# Robust Designs for Response-Adaptive Randomization

## ENAR 2020

Thevaa Chandereng

Biostatistics & Medical Informatics
University of Wisconsin–Madison

Joint work with Rick Chappell
github.com/thevaachandereng

# What is a Response Adaptive Randomization (RAR) Design?

- RAR designs utilize accumulating information on outcomes as the trial progresses in order to alter the treatment assignment probabilities and assign more patients to the better-performing treatment arm.
- Previous outcomes are used to tilt the randomization probabilities in favor of the treatment currently observed to have superior results.

# What is a Response Adaptive Randomization (RAR) Design?

- RAR designs utilize accumulating information on outcomes as the trial progresses in order to alter the treatment assignment probabilities and assign more patients to the better-performing treatment arm.
- Previous outcomes are used to tilt the randomization probabilities in favor of the treatment currently observed to have superior results.
- The trial needs to be short to be able to obtain the outcome of the trial for future randomization

# What is the issue with RAR?

Peto (1985):
" . . . patients entering a trial are often not uniform in prognosis. For example, both in the European Coronary Bypass Trial and in the Medical Research Council's 8th Acute Myeloid Leukaemia Trial there was a statistically significant trend towards better prognosis as time went by. If, therefore, there had also been a trend in allocation proportions towards the apparently better treatment, an appreciable bias might have been engendered."

Time- Trends in Clinical Trials!

# Why Time-trends are Common?

- Centers can change, such as when VAs enter the trial earlier or later than academic institutions.

# Why Time-trends are Common?

- Centers can change, such as when VAs enter the trial earlier or later than academic institutions.
- Patients within centers can change, especially but not only with chronic diseases, due to the phenomenon of “A queue of desperate patients lining up at the door”.

# Why Time-trends are Common?

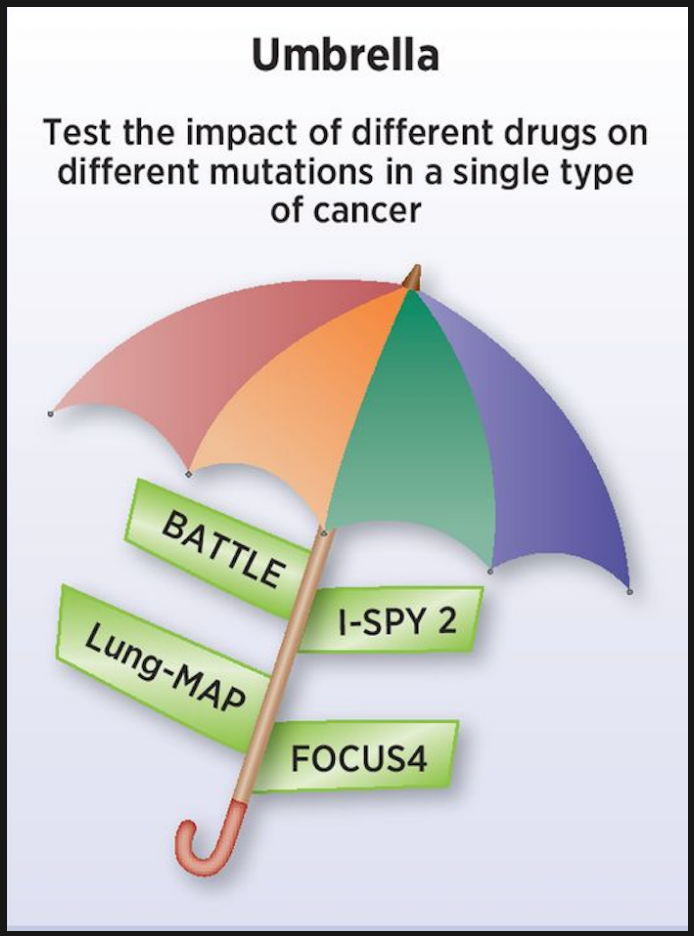
- Centers can change, such as when VAs enter the trial earlier or later than academic institutions.
- Patients within centers can change, especially but not only with chronic diseases, due to the phenomenon of “A queue of desperate patients lining up at the door”.
- In addition to these examples, an investigator who wants to game the system could cross his/her fingers that his favored treatment arm is ahead, then progressively enroll better prognosis patients over time.

# Time-trends in RAR

- Thall et al. (2015) investigated type-I error under a linear time-trend induced in the traditional response-adaptive randomization design and showed that the type-I error is significantly above the nominal level.

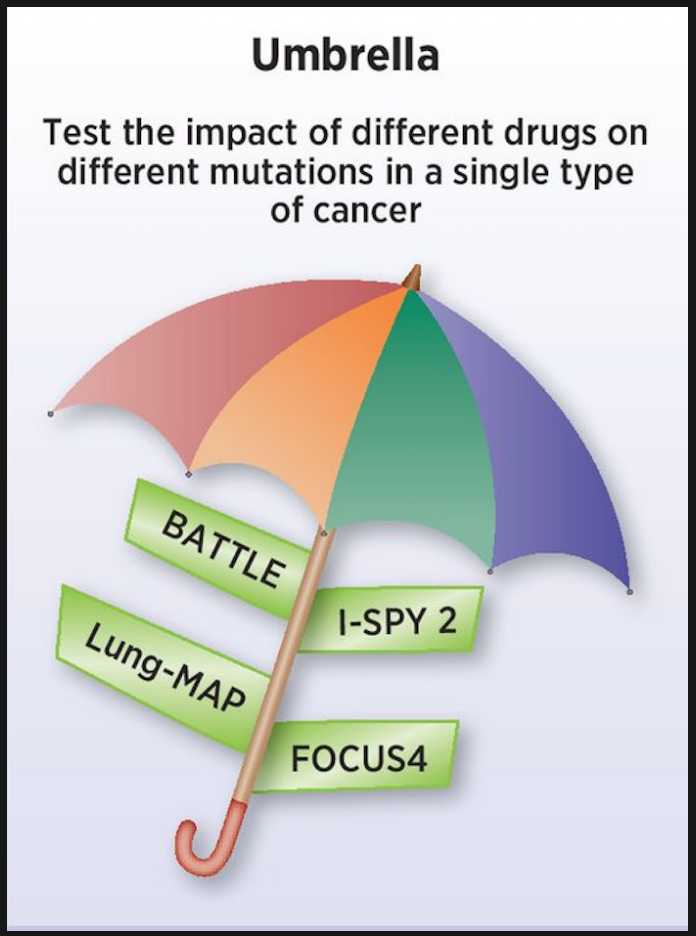
# BATTLE-1 Trial

Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE - 1)



# BATTLE-1 Trial

Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE - 1)




- 4 treatments, 5 biomarker;
- primary end point, 8-week disease control rate;
- treat more patients in promising groups based on biomarker profile;
- suspend ineffective groups early

# BATTLE-1 Trial

# BATTLE-1 Trial

“More smokers enrolled in the latter part of the study compared to the beginning of the study”.

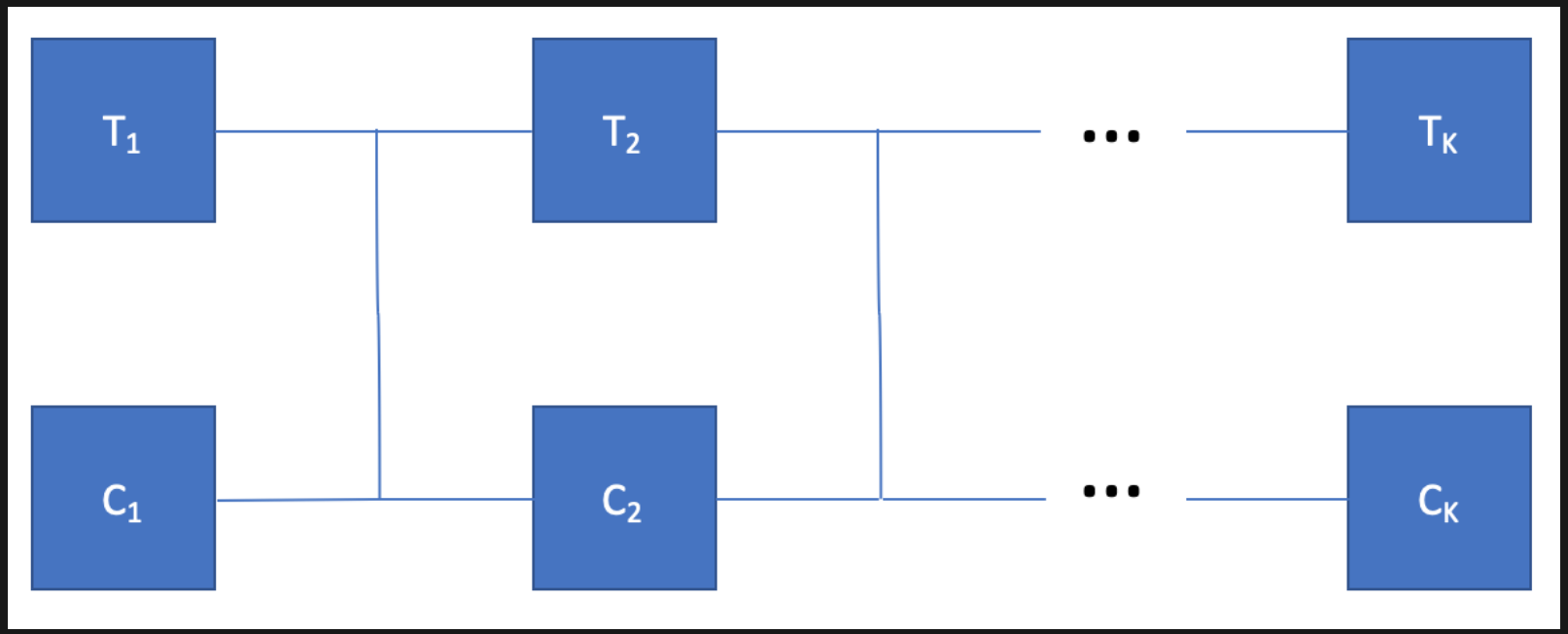
# BATTLE-1 Trial

"More smokers enrolled in the latter part of the study compared to the beginning of the study".

Time trends are nearly universally ignored among RAR proponents.

# Blocked Design for RAR

- Group-sequential procedure.
- Update allocation ratio for groups rather than patients.



# Simulation Setup

- $N$ = 200 subjects
- $K$ = 1, 2, 4, 5, 10, 20, 100, 200.
- 10,000 independent simulations were performed
- $p_A$ is set to 0.25, $p_B$ is set to 0.25 (null case), 0.45.
- Run-in period of 20 patients (fixed randomization).
- Allocation probability bounded between 0.2 and 0.8.

# Simulation Setup

To examine the effects of time-trends, we increased both $p_A$ and $p_B$ linearly from their initial values to a final value of 0.25 larger.

- $p_A(n) = 0.25 + 0.25(n/N)$, where $n$ is the patient that have been accrued and $N$ is the total sample size.
- for null case, $p_B(n) = 0.25 + 0.25(n/N)$.
- for alternative case, $p_B(n) = 0.45 + 0.25(n/N)$.

