Adaptive Designs in Clinical Trials

Thevaa Chandereng

Biostatistics & Medical Informatics
University of Wisconsin–Madison

thevaachandereng.github.io
github.com/thevaachandereng

Overview

- Adaptive Process
- Introduction to RAR
- BATTLE-1 Trial
- Other Issues with RAR
- Blocked RAR
- Incorportion of Historical Data
- bayesCT
- Future Work

Adaptive Designs in Clinical Trials

FDA

Adaptive Design Clinical Trials for Drugs and Biologics
Guidance for Industry, November 2019

Adaptive Designs in Clinical Trials

FDA

Adaptive Design Clinical Trials for Drugs and Biologics
Guidance for Industry, November 2019

"An adaptive design is defined as a clinical trial design that allows for prospectively planned modifications to one or more aspects of the design based on accumulating data from subjects in the trial. "

Adaption on-the-fly

- Randomize more subjects to better treatment.

Adaption on-the-fly

- Randomize more subjects to better treatment.
- Can drop unpromising treatment arms.

Adaption on-the-fly

- Randomize more subjects to better treatment.
- Can drop unpromising treatment arms.
- Alter inclusion criteria.

Adaption on-the-fly

- Randomize more subjects to better treatment.
- Can drop unpromising treatment arms.
- Alter inclusion criteria.
- Adding an arm to an ongoing trial

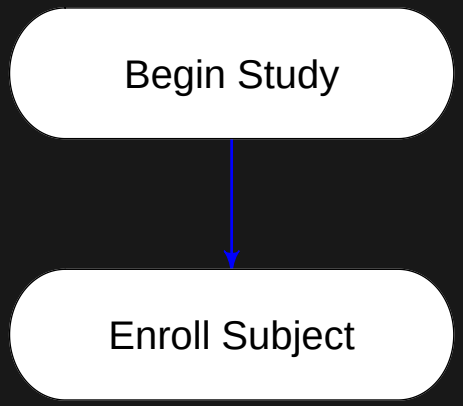
Adaption on-the-fly

- Randomize more subjects to better treatment.
- Can drop unpromising treatment arms.
- Alter inclusion criteria.
- Adding an arm to an ongoing trial
- Stop trial early for futility or success!

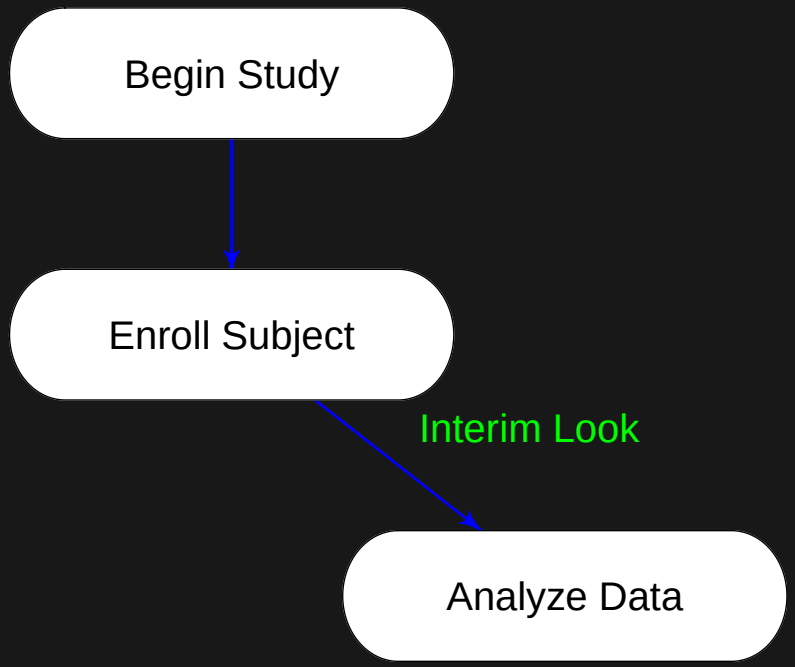
Adaption on-the-fly

- Randomize more subjects to better treatment.
- Can drop unpromising treatment arms.
- Alter inclusion criteria.
- Adding an arm to an ongoing trial
- Stop trial early for futility or success!
- Dynamically incorporate historical data.

