A burn-in(g) question: How long should an initial equal randomization stage be before Bayesian response-adaptive randomization?

Stef Baas
Collaborators: Edwin Y.N. Tang, Daniel Kaddaj, Lukas Pin, David S. Robertson, and Sofía S. Villar

October 24, 2025

Bayesian response-adaptive designs and burn-in period

- Clinical trial with binary data (success/failure), two treatments (**C**ontrol or **D**evelopmental), using Bayesian *response-adaptive randomization* (RAR):

  Probability of allocating participant $i+1$ to control is $P(\theta_C \geq \theta_D | \text{Data}_i)$,

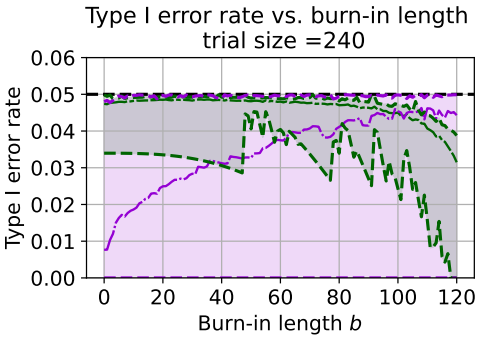
  where θ_a success probability for arm $a \in \{C, D\}$.
- Problems with inference under standard tests: type I error rate inflation, power deflation, "random highs" for the inferior treatment arm.
- Burn in period: initial non-adaptive stage of trial, used to "stabilize RAR".
- Usually no justification of the choice of burn-in length given, however it can negatively influence patient benefit (too long) or inference (too short).

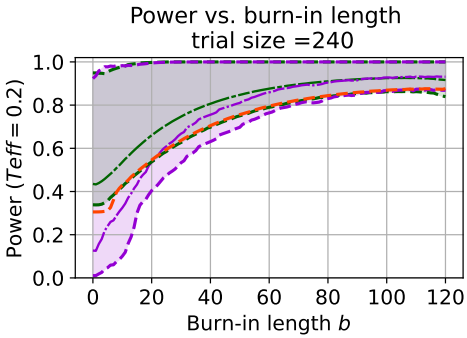
Exact analysis approach

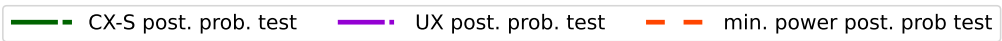
- In Baas et al. (2025) [1] we developed an efficient method for exact computation operating characteristics (e.g., type I error rate, power) and exact tests (i.e., tests that ensure a type I error rate bounded by the nominal significance level):
  - conditional exact test based on total successes (CX-S), extending Fisher's exact test,
  - unconditional exact (UX) test, extending Barnard's test.
- Suggestion: evaluate operating characteristics (OCs) Bayesian RAR for all burn-in lengths, choose the burn-in length that provides an optimal trade-off of OCs.
- In paper:
  - standard (calibrated) tests show type I error rate inflation to almost three times the significance level (15% vs. 5%) under misspecifcation of assumed parameters,
  - having burn-in length least half the trial size yields well-behaved type I error rates,
  - for robust type I error control: use exact tests.

[1] S. Baas, P. Jacko, and S.S. Villar. Exact statistical analysis for response-adaptive clinical trials: A general and computationally tractable approach. Computational Statistics & Data Analysis, 211:108207, 2025. URL https://doi.org/10.1016/j.csda.2025.108207.

Evaluation type I error rate and power exact tests







Note: CX-S: conditional exact test; UX: unconditional exact; post. prob. test: posterior probability test rejecting if $P(\theta_C \geq \theta_D | \text{Data}_i)$ is large or small

Thank you for your attention!

- Any questions?
- Link to paper (with results for other measures such as patient benefit and estimation bias, and application to real-life ARREST trial)