# Simulation Setup: Bayesian Design

- The probability of randomizing subjects to treatment A in stratum j, $\pi_{j,A}$ is defined as

$$\pi_{j,A} = \frac{(p_{A>B,j-1}(data))^{n/2N}}{(p_{A>B,j-1}(data))^{n/2N} + (p_{B>A,j-1}(data))^{n/2N}},$$

  where $p_{A>B,j-1}(data)$ is the posterior probability that treatment A has a higher success rate than treatment B, $p_{A>B,j-1}(data)) = 1 - p_{B>A,j-1}(data))$ after the $j-1^{st}$ block, $n$ is the number of accrued patients and $N$ is the maximum sample size. .

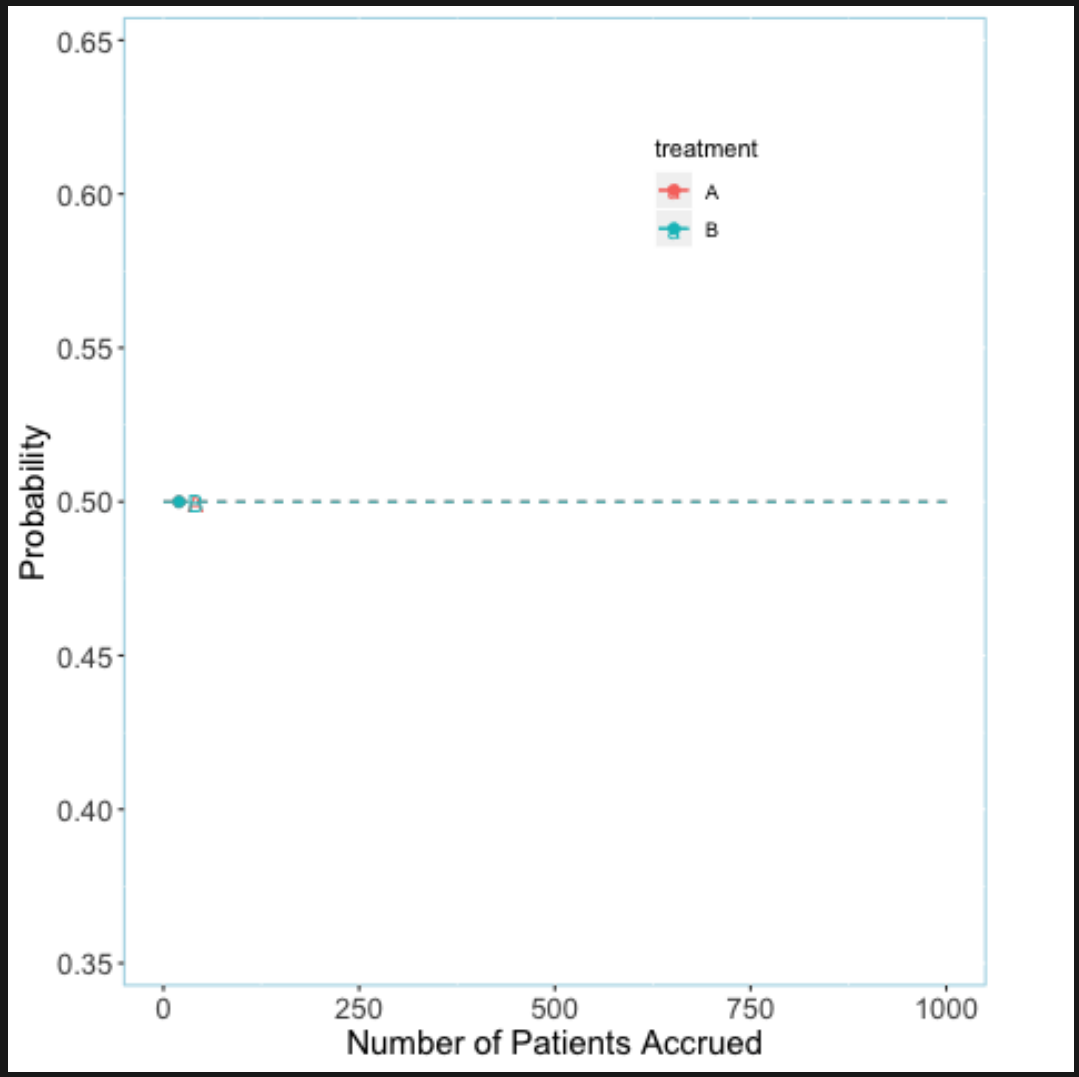
- Beta-binomial conjugate prior is used with $\mathcal{B}eta(0.5, 0.5)$ priors.

# Simulation Setup: Bayesian Design

- If treatment B is selected to be better than treatment A, then if $P_{B>A}(data) > .99$, the trial is stopped early for success and if $P_{B>A}(data) < .01$, the trial is stopped early for failure.
- For non-stratified design, where $K$ = 1, 200, if the final posterior probability $P_{B>A}(data) > .95$, then treatment B is declared superior to treatment A.
- For stratified design by block, Bayesian logistic regression was implemented to estimate the posterior probability of treatment difference and block effects.

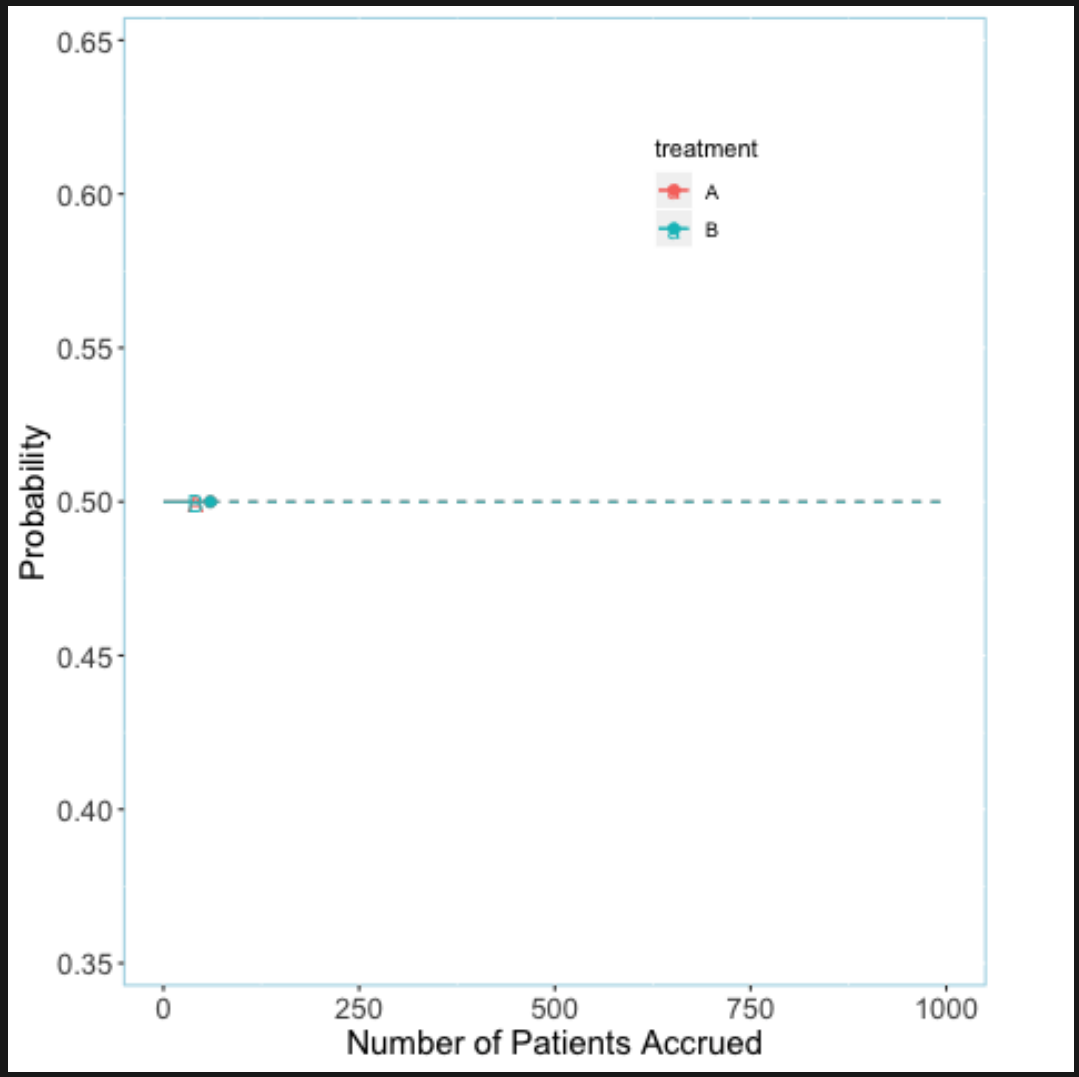
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



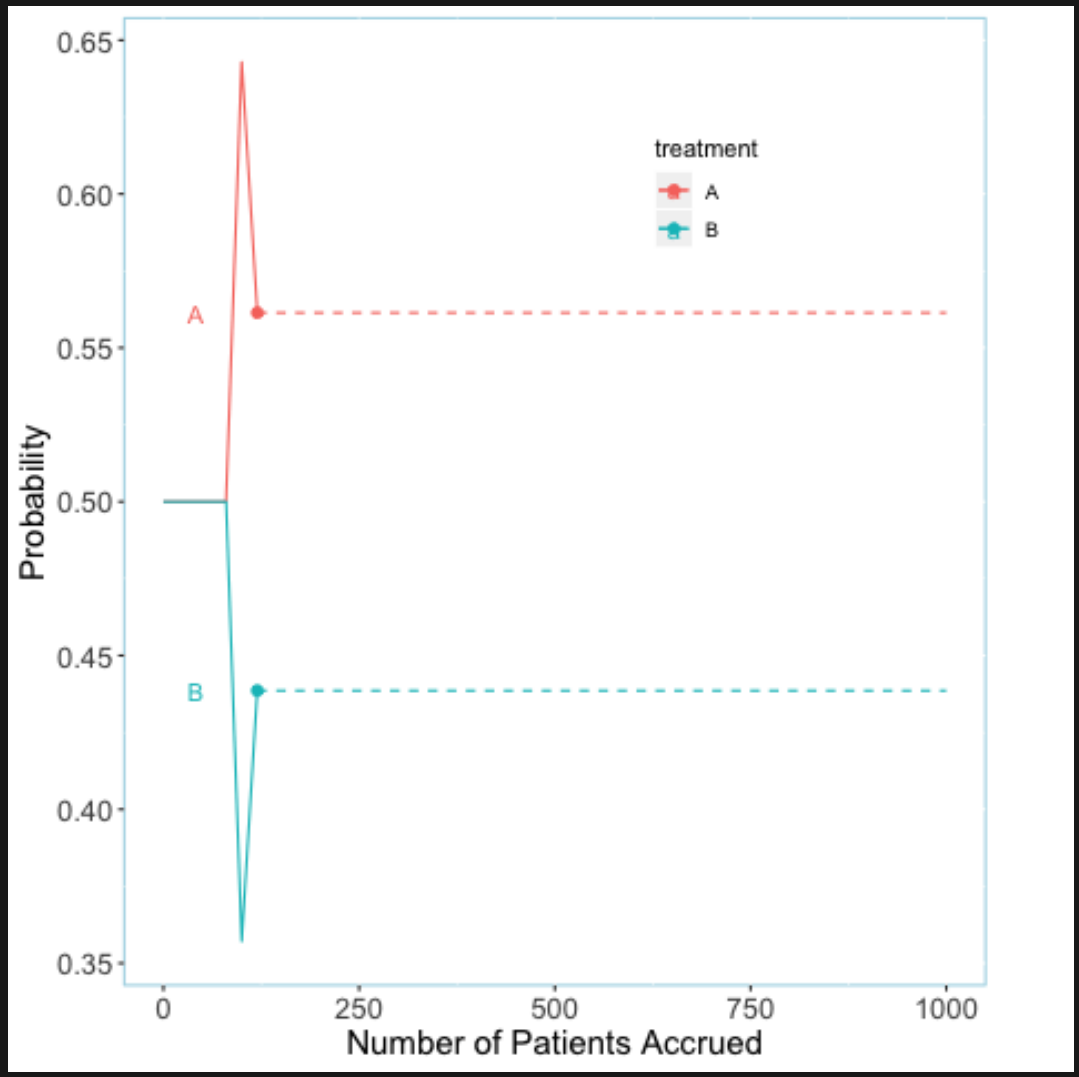
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



