Exploring re-randomisation tests in an equivalence trial

Holly Prescott, Marsha Kabeleva and Niccolo' Bassani

Quanticate
The Clinical Data Experts

Introduction

- Randomisation forms the foundation for validity of all statistical tests conducted in the trial. This is particularly true in trials where stratification factors need to be accounted for.
- As an alternative to 'static' methods, dynamic randomisation methods 'create' the list as the patients are enrolled, using an algorithm to balance weight of stratification.
- Regulators are concerned type I error control might not always be achieved through dynamic randomisation, hence they advise on using re-randomisation methods [1]
- Whilst the above test is particularly well suited for superiority trials, can it be used within an equivalence trial?

Methodology

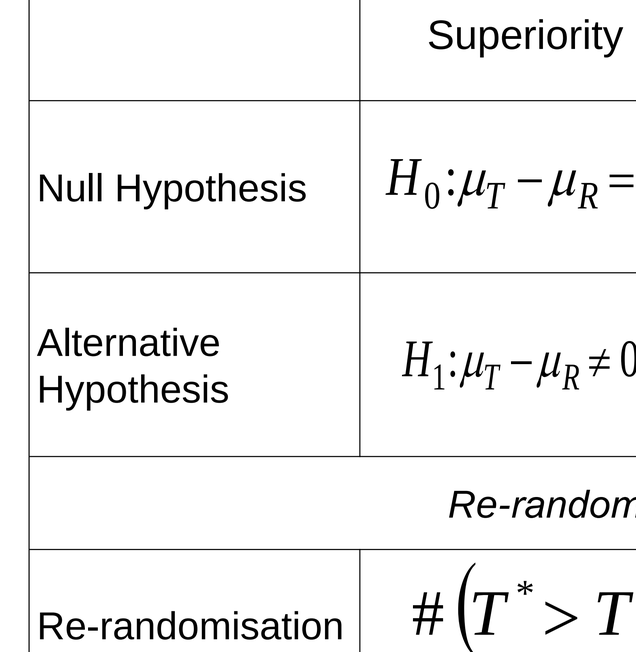
- In a superiority trial, under the null hypothesis of no difference, treatment labels are randomly re-shuffled across patients and the test statistic estimated for each re-shuffled dataset.
- The number of re-shuffled statistics which is more extreme then the 'original' one is the re-randomisation p-value. The higher (e.g. >0.05), the less robust is the original finding

Table 1. *Re-randomisation test framework*

	Superiority	Equivalence
Null Hypothesis	$H_0: \mu_T - \mu_R = 0$	$H_{01}: \lvert\mu_T - \mu_R\rvert \geq \delta^U \cup H_{02}: \lvert\mu_T - \mu_R\rvert \leq \delta^L$
Alternative Hypothesis	$H_1: \mu_T - \mu_R \neq 0$	$H_{11}: \lvert\mu_T - \mu_R\rvert < \delta^U \cap H_{12}: \lvert\mu_T - \mu_R\rvert > \delta^L$
Re-randomise…		
Re-randomisation test	$\frac{\#\left(T^* > T\right)}{N}$	$\#D^* > \left(\delta^U + D\right)/N$ $\#D^* > \left(\delta^L + D\right)/N$

NOTE: T(D) is the value of test statistic (difference between treatments)) on the original dataset, T*(D*) is the value over the re-shuffled datasets. δ^U and δ^L are, respectively, the upper and lower margin of the equivalence margin.

Figure 1. Superiority trial (N = 200)



The histogram represents the distribution of the re-randomised test statistics.

Re-randomisation p-value is based on how many re-shuffled test statistics lie outside of the vertical red lines (original results)

Since we get a 'significant' number of more extreme results by re-randomising, the original p-value is not conclusive of a difference between treatments

How about equivalence?

