. How long did this simulation study took? What is the advantages and disadvantages of this approach compared to the one tried before?

(c) Is there a quicker way to run such a simulation study? Can one adjust σ above knowing the correlation and number of observations taken?