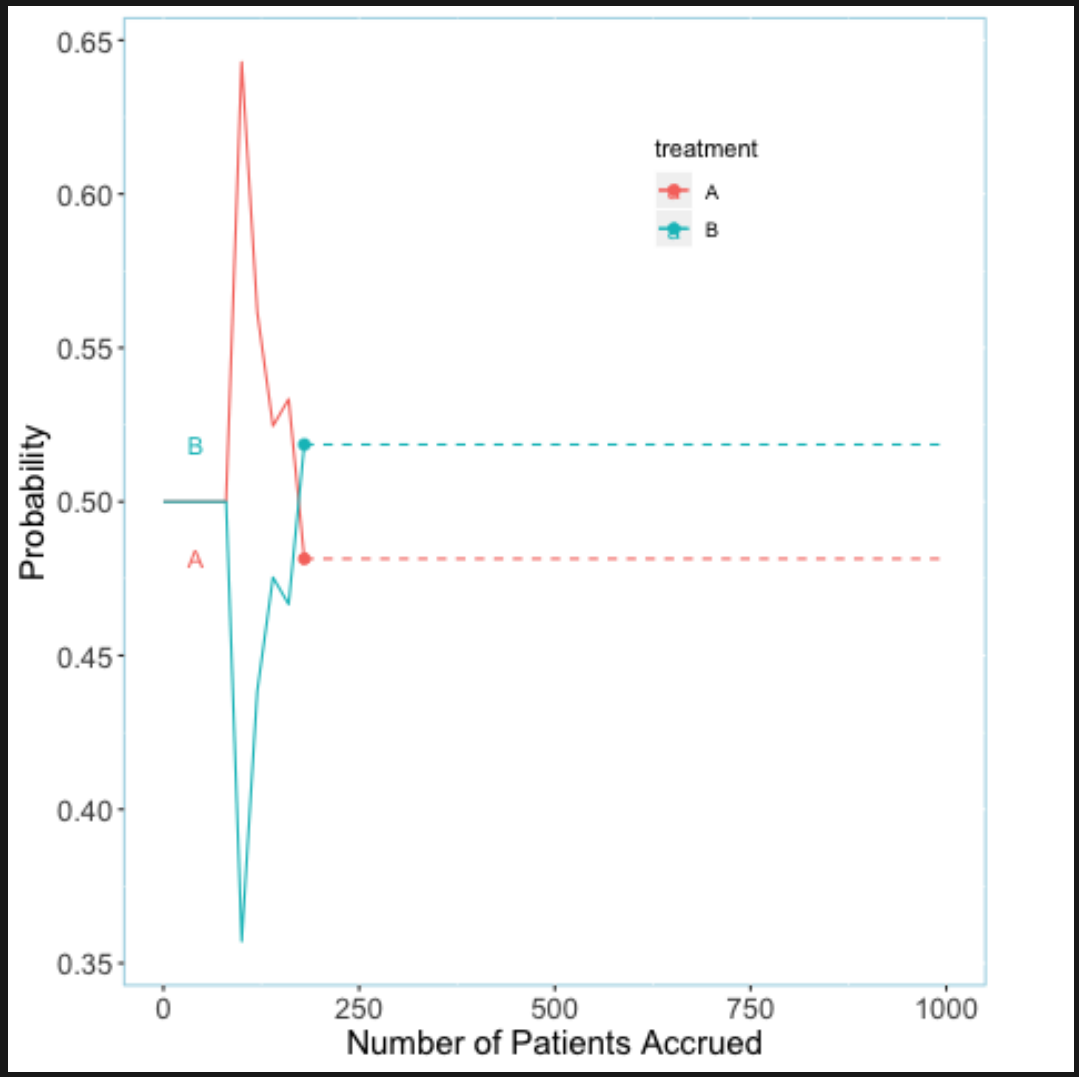
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



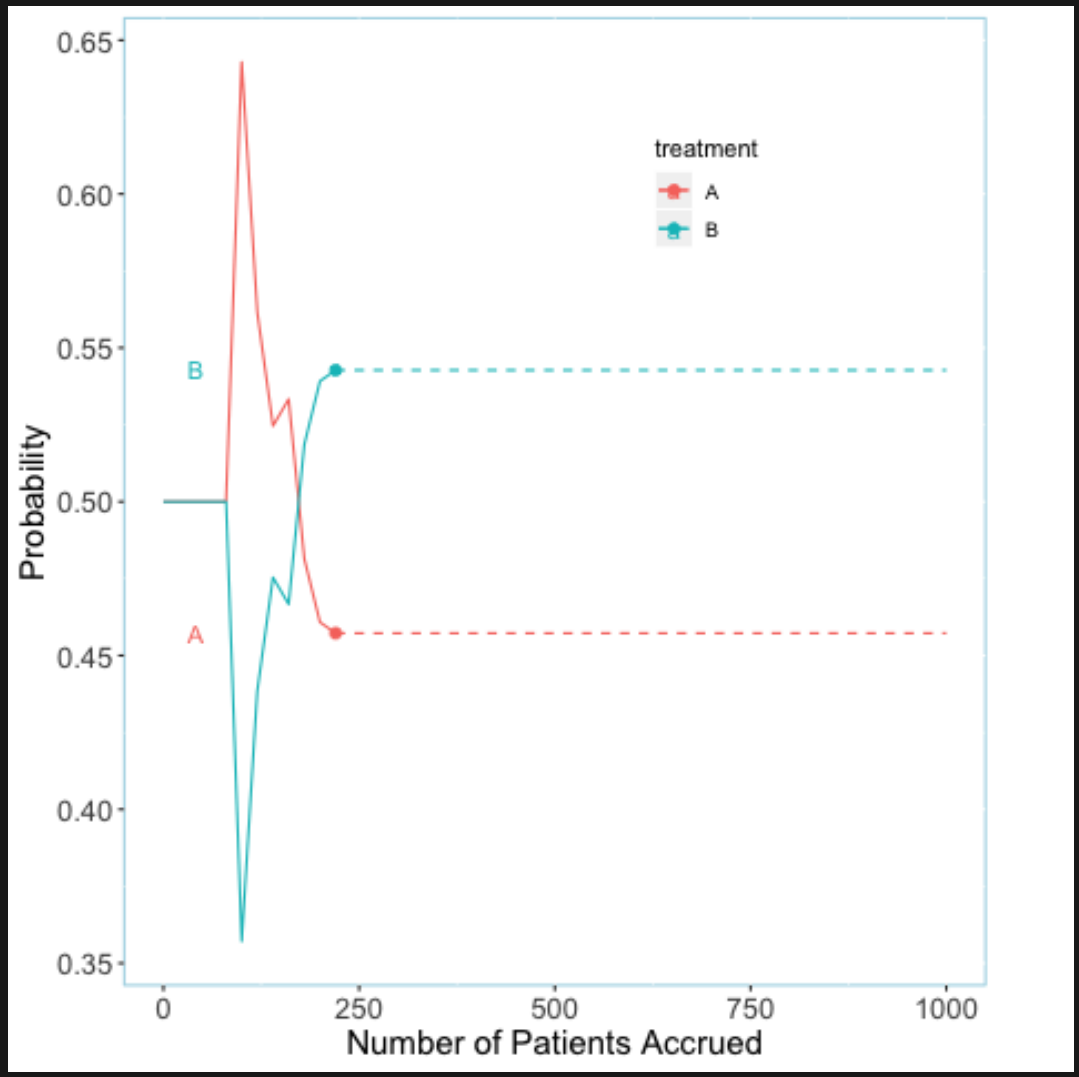
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



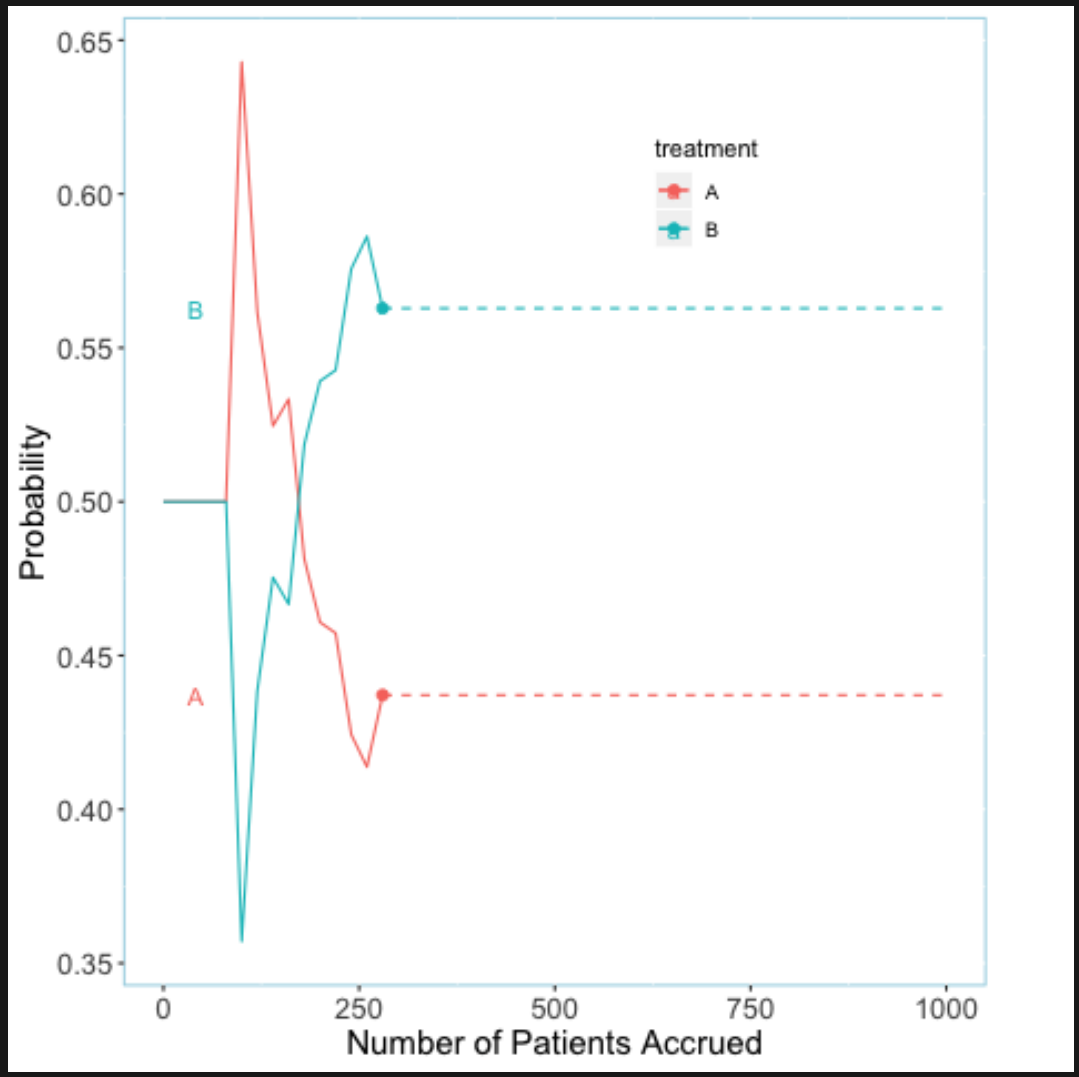
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