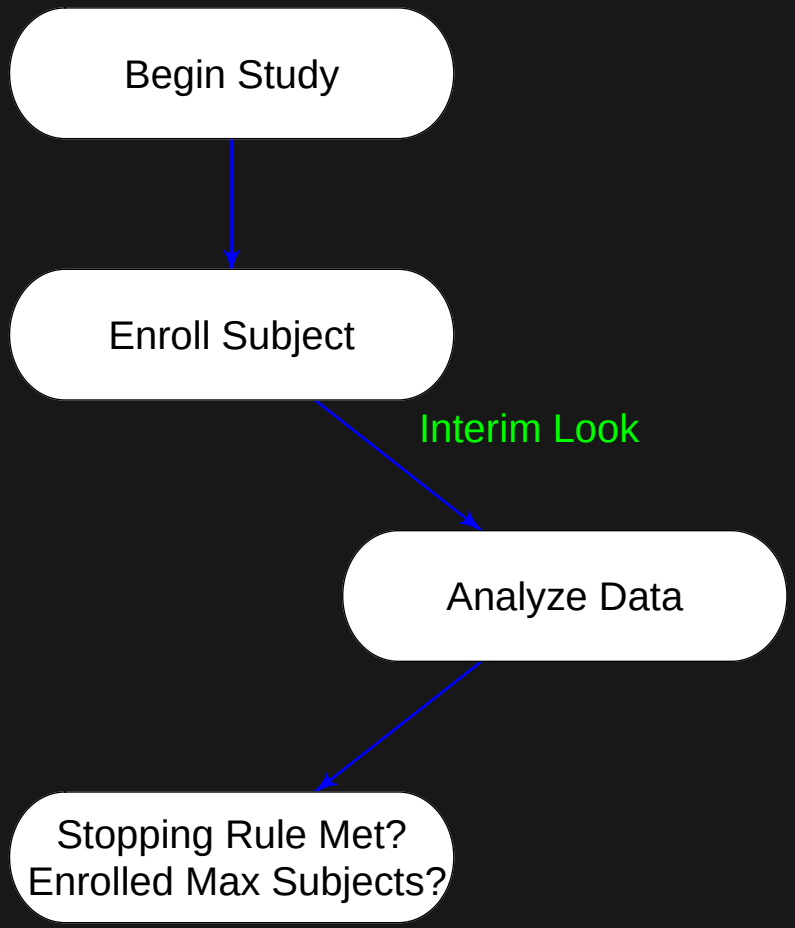
Adaptive Process



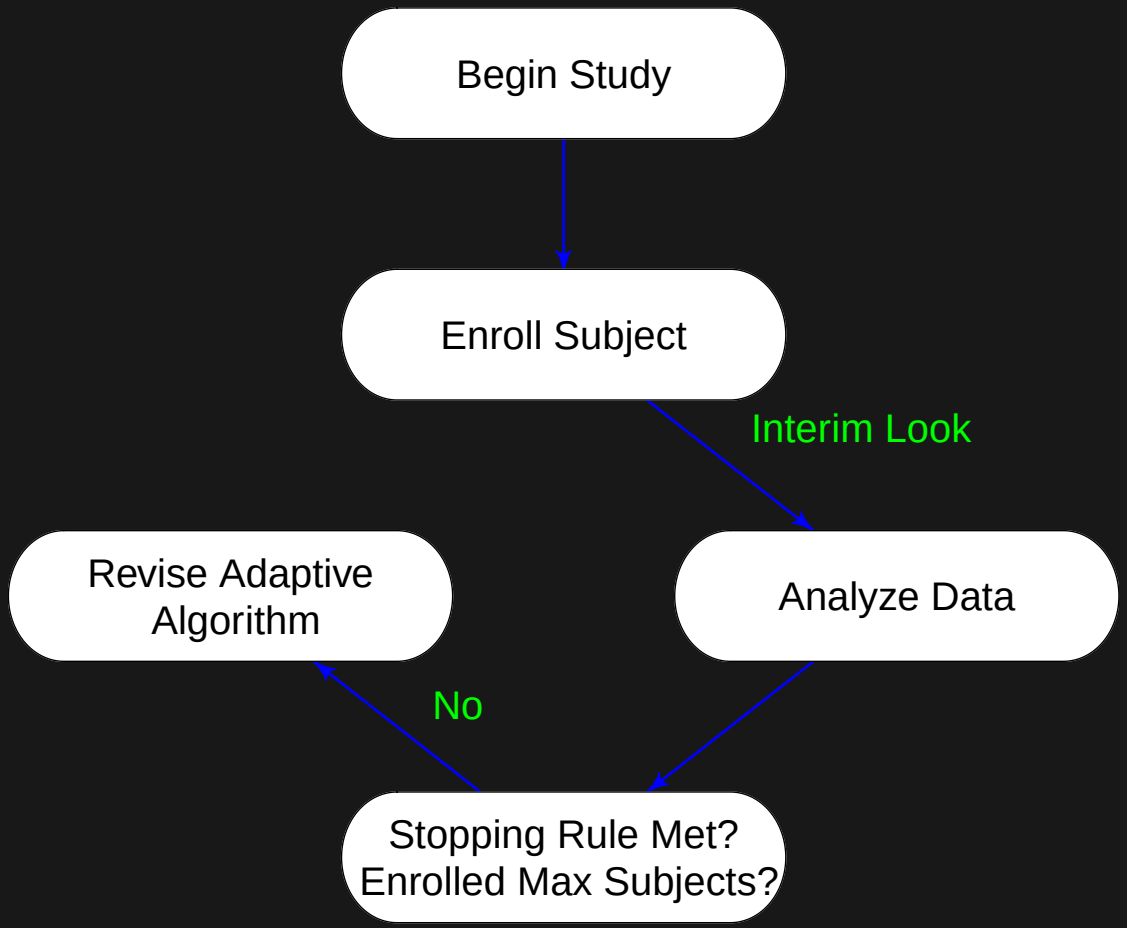
Adaptive Process



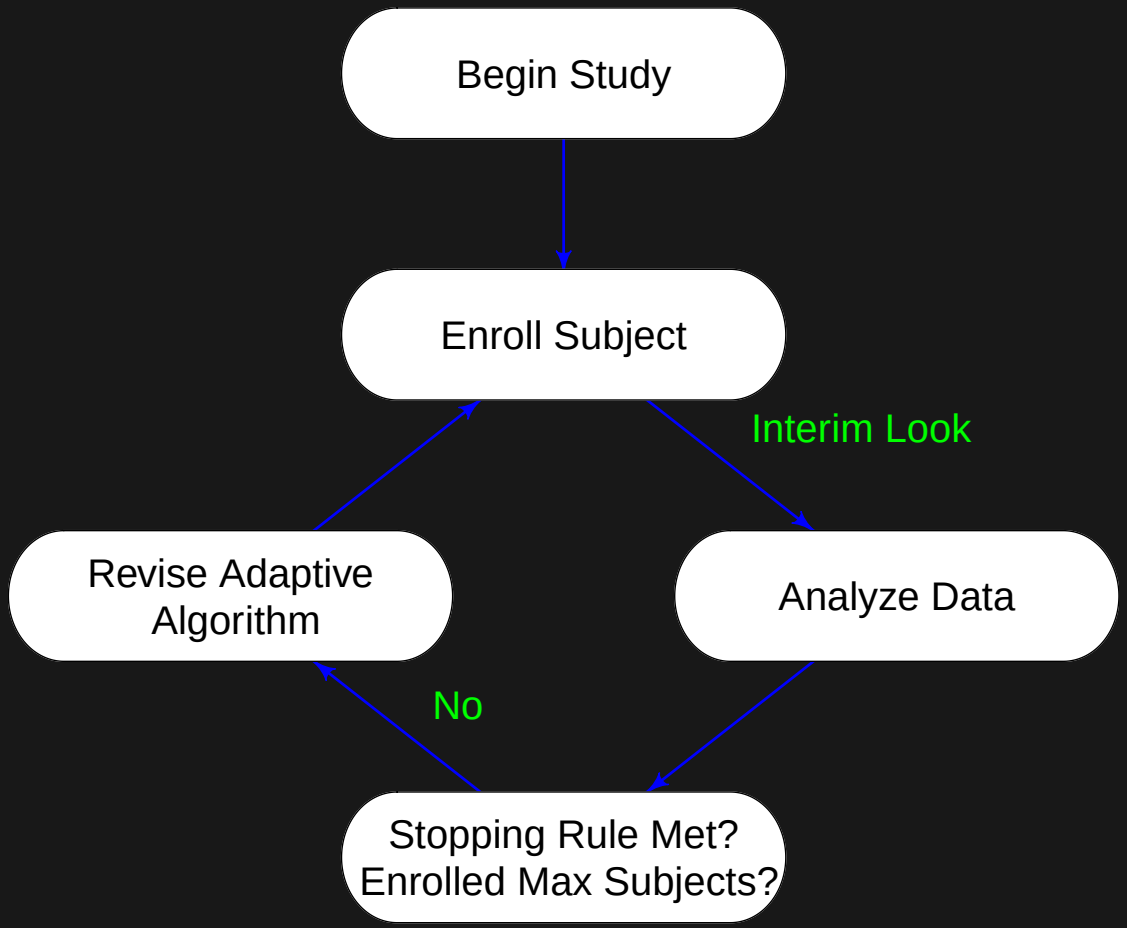
Adaptive Process



Adaptive Process



Adaptive Process



Adaptive Process

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

Objective of Trials

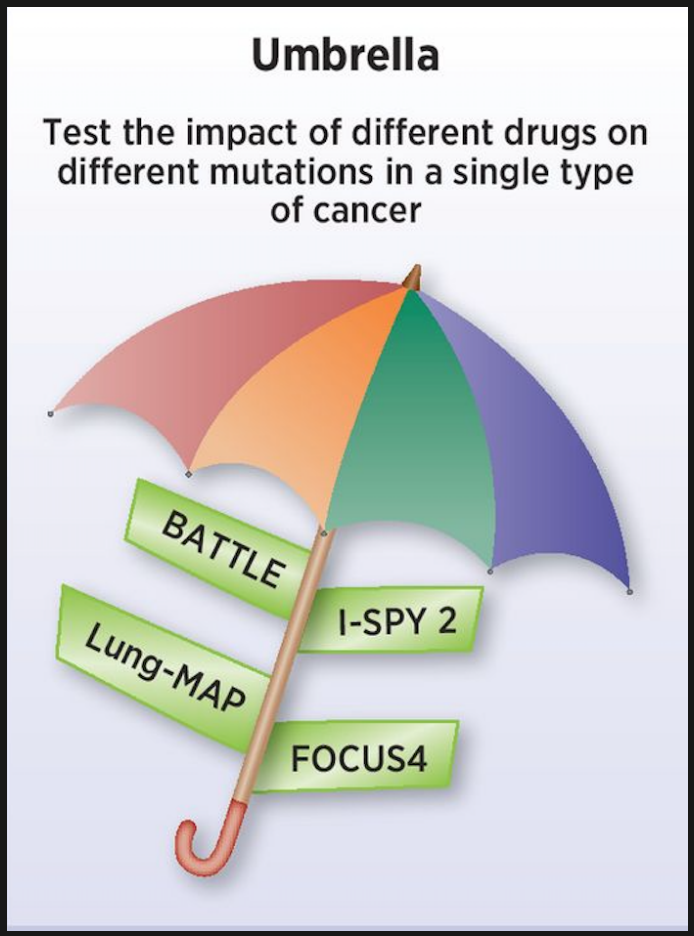
- maximize power to detect clinically relevant difference;
- minimize total cost of the trial;
- minimize the number of failures;
- control type I error under nominal level;
- maximize the individual's experience in the trial

Objective of Trials

- maximize power to detect clinically relevant difference;
- minimize total cost of the trial;
- minimize the number of failures;
- control type I error under nominal level;
- maximize the individual's experience in the trial
- maximize the probability to provide the best treatment