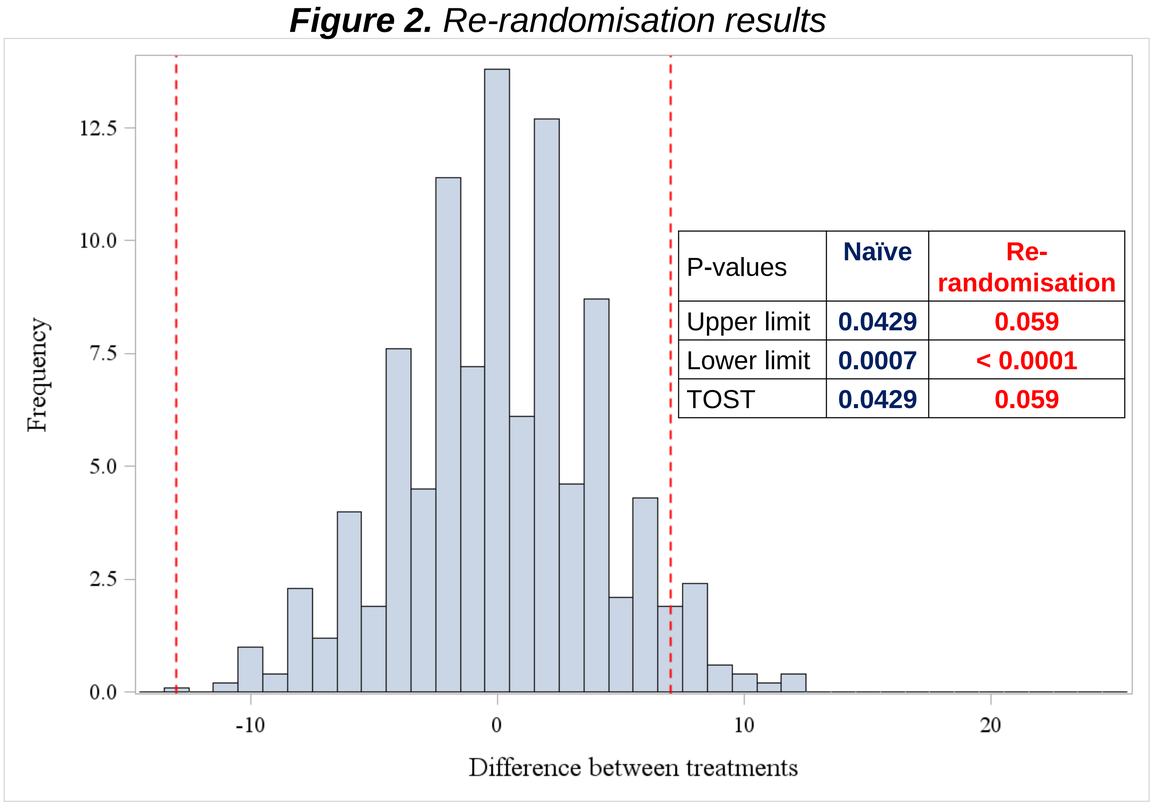
- Thinking is reversed: the null hypothesis of no difference becomes the alternative, so that re-shuffling is done under the alternative, rather than the null.
- Thus the primary goal of re-randomisation wouldn't apply here, in that we'd be 'testing' the system under the alternative.
- The framework illustrated in Table 1 might still be used, its results actually being thought more as a 'back-up' to the original study

Table 2. *Simulation results - TOST*

	Proportion of Responders		Difference	90% CI
	Treatment A	Treatment B		
True	0.5	0.6	0.1	-
*Simulated**	0.51	0.54	-0.03	-0.097, 0.037†

*300 patients for each group. Seed used for simulation = 5.
† TOST p-value = 0.0429 (using a margin of -0.1 to 0.1).

Figure 2. *Re-randomisation results*



- Following suggestions from [2] and [3], the alpha level would need to be calibrated/corrected to α^c to ensure asymptotic tests maintain the nominal α level.
- The R package EQUIVNONINF, via function `bi2diffac`, provides solution for cases where Central Limit Theorem holds (e.g. the standard TOST above), and in our simulation would lead to $\alpha^c = 0.036$, thus 'invalidating' the claim of equivalence based on naïve TOST results.
- For non-parametric permutation-like tests like the one described here, simulations are required (future work)

Conclusion

- Re-randomisation techniques, whilst largely applicable in superiority trials, are difficult to apply in equivalence due to the reversing of null and alternative hypotheses
- Nevertheless, the proposed solution can be implemented, to back up study results, and due to its non-parametric foundation can be useful in scenarios where the underlying test assumptions might be violated
- As such this approach is a valuable alternative/support to standard parametric testing, and can be implemented in most standard software (though feasibility depends on sample size, ultimately)

References

1. EMA/CHMP/295050/2013 *Guideline on adjustment for baseline covariates in clinical trials.*
2. Wellek S. *Testing statistical hypotheses of equivalence and noninferiority. Second edition.* Boca Raton: Chapman & Hall/CRC Press, 2010
3. R. Arboretti, E. Carrozzo, F. Pesarin, L. Salmaso, *Testing for equivalence: an intersection-union permutation solution.* arXiv:1802.01877

All correspondence to:
enquiries@quanticate.com