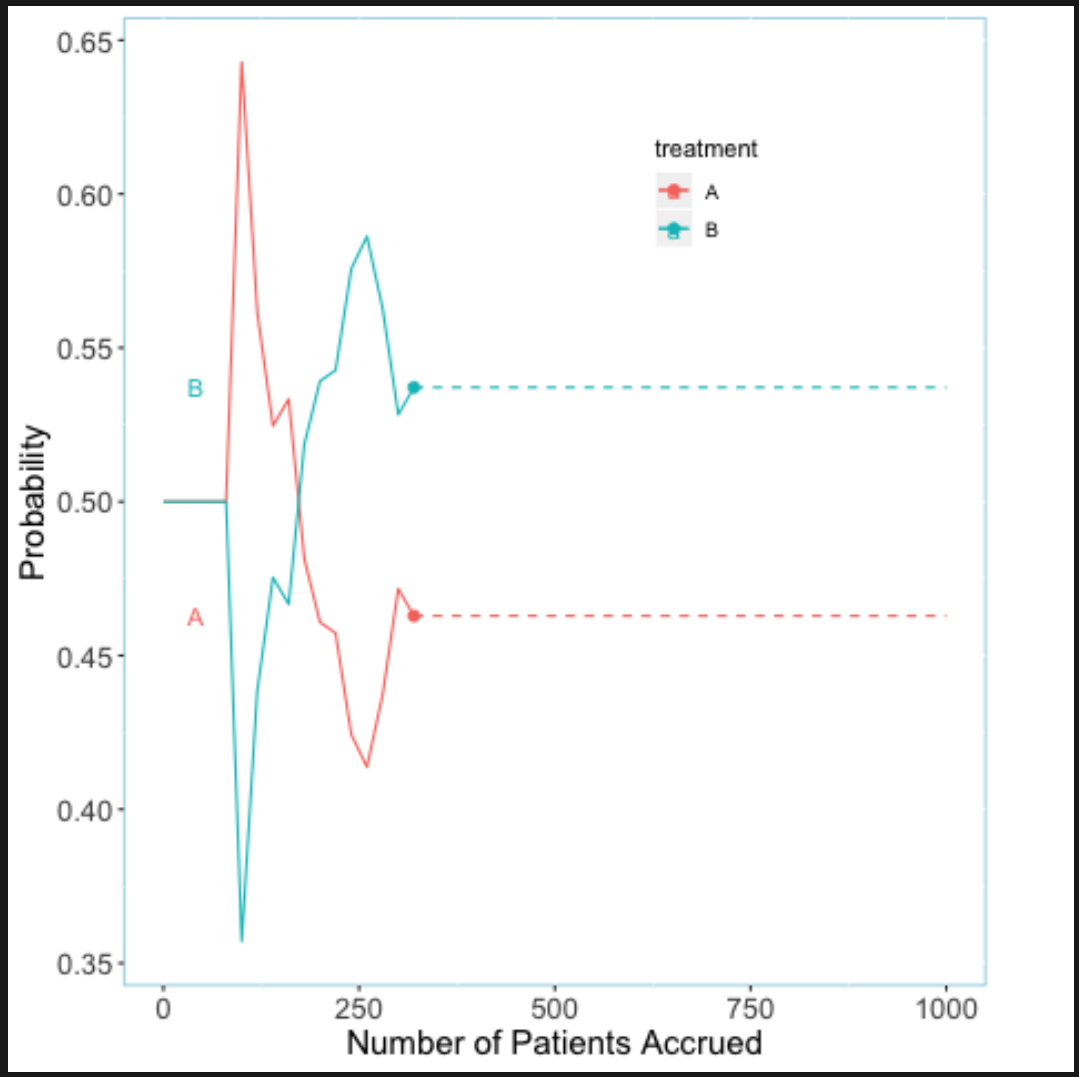
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



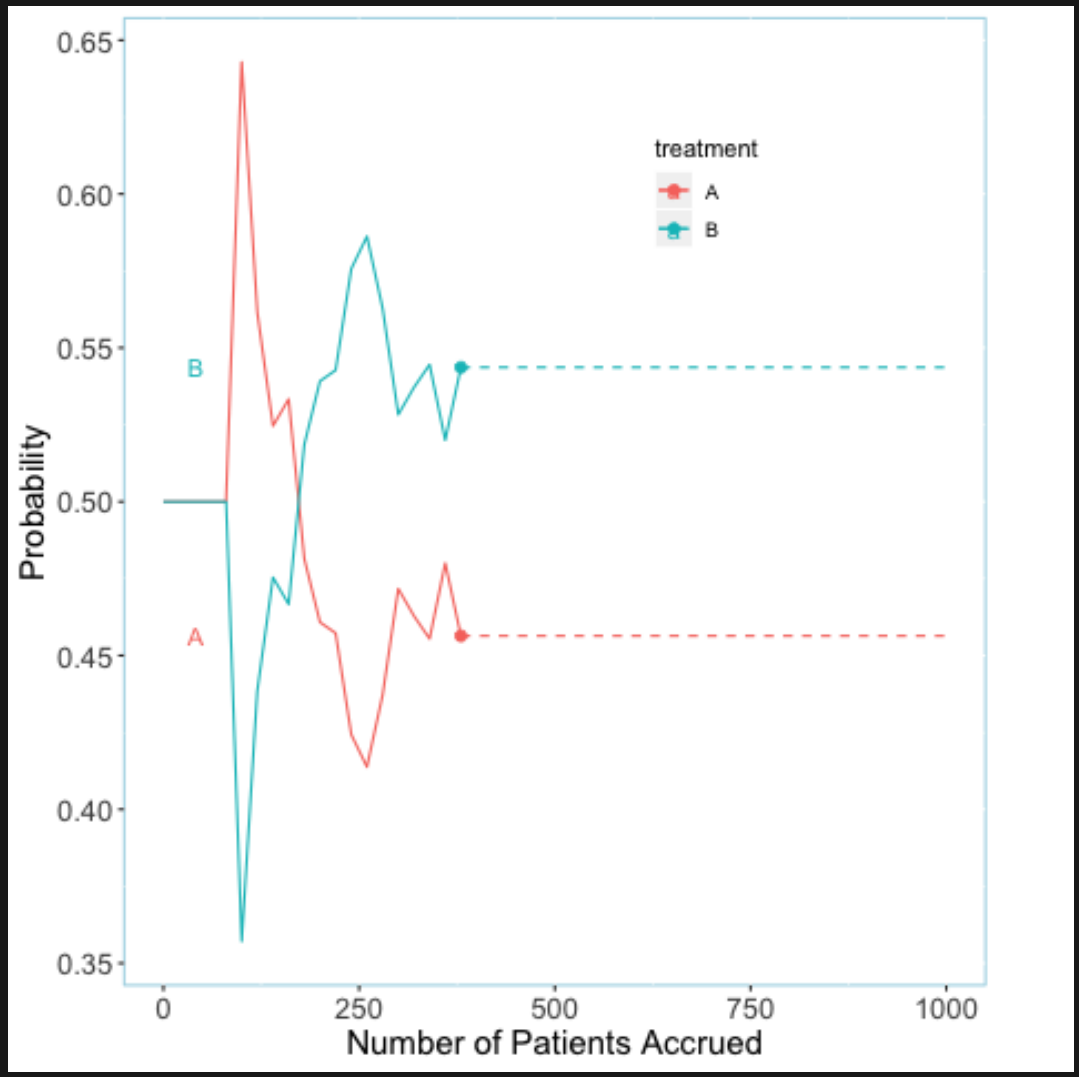
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



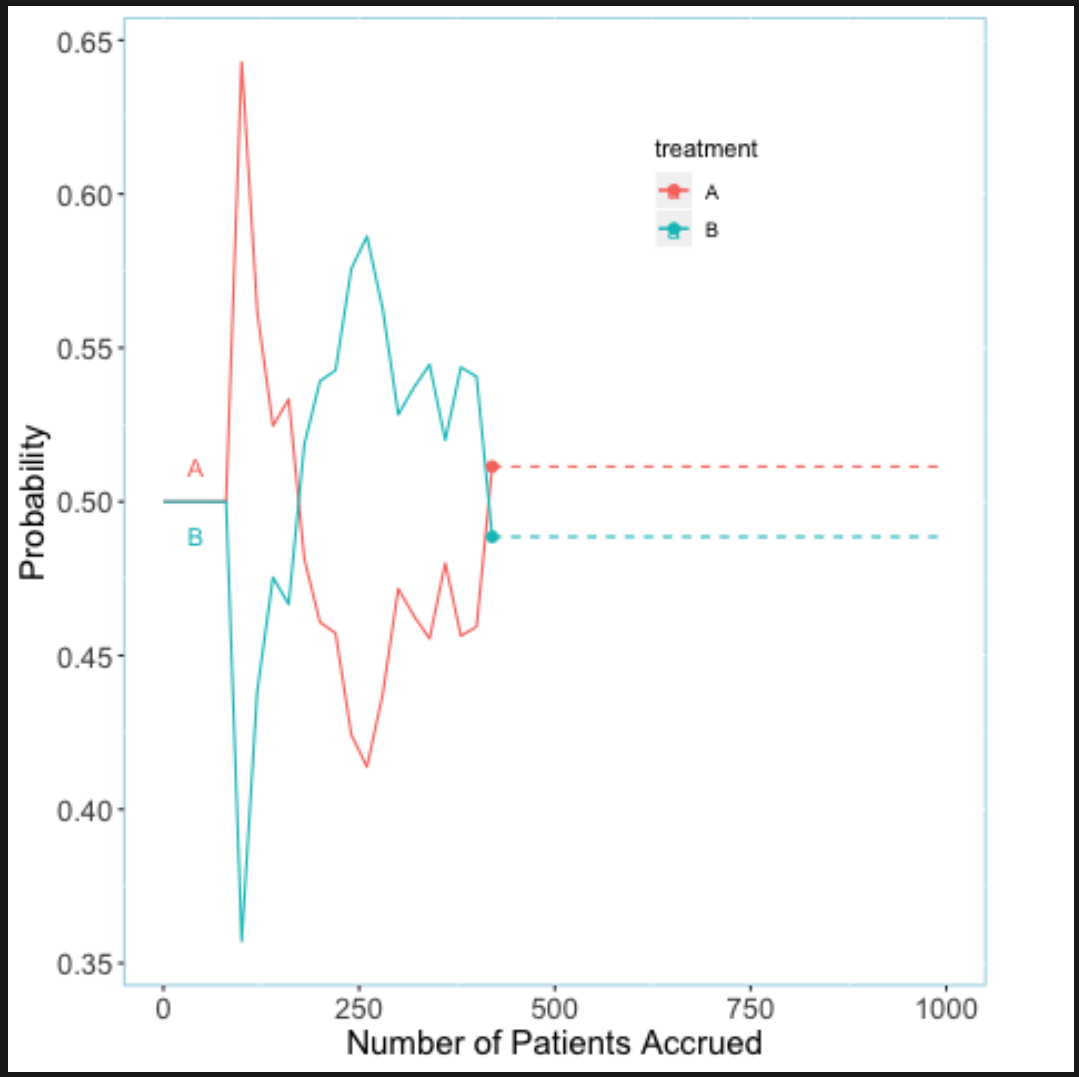
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



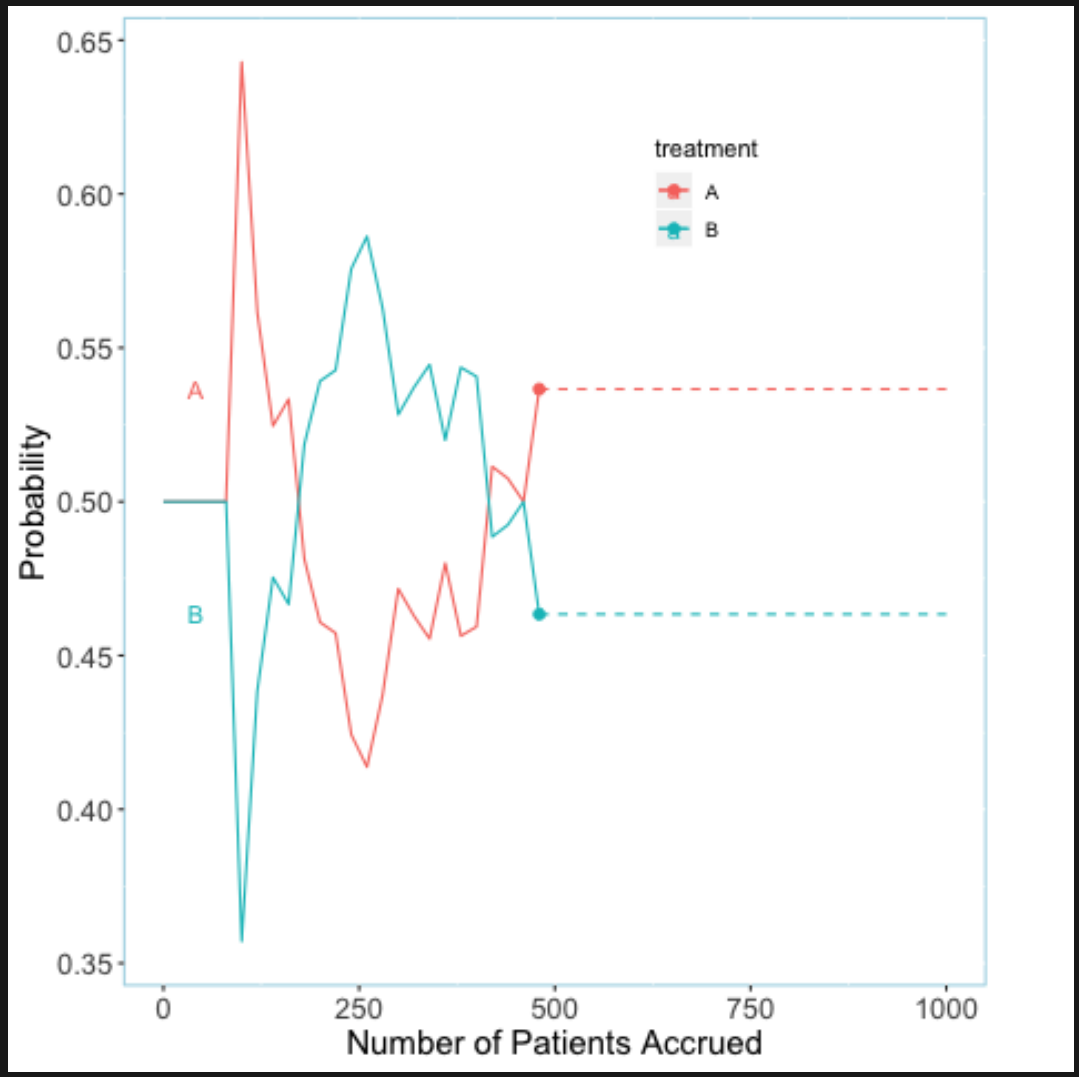
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



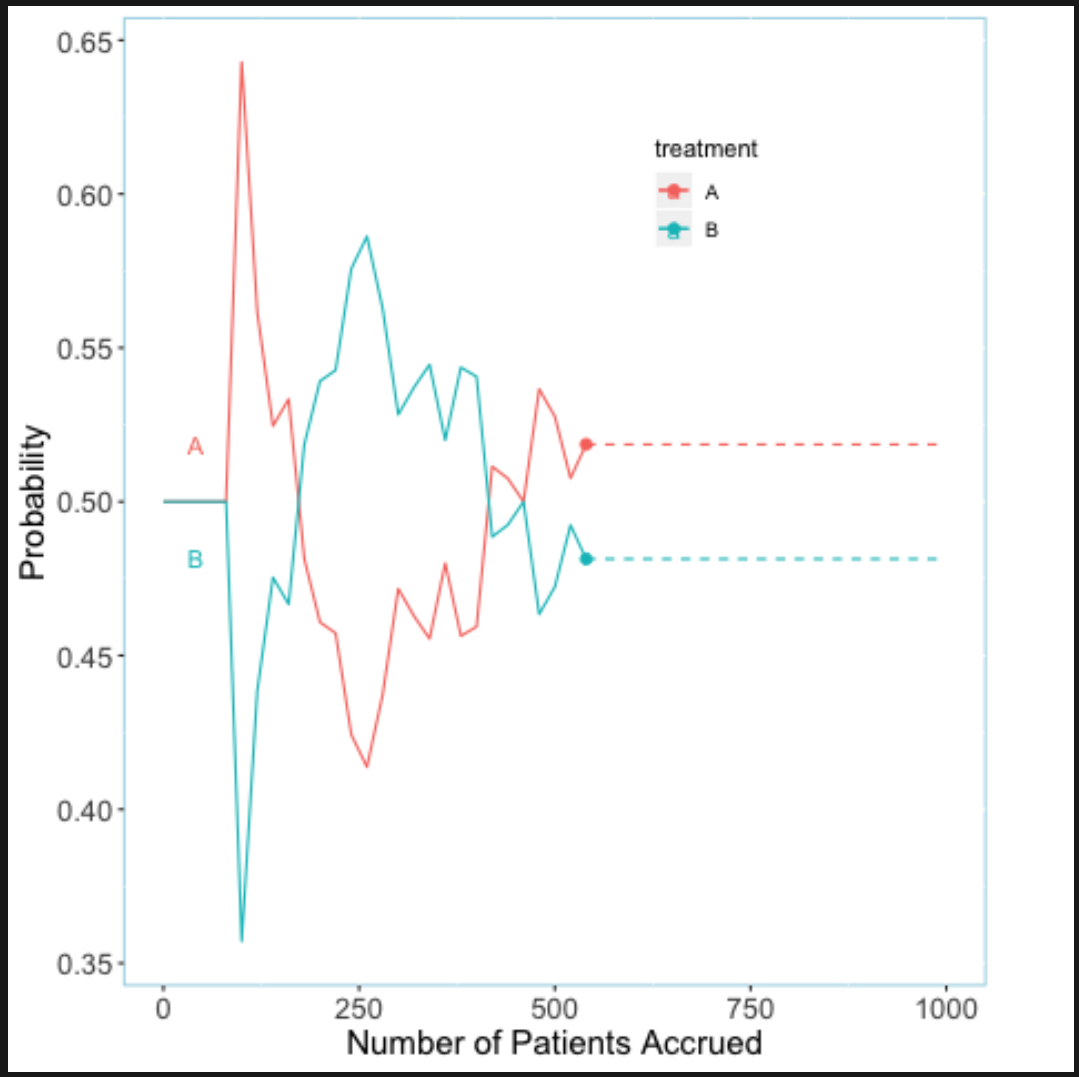
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



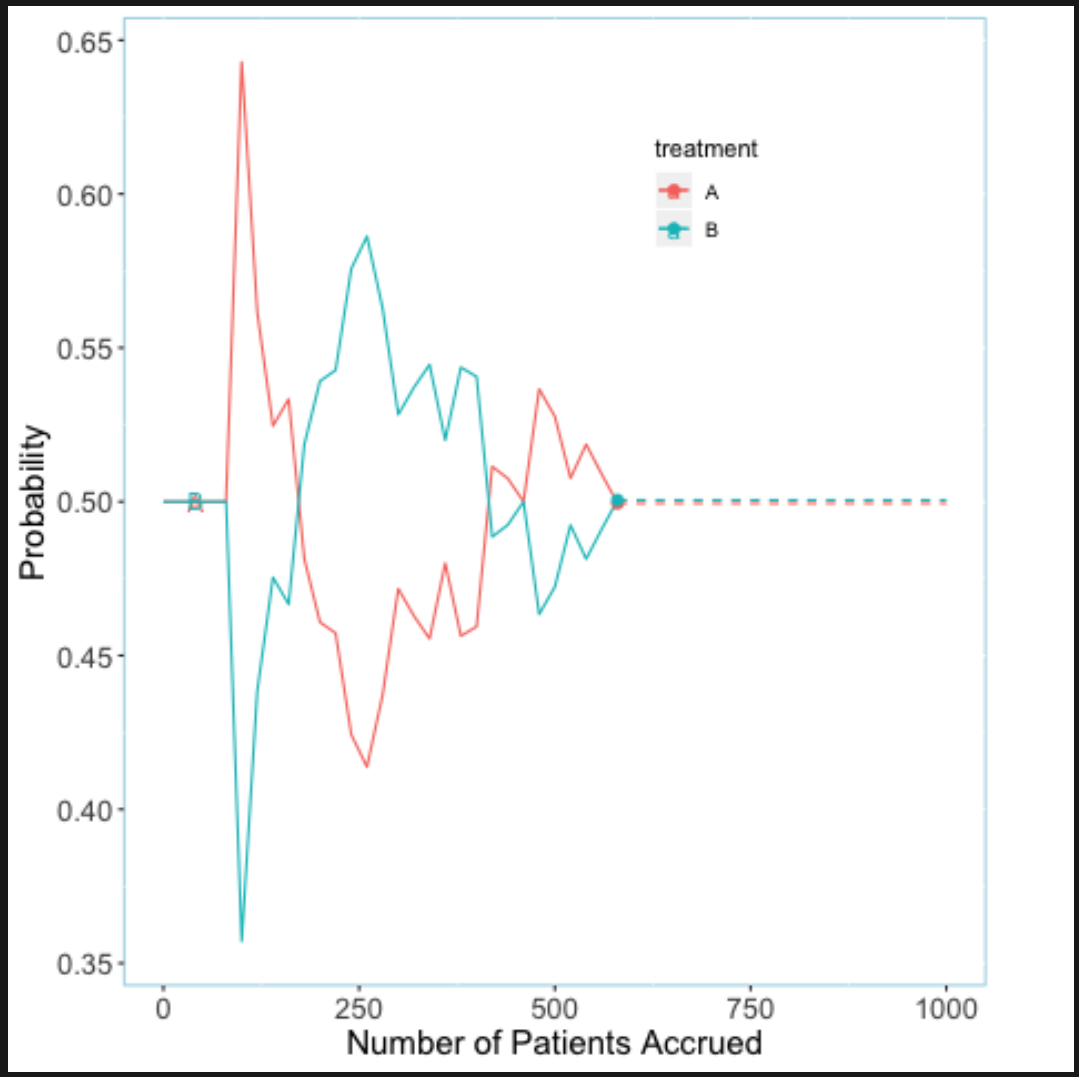
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



