Response-Adaptive Randomisation

Practical 4: Introduction to RAR design and analysis

Instructor: Lukas Pin

In this practical we consider the re-design of the CALISTO trial (Decousus et al., 2010), a randomised study comparing a new drug (Arixtra) for treating patients with acute symptomatic thrombophlebitis of the lower limbs against Placebo. The primary efficacy outcome was a composite of events at day 47: death from any cause, symptomatic pulmonary embolism, symptomatic deep-vein thrombosis, symptomatic extension to the saphenofemoral junction or symptomatic recurrence of superficial-vein thrombosis. A treatment success in this case is the absence of all of these events at day 47. Observed success probabilities were $p_0 = 0.941$ in placebo arm and $p_1 = 0.991$ in the Arixtra arm.
A well designed response adaptive design could be beneficial for trials like these because of their potential to increase the number of successes on average. Ideally, we would like to achieve this while preserving (or even increasing) power of treatment comparison. There are additional reasons to consider the use of a response-adaptive design in this kind of settings. (Tehranisa and Meurer, 2014). Using the estimated success rates as if they were the underlying (expected) true rates for the future trial, we will compare equal randomisation (ER) to the following response-adaptive designs:

(1) Randomised play the winner rule (as described in the lecture 1).

(2) Play the winner rule: (or the non-randomised procedure version of the randomised play the winner rule). The first patient[1] is allocated to an arm randomly. If the outcome of that patient is a success the next patient will be allocated to the same arm. If it is a failure the next patient will be allocated to the other arm.

(3) Bayesian response-adaptive RAR design with no tuning as in the Slides.

Exercise 1: Fill in the theoretical expected limiting allocation proportion for the 4 procedures in Table 1 below taking the reported values as the true values. The play the winnner rule (PW) and randomised play the winner rule (RPTW) have the same limiting allocation proportion (see Lecture 1 notes for RPTW), which can be found on the lecture slides. For BRAR (burn-in= 2) compute the limiting allocation by simulation using the R Shiny app.
http://shiny.mrc-bsu.cam.ac.uk/RAR/

Design	ER	PW	RPW	BRAR
Limiting Proportion	0.5			1

Table 1: Limiting Proportions for ER and RAR designs.

[1] Or patients if the *burn-in* is bigger than 1.

You are asked for advice on which of the above designs should be considered for trial design in a very similar setting.[2] Assume the (fixed) trial size $n = 366$ was chosen to achieve 80% power in a design using equal randomisation. To implement and compare the aforementioned options we need to decide how to start the allocation before any outcome data is collected. One option to do this is to set have an initial phase of the trial where allocation probabilities are fixed and equal and only after that phase allocation probabilities start adapting based on data accumulation. This initial phase is called a *burn-in* period. Please allocate 30 patients per arm using ER before starting with the RAR procedure. This leads to a burn-in period of 60 patients in total out of $n_{max} = 366$ patients.

Exercise 2: Simulate the limiting mean allocation to the treatment arm for trial replicates (under the observed rates in the CALISTO trial) using the R shiny app. Fill them into column 3 of Table 2. Compare the simulated value to the theoretical values calculated above. How do they compare to the theoretical values above? Take some notes to report back to all at the end of the practical.

Exercise 3: Use the R shiny app to compute the gaps in Table 2 for power and expected mean responses (EMR) reported in Table 2.

Exercise 4: Engage in a discussion with others to understand which design you would recommend within the context of the trial provided above. Why would you choose your preferred design? Why would you not use other options? Take some notes to report back to all at the end of the practical.

Table 2: Run 10,000 Simulations for each Scenario (1000 if it takes to long) and report the Power of the Wald-test for the simple mean difference, which percentage of the patients gets assigned to the superior arm and what the expected mean response is. For the last two metrics we are also interested in their variance across simulations.

Procedure	$Power_Z$	% Arm1 (Var)	EMR (Var)
ER	80%	50% (0)	0.966 (0.0001)
PW			
RPW			
BRAR			

Exercise 5: Does your recommendation change if you consider the provided type-I error rates for $p_0 = p_1 = 0.941$ (no treatment difference) from Table 3?

Table 3: Wald test Type-I error rates for $p_0 = p_1 = 0.941$ for $n = 366$ and a burn-in period of 30 patients per arm (10,000 Simulations for each Scenario)

Procedure	ER	PW	RPW	BRAR
Type-I Error Z-test	5%	2.8%	5.1%	4.5%

[2]For example, Trials comparing therapies for acute myocardial infarction, these trials yield around 93%-95% non-failure rate and commonly use a 30-day mortality binary primary endpoint.

Bonus Question 1: How would your conclusions change for a trial much larger than 366?
Bonus Question 2: Try smaller and larger burn-in periods. Minimum burn-in is 2 patients per arm and maximum is $n/2$ patients per arm. How does this impact your conclusions?
Bonus Question 3: Re-run the whole analysis for very small success probabilities $p_0 = 0.009$ and $p_1 = 0.059$. Do you expect similar results? Why? If you have further questions, you can ask the RAR experts present in the room. Just ask!

Space for your own notes!

References

H. Decousus, P. Prandoni, P. Mismetti, R. M. Bauersachs, Z. Boda, B. Brenner, S. Laporte, L. Matyas, S. Middeldorp, G. Sokurenko, and others. Fondaparinux for the treatment of superficial-vein thrombosis in the legs. *New England Journal of Medicine*, 363(13):1222–1232, 2010.

J. S. Tehranisa and W. J. Meurer. Can response-adaptive randomization increase participation in acute stroke trials? *Stroke*, 45(7):2131–2133, 2014.