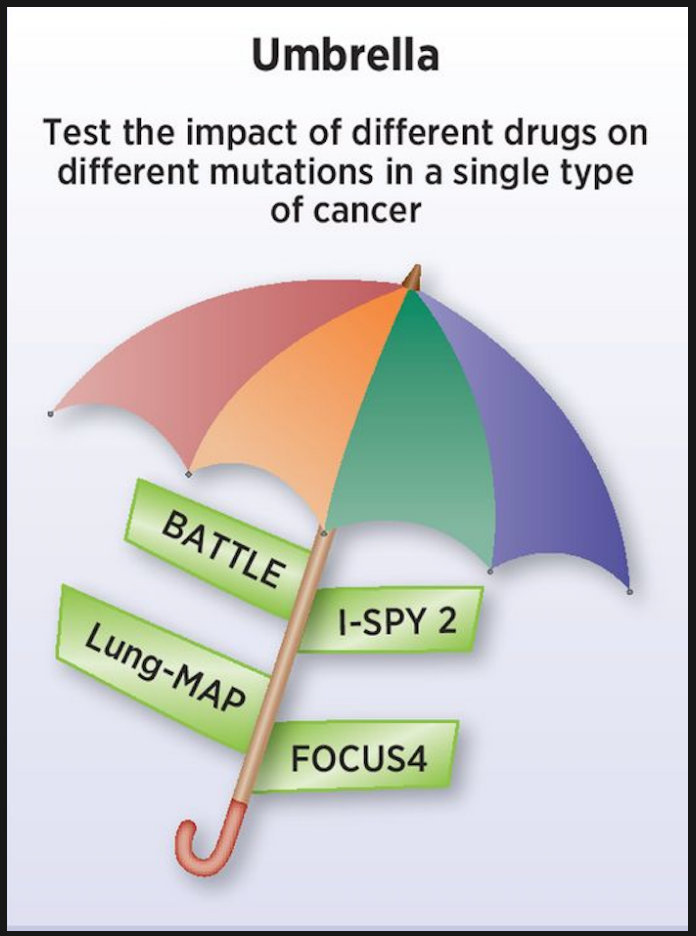
BATTLE-1 Trial

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



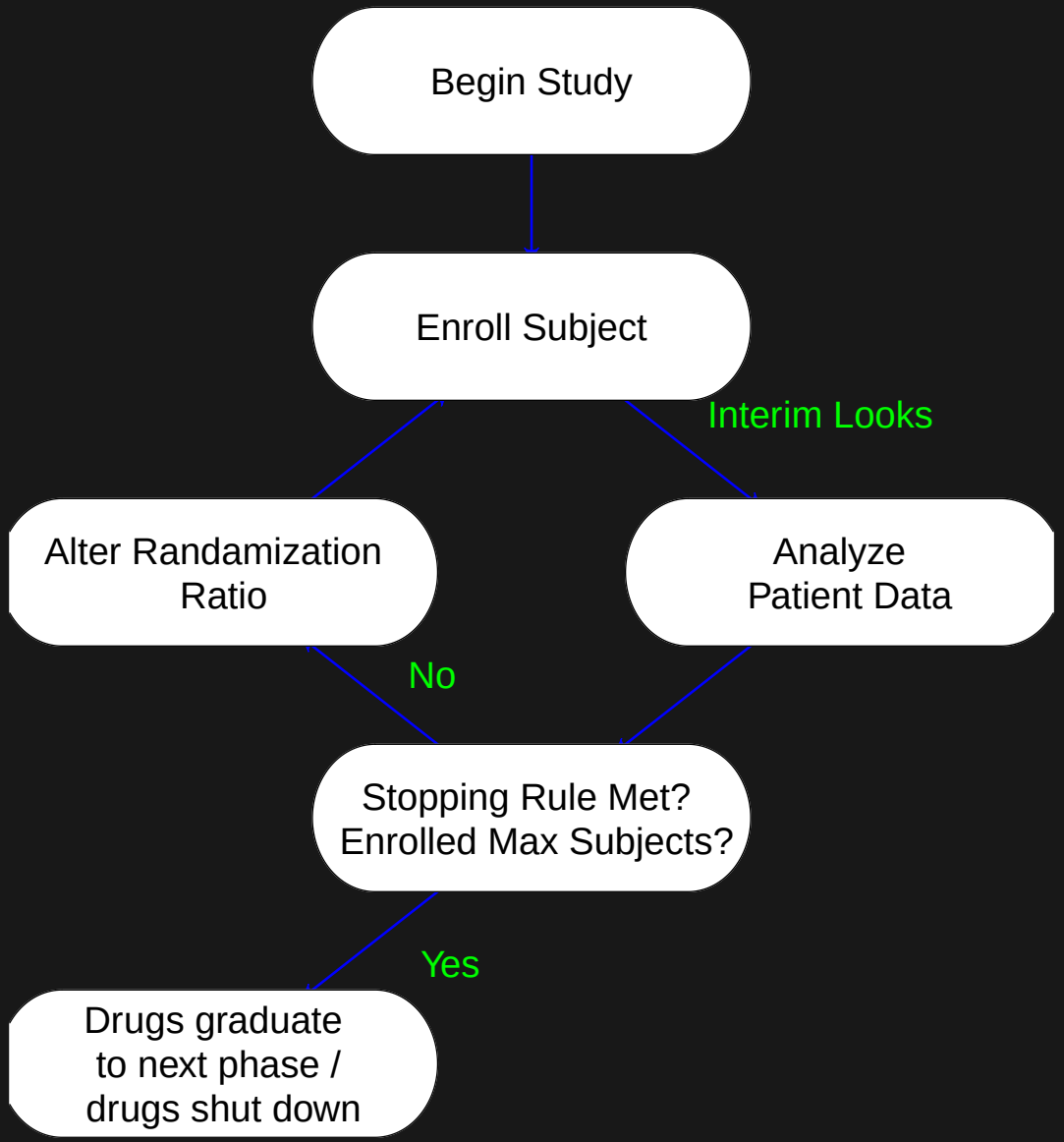
BATTLE-1 Trial

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

Design of BATTLE-1 Trial



- 1st interim look: at least one patient in each treatment and biomarker groups
- All subsequent interim looks: patient by patient

BATTLE-1 Trial

BATTLE-1 Trial

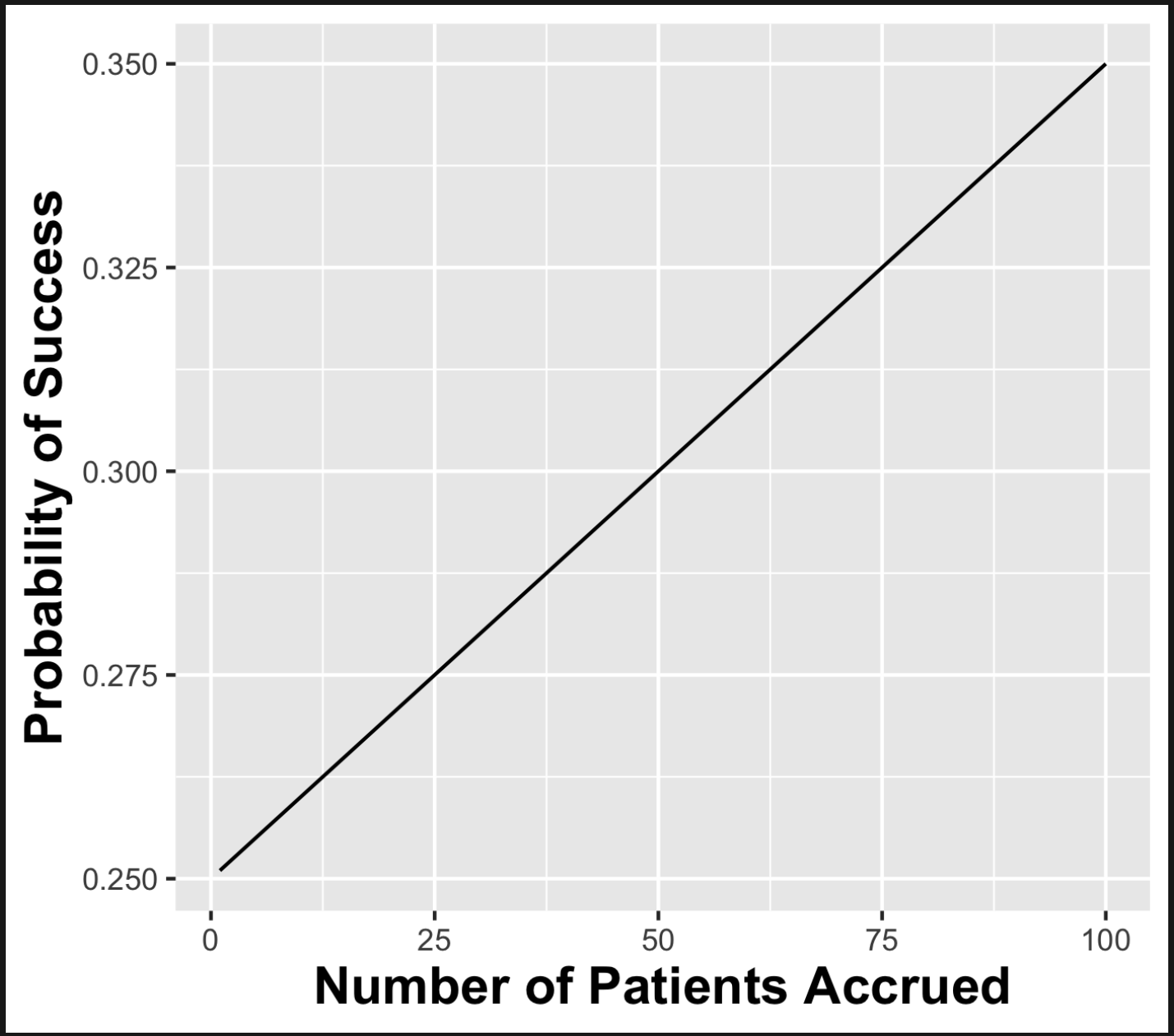
“More smokers enrolled in the latter part of the study compared to the beginning of the study”.

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.

Time-trends in Clinical Trials



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.

- Chappell (2011) used an artificial example to illustrate changes in proportion of success over time in a binomial trial draws a wrong conclusion.

Issues with RAR

Common problems with RAR methods that are proposed (besides time-trends):

- The type-I error rate is usually not controlled at the nominal level under some traditional Bayesian RAR design.

Issues with RAR

Common problems with RAR methods that are proposed (besides time-trends):

- The type-I error rate is usually not controlled at the nominal level under some traditional Bayesian RAR design.
- Thall et al. (2015): "... AR may behave pathologically in that it carries a nontrivial risk of creating a large sample size imbalance in favor of the inferior treatment".

Issues with RAR

Common problems with RAR methods that are proposed (besides time-trends):

- The type-I error rate is usually not controlled at the nominal level under some traditional Bayesian RAR design.
- Thall et al. (2015): "... AR may behave pathologically in that it carries a nontrivial risk of creating a large sample size imbalance in favor of the inferior treatment".
- Poor/extreme schemes to alter the allocation ratios.

Blocked Design for RAR

Karrison et al. (2003) introduced a stratified group-sequential method for RAR.

Blocked Design for RAR

Karrison et al. (2003) introduced a stratified group-sequential method for RAR.

But never expanded it on the characteristics of blocked RAR!

Blocked Design for RAR

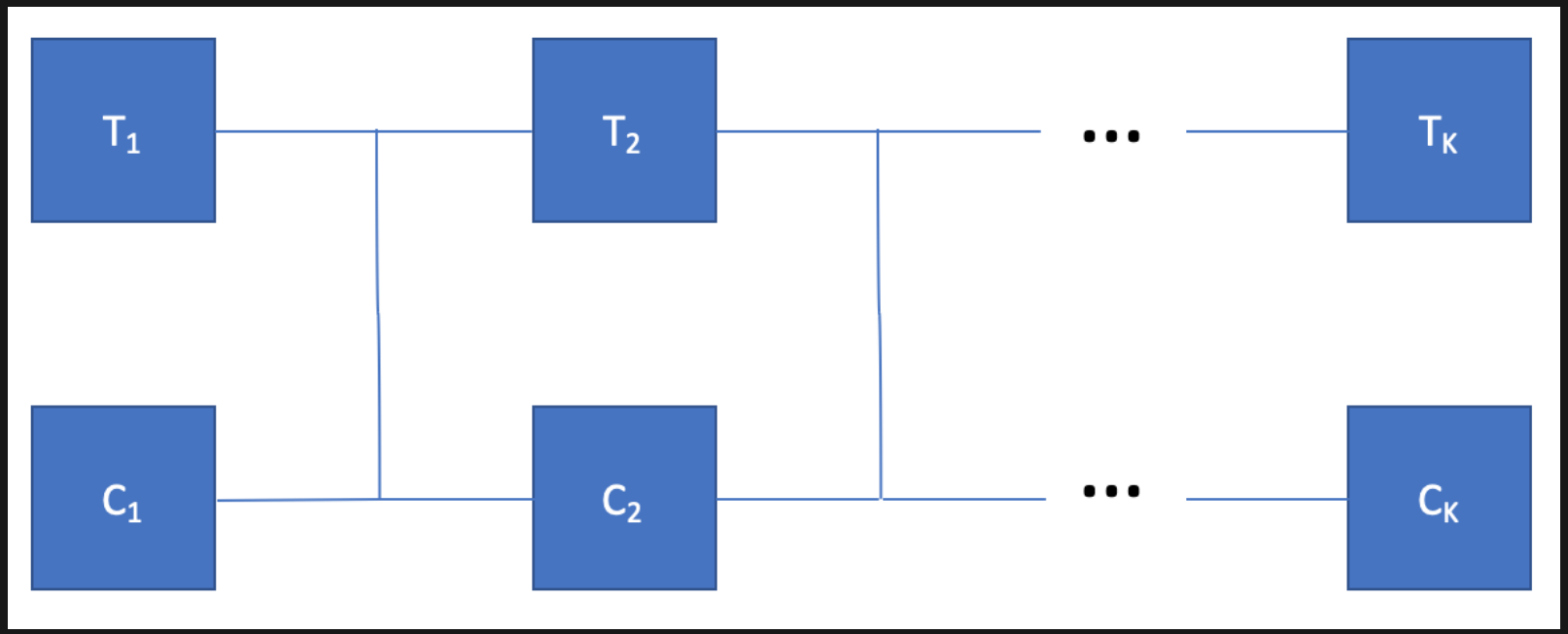
Karrison et al. (2003) introduced a stratified group-sequential method for RAR.

But never expanded it on the characteristics of blocked RAR!

What is the optimal block size?

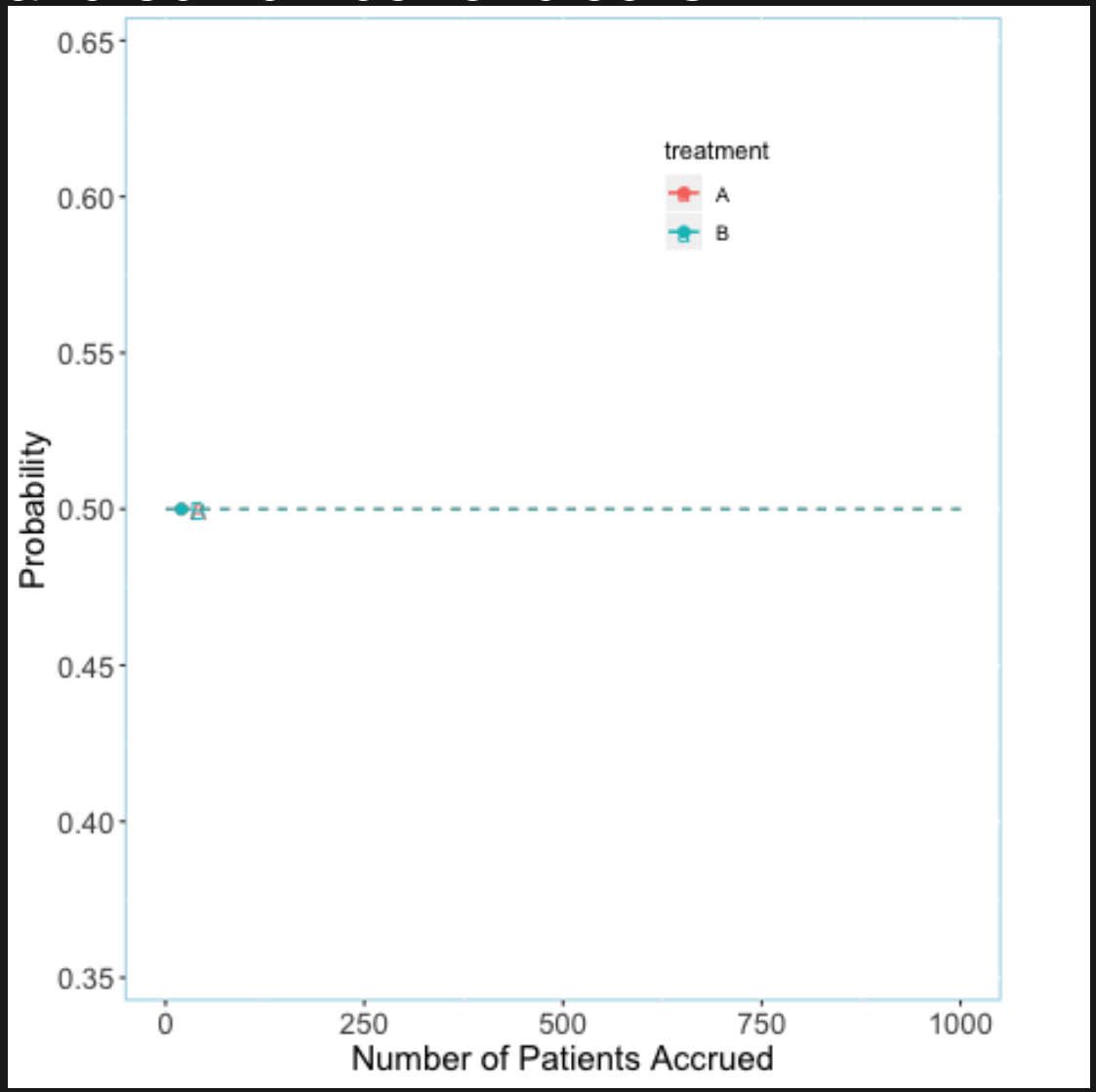
Blocked Design for RAR

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



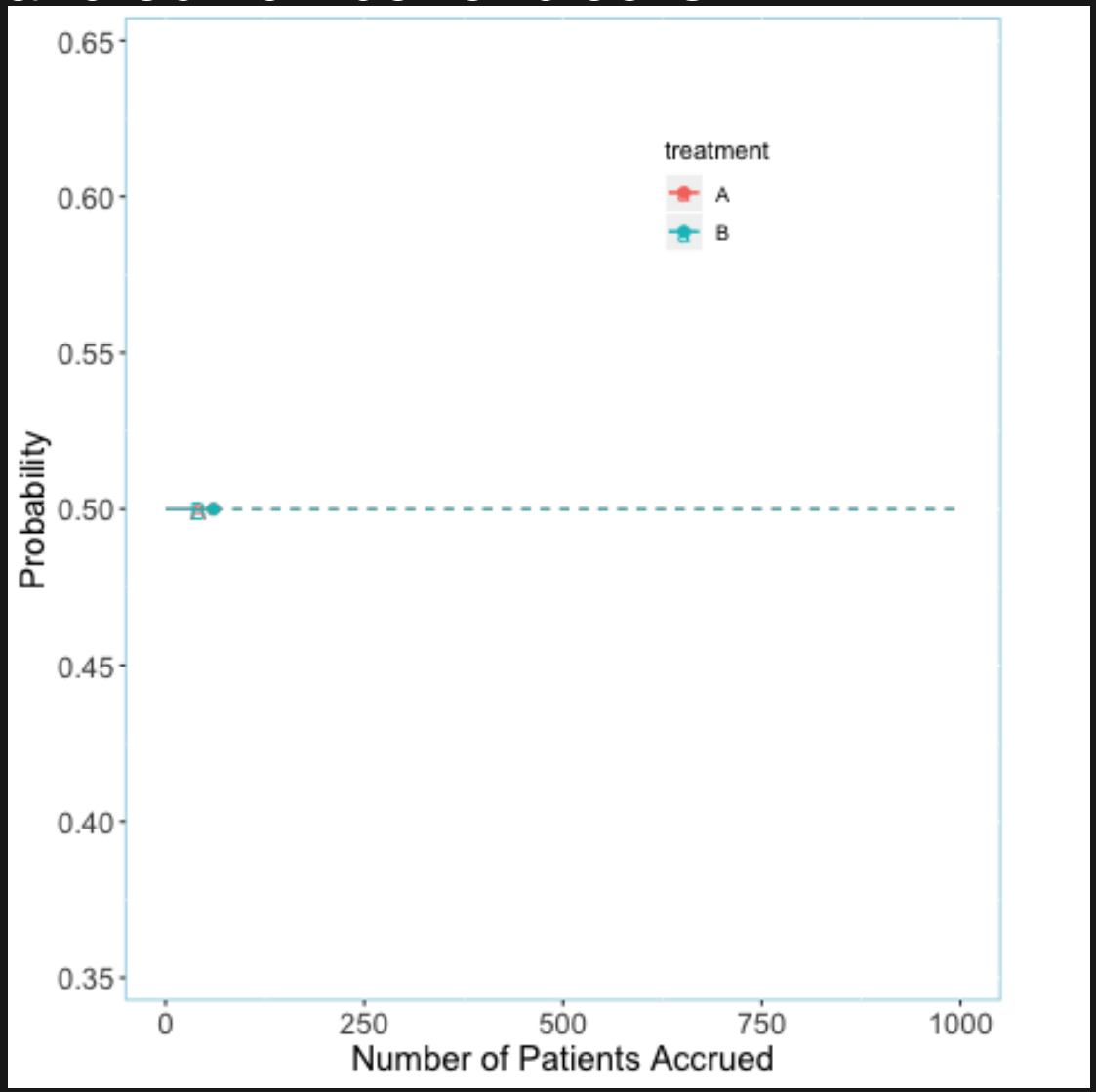
Example RAR

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



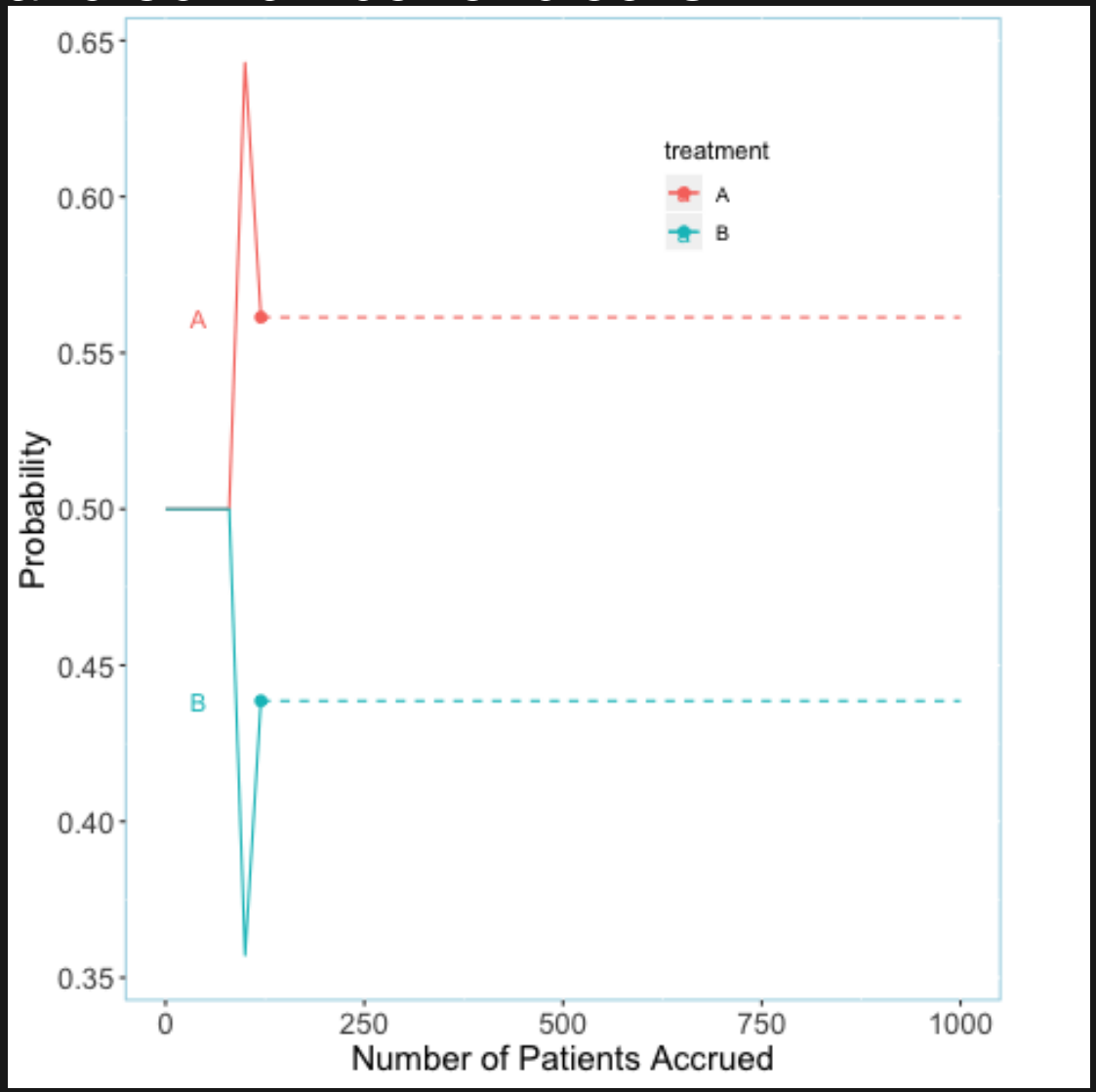
Example RAR

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



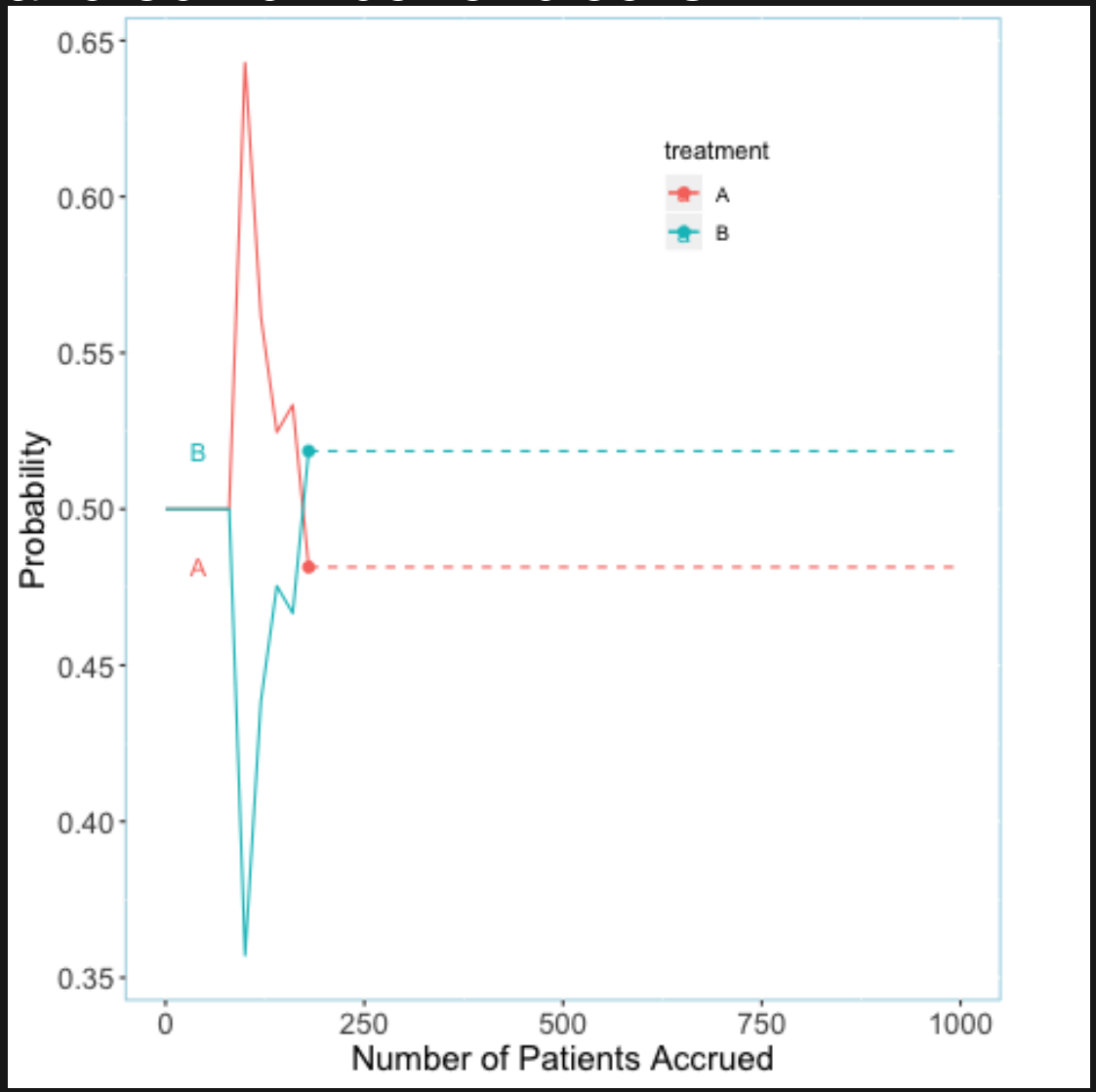
Example RAR

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



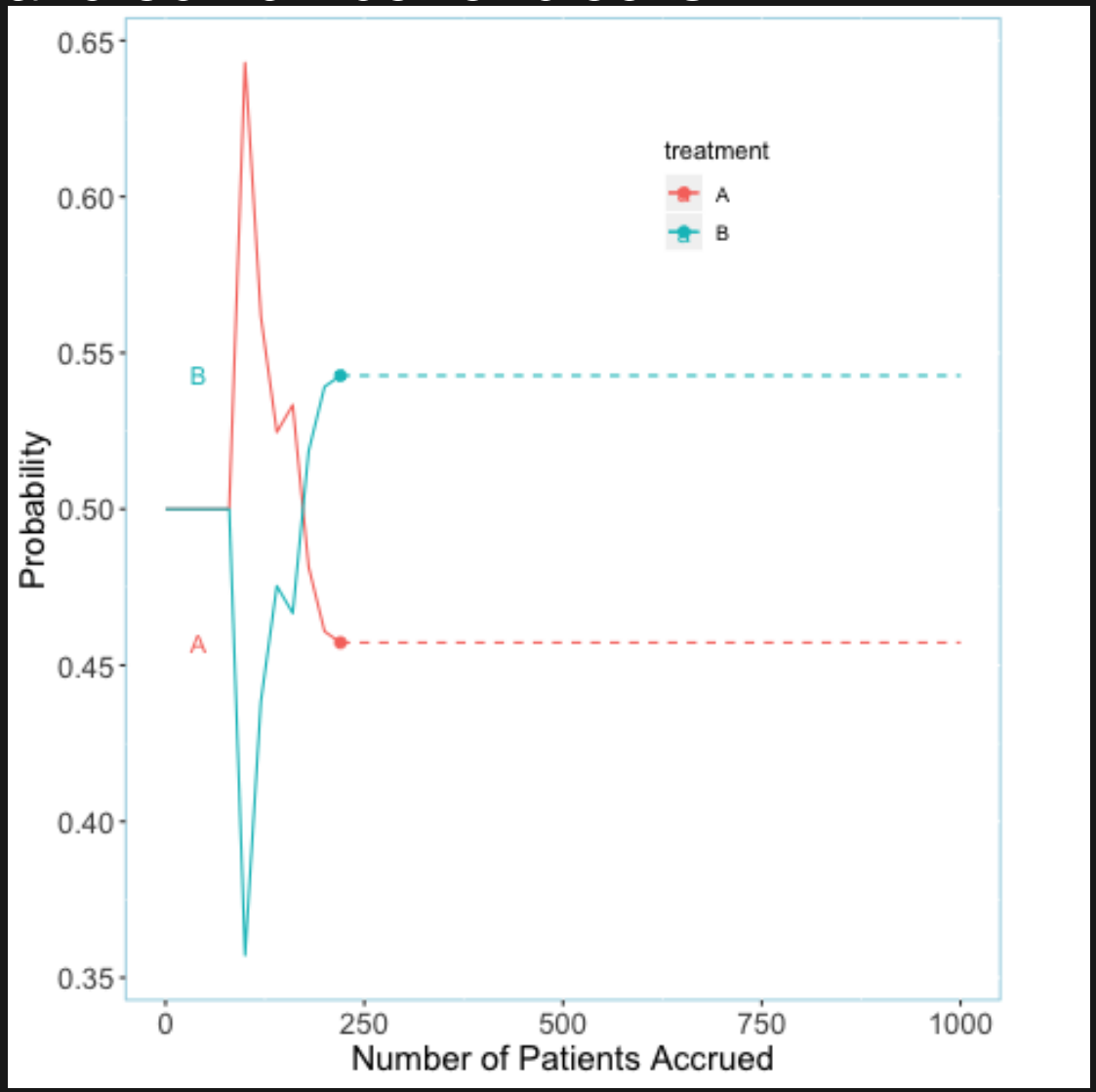
Example RAR

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



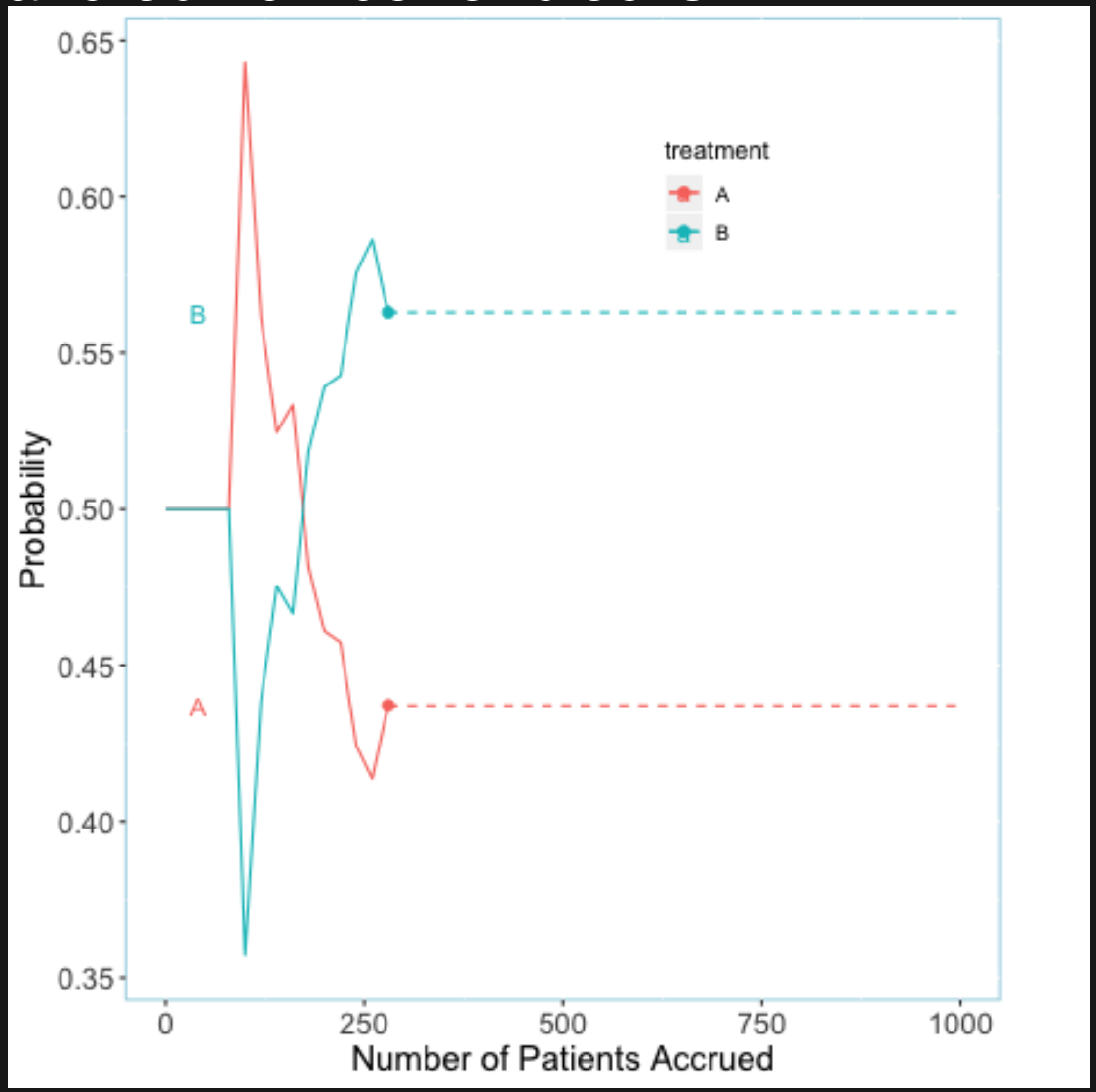
Example RAR

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



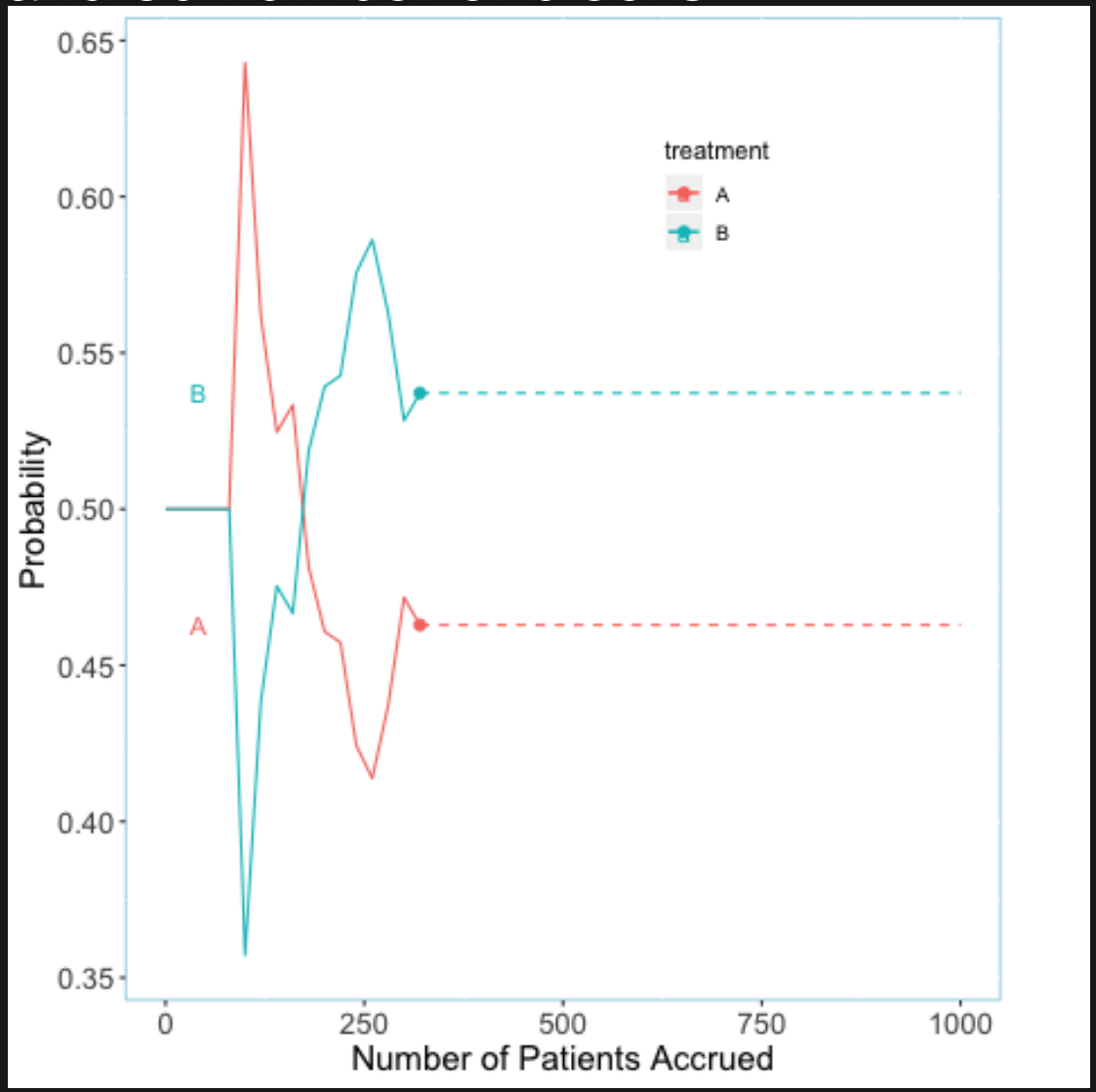
Example RAR

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



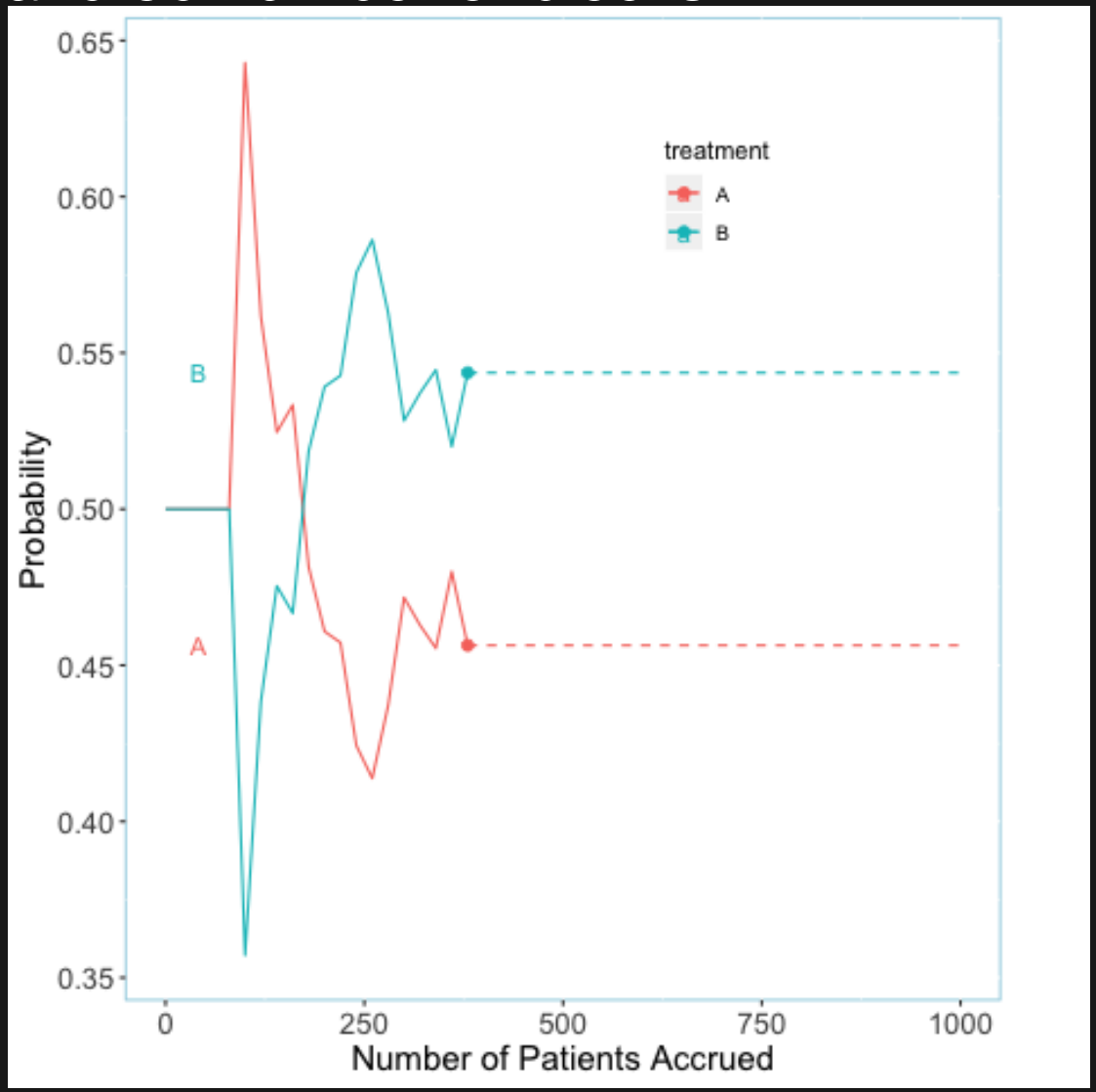
Example RAR

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



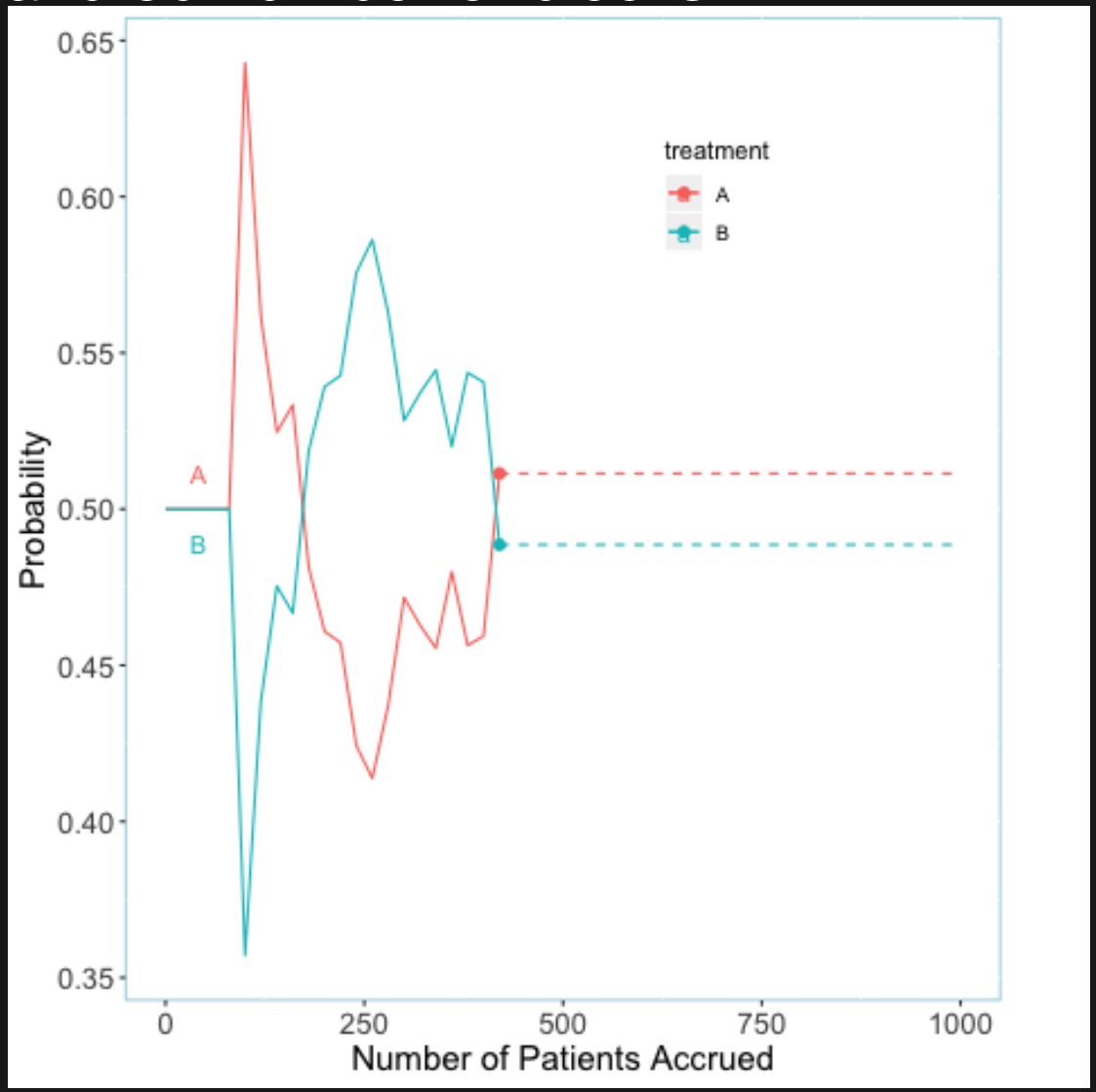
Example RAR

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



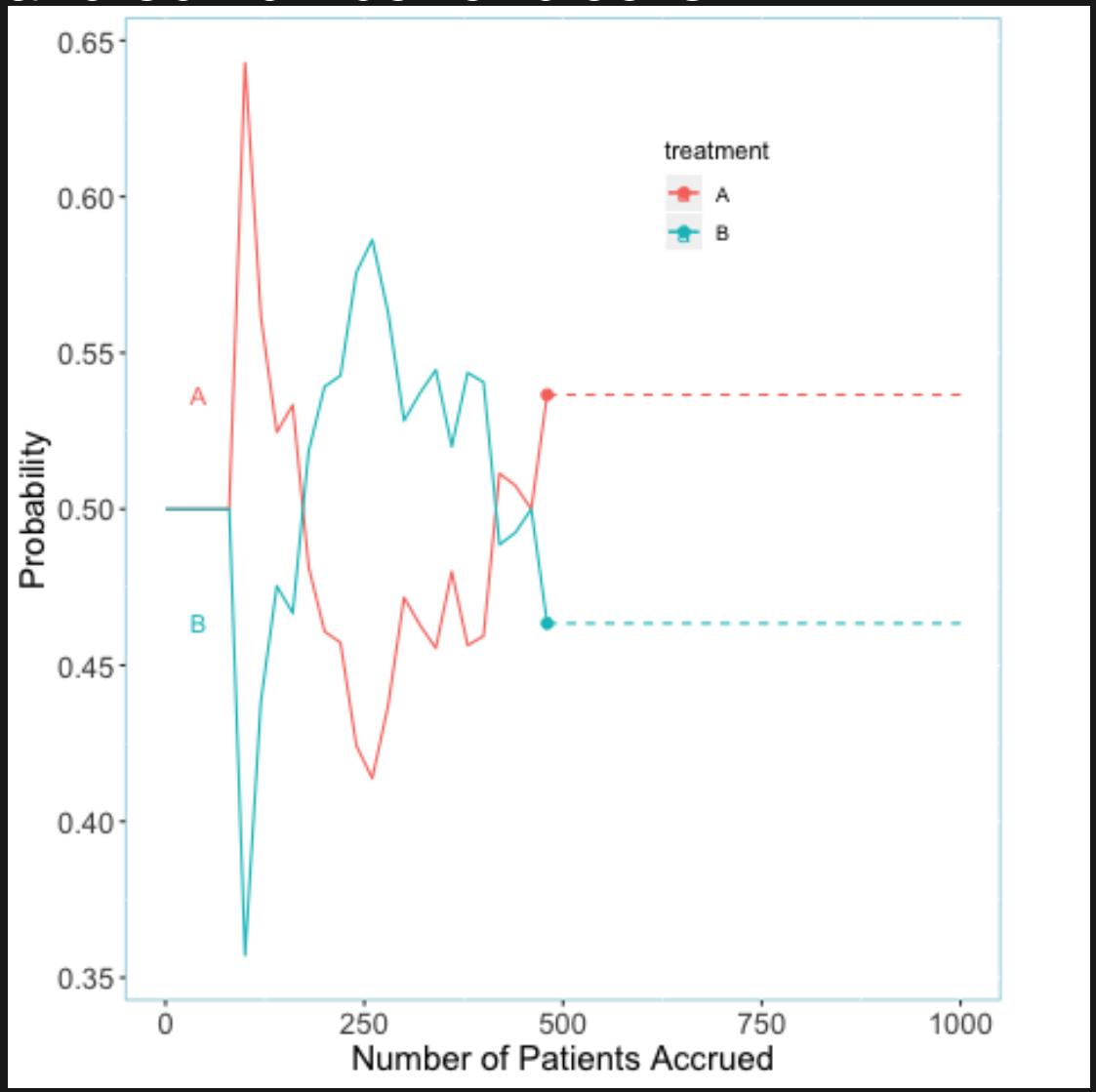
Example RAR

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