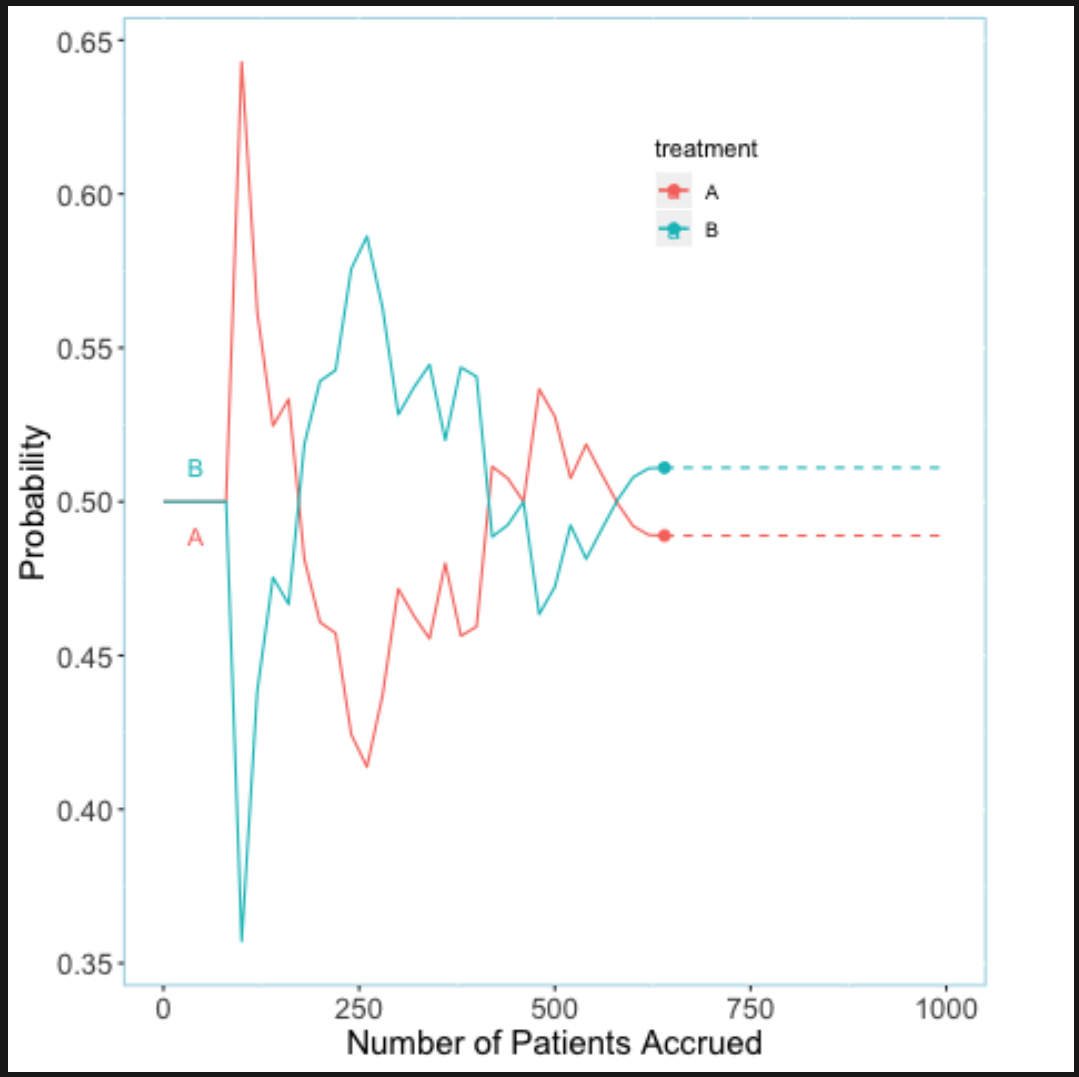
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



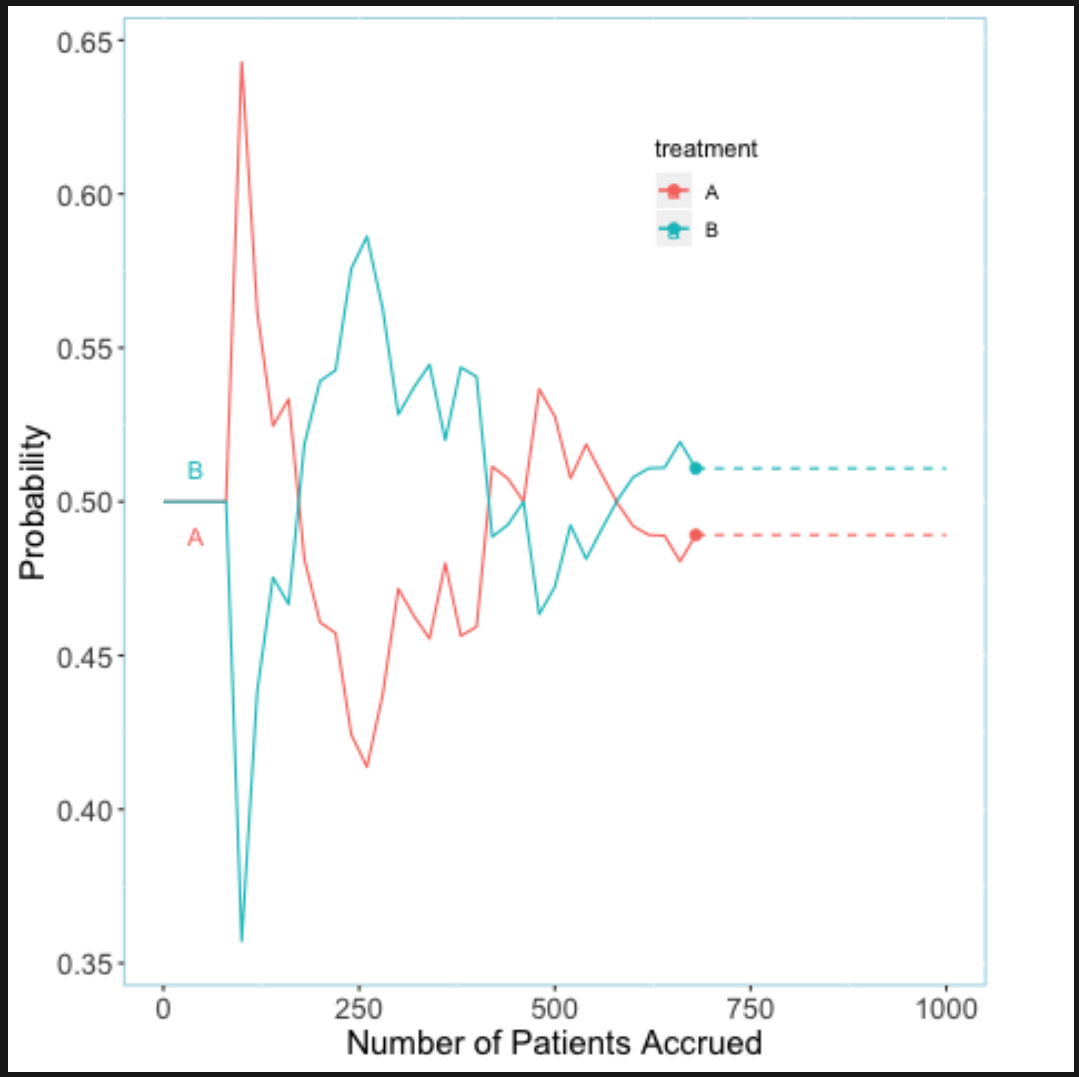
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



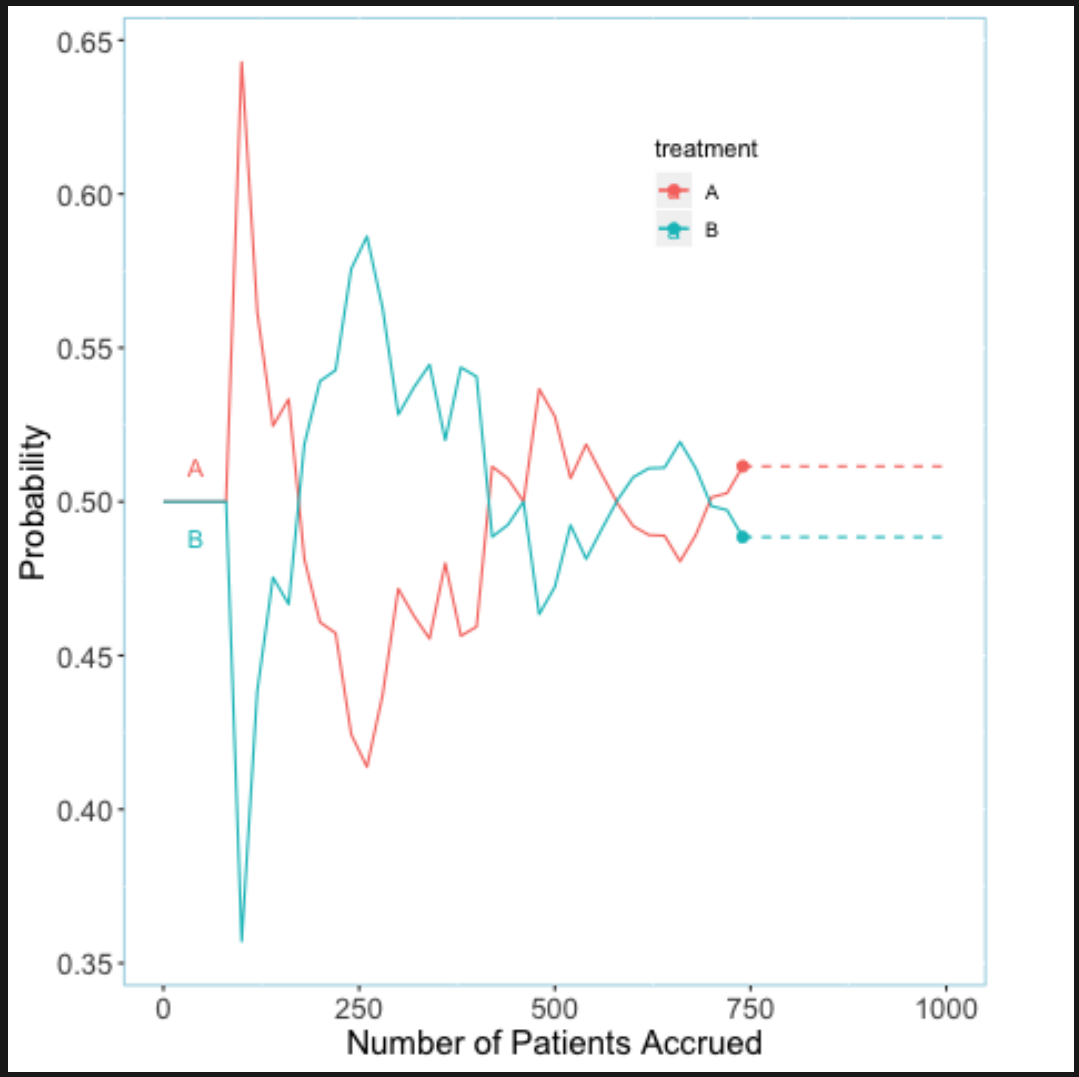
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



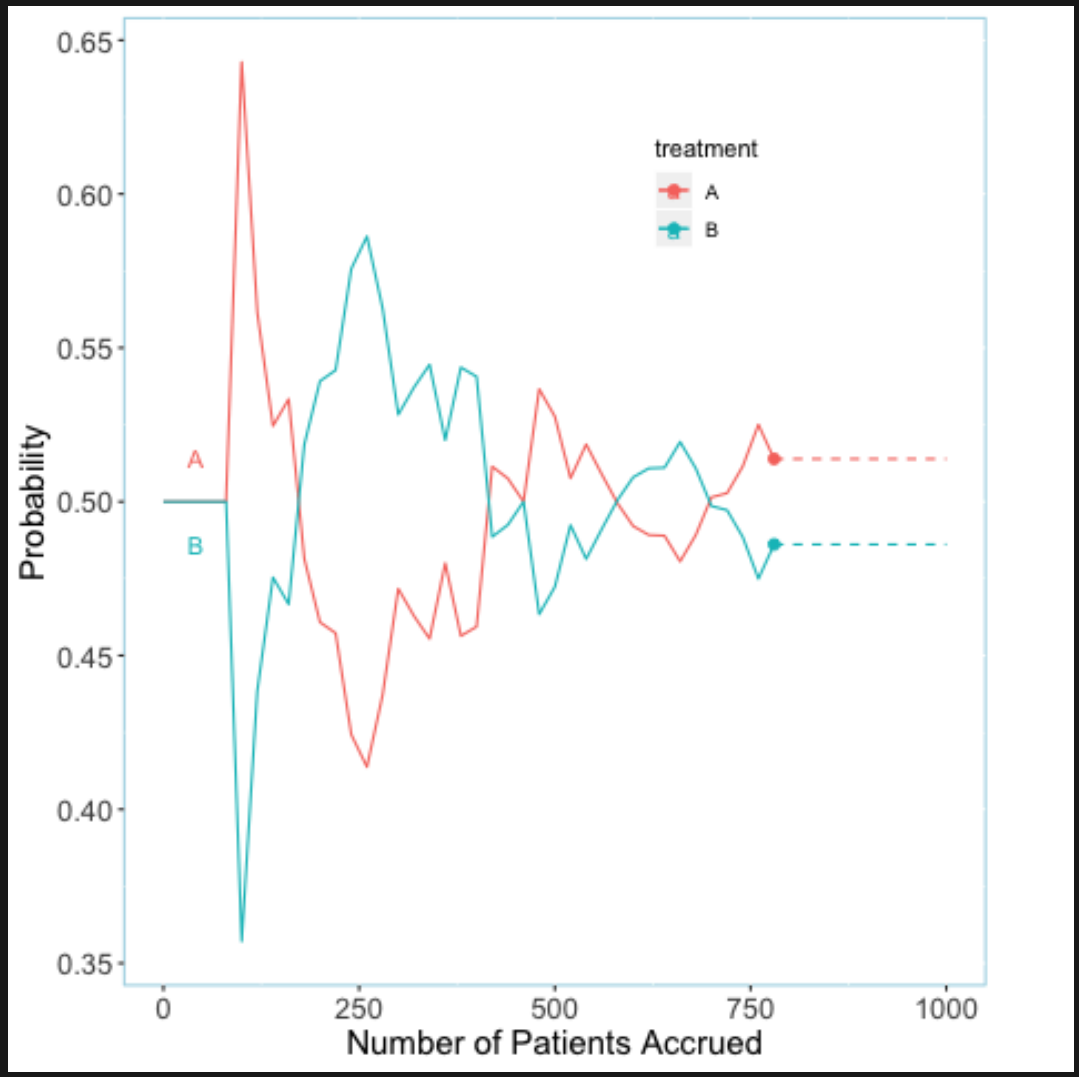
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



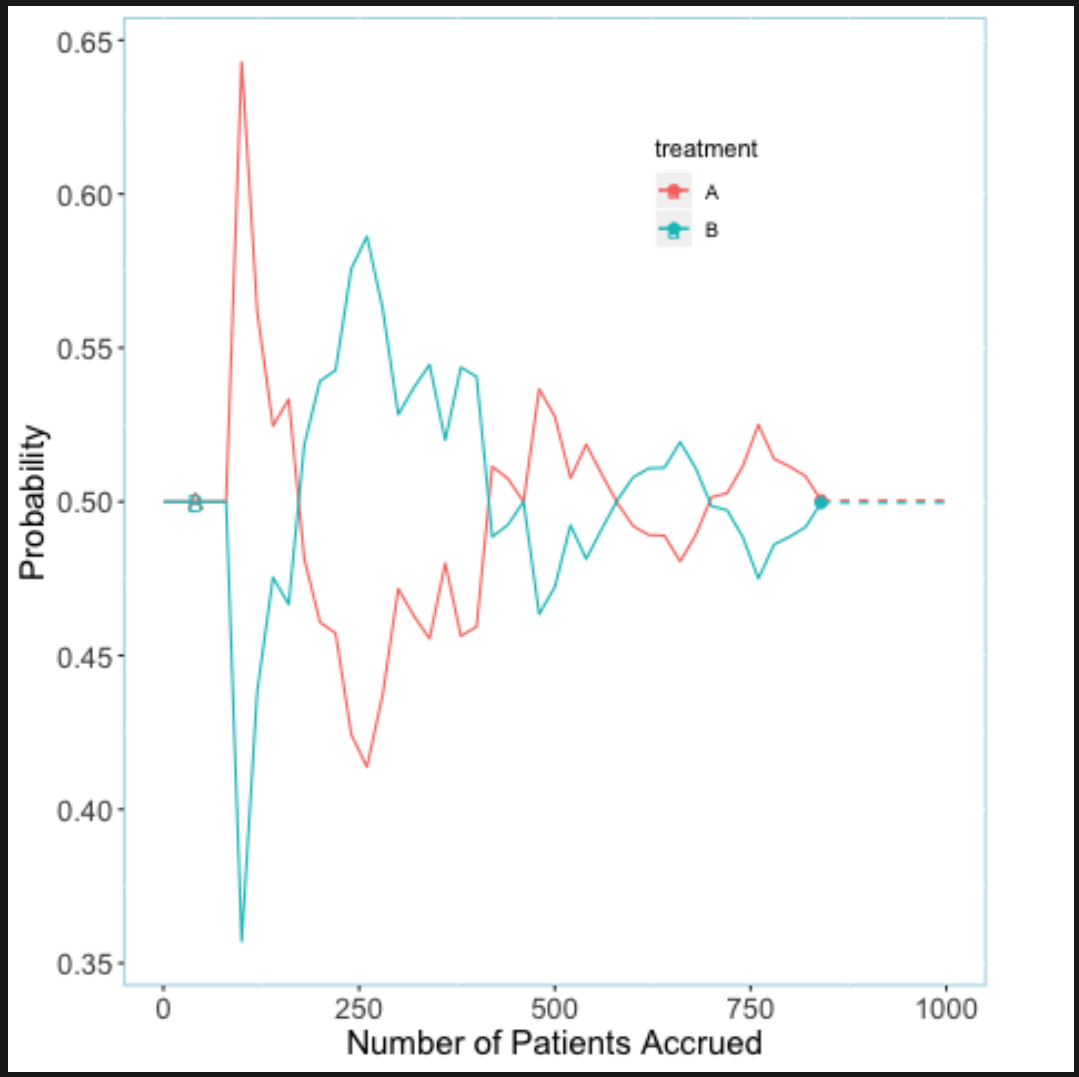
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



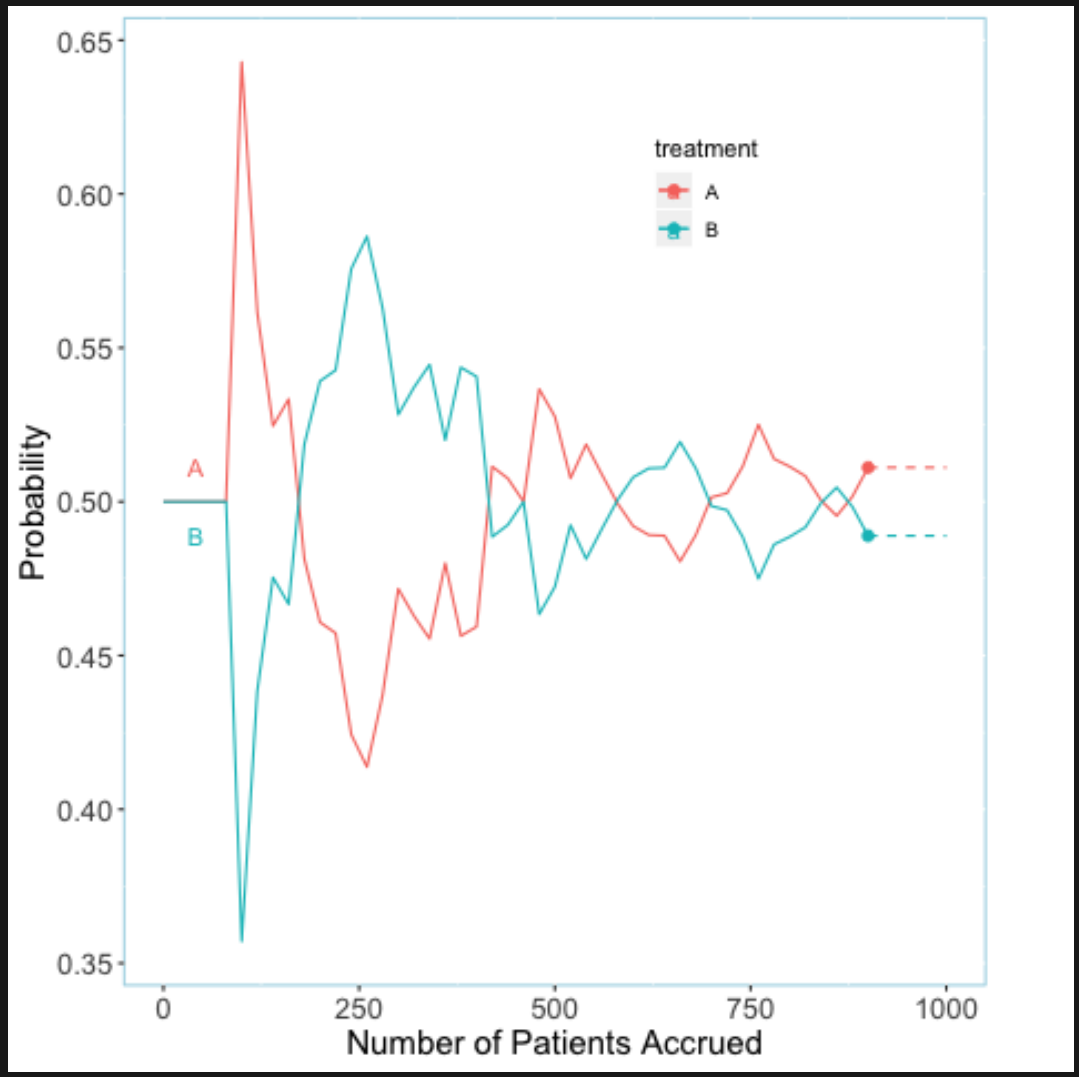
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



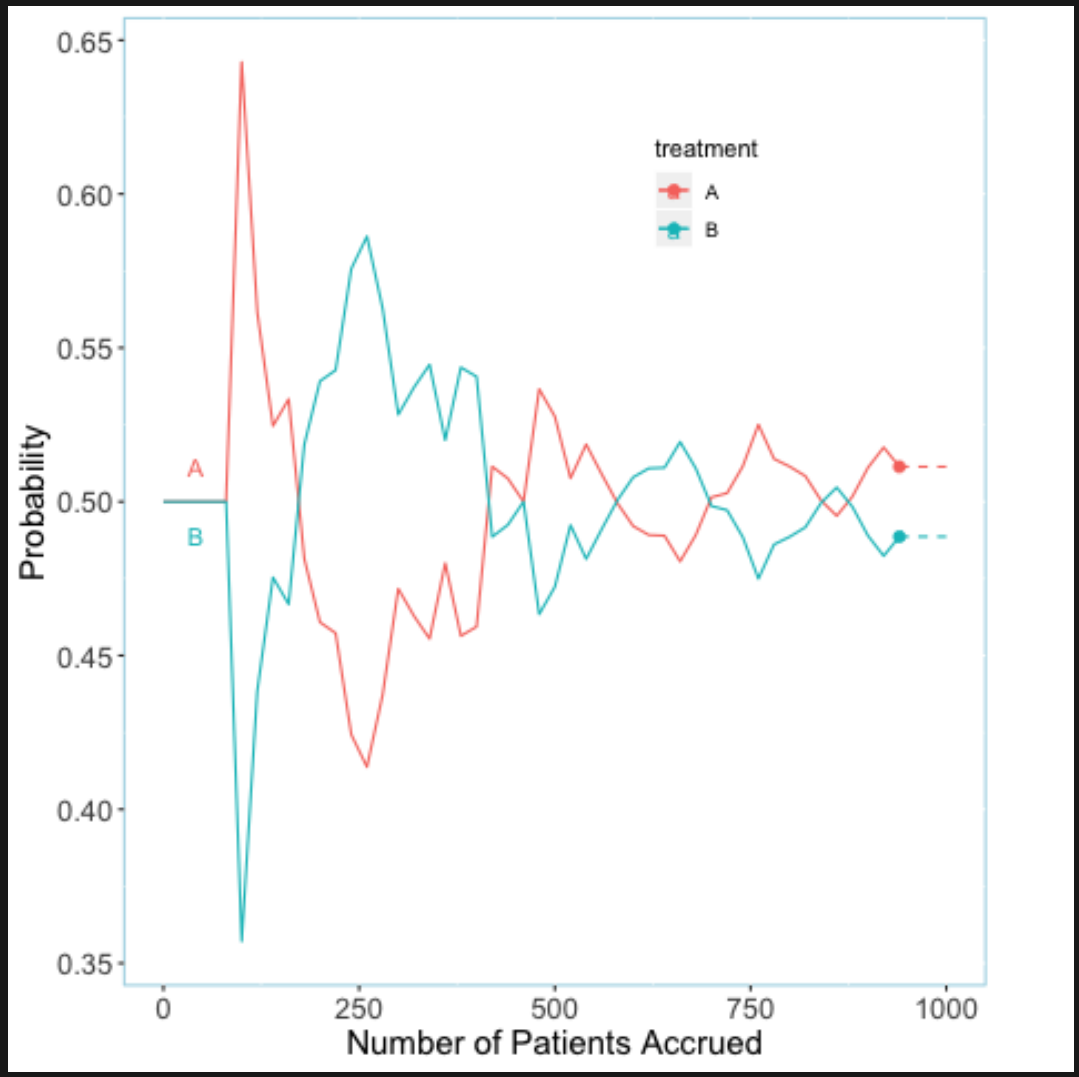
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



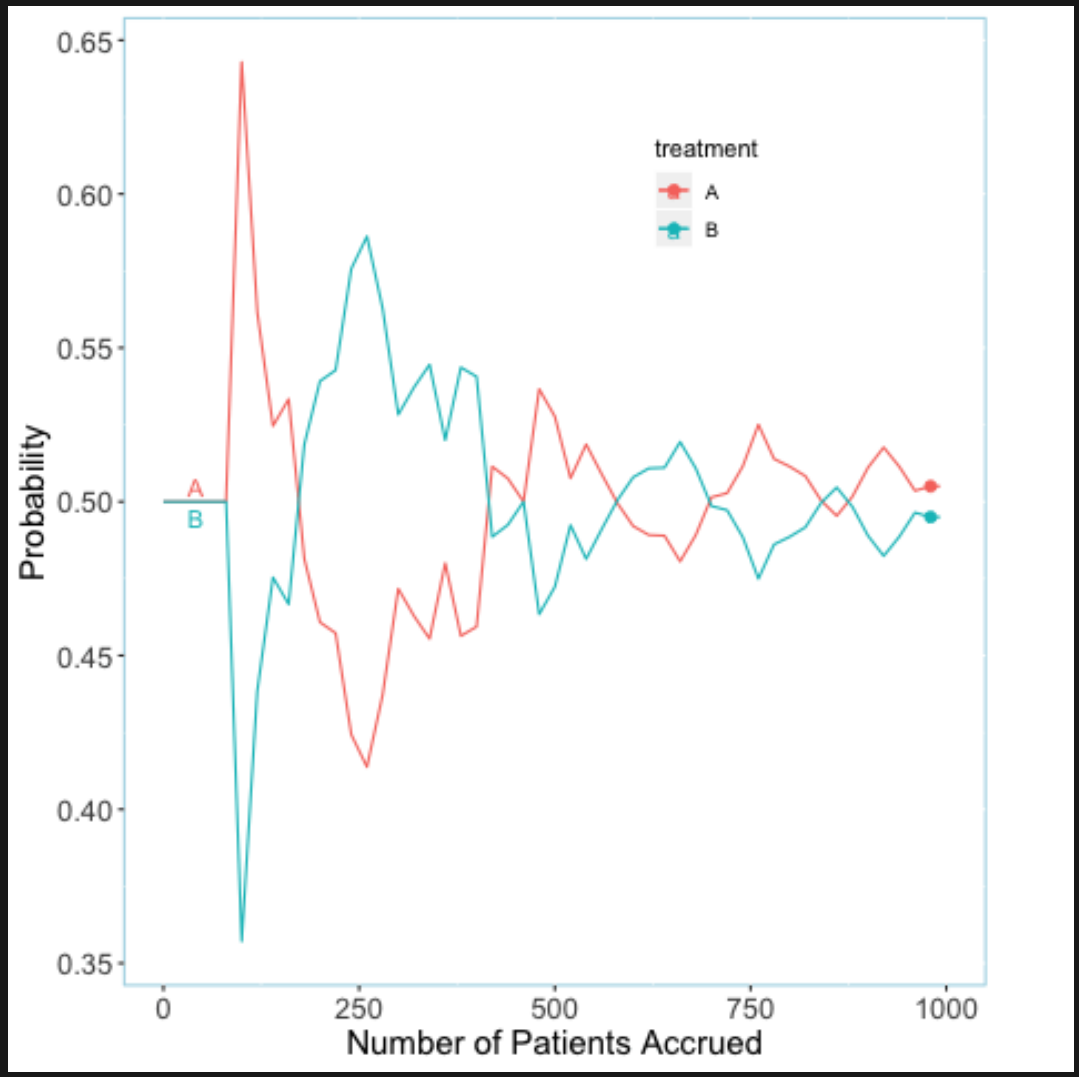
# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



# Simulation

A single simulation for RAR with $p_A = 0.07$, $p_B = 0.05$ with N of 1000 and 50 number of blocks.



# Results using Bayesian Approach: With Drift

| Block | Type I error | Bias | E(N) | $\pi_{20}$ | $N_B - N_A$ |
|---|---|---|---|---|---|
| FR | 0.05 | 0.00 | 180 | 0.06 | 0.27 (-28, 28) |
| 2 | 0.05 | 0.00 | 178 | 0.20 | -0.14 (-48, 48) |
| 4 | 0.07 | 0.00 | 178 | 0.24 | -0.09 (-56, 58) |
| 5 | 0.08 | 0.00 | 180 | 0.25 | 0.03 (-58, 58) |
| 10 | 0.08 | 0.00 | 179 | 0.27 | -0.39 (-62, 62) |
| 20 | 0.08 | 0.02 | 179 | 0.27 | 0.01 (-62, 62) |
| 100 | 0.03 | 0.01 | 177 | 0.25 | 0.02 (-60, 60) |
| trad RAR | 0.16 | 0.00 | 175 | 0.26 | -0.24 (-60, 60) |

Table: RAR using Bayesian approach and with 0.25 drift. $E(N)$ represents the mean sample size. $\pi_{20} = P(N_A - N_B > 20)$. $N_A$ and $N_B$ denotes the number of patients assigned to treatment A and B. The mean (2.5%, 97.5%) of $N_B - N_A$ is reported.

# Results using Bayesian Approach: With Drift

| Block | Power | Bias | E(N) | $\pi_{20}$ | $N_B - N_A$ |
|---|---|---|---|---|---|
| FR | 0.88 | 0.01 | 98 | 0.07 | -0.13 (-28, 28) |
| 2 | 0.88 | 0.02 | 98 | 0.01 | 18.46 (-18, 62) |
| 4 | 0.88 | 0.03 | 99 | 0.01 | 20.91 (-12, 68) |
| 5 | 0.89 | 0.04 | 99 | 0.00 | 21.48 (-10, 68) |
| 10 | 0.87 | 0.05 | 99 | 0.00 | 20.9 (-8, 68) |
| 20 | 0.87 | 0.06 | 98 | 0.00 | 19.36 (-8, 68) |
| 100 | 0.35 | -0.07 | 97 | 0.00 | 16.5 (-8, 66) |
| trad RAR | 0.95 | 0.08 | 93 | 0.00 | 15.4 (-9, 64) |

Table: RAR using Bayesian approach and with 0.25 drift. $E(N)$ represents the mean sample size. $\pi_{20} = P(N_A - N_B > 20)$. $N_A$ and $N_B$ denotes the number of patients assigned to treatment A and B. The mean (2.5%, 97.5%) of $N_B - N_A$ is reported.

# Discussion

- Blocks with a large number of patients also help reduce the large probability of imbalance in the wrong direction, where more patients are assigned to the inferior treatment compared to traditional RAR.
- Blocks with fewer patients can reduce the power of the trial because if patients are not randomized to both treatments in a block, the block becomes uninformative.
- Small number of blocks (K = 2, 4 and 5) has a good tradeoff between efficiency and ethically treating patients to the best known superior treatment.

# Discussion

- Large number of blocks should be clearly avoided for both ethical reason and poor design.
- Statisticians need to be careful with issues of time-trend and methods to alter the randomization ratio.
- Time-trend can significantly impact the type-I error rate.
- R package **blockRAR**

# Acknowledgement


- Rick Chappell (UW-Madison)
- David DeMets (UW-Madison)
- Peter Thall (MD Anderson)

# References


- Karrison TG, Huo D and Chappell R. A group sequential, response-adaptive design for randomized clinical trials. Controlled Clinical Trials 2003; 24(5): 506–522.

- Thall P, Fox P and Wathen J. Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. Annals of Oncology 2015; 26(8): 1621–1628.

- Bashir Q, Munsell MF, Giralt S et al. Randomized phase ii trial comparing two dose levels of thymoglobulin in patients undergoing unrelated donor hematopoietic cell transplant. Leukemia & lymphoma 2012; 53(5): 915–919.

# THANK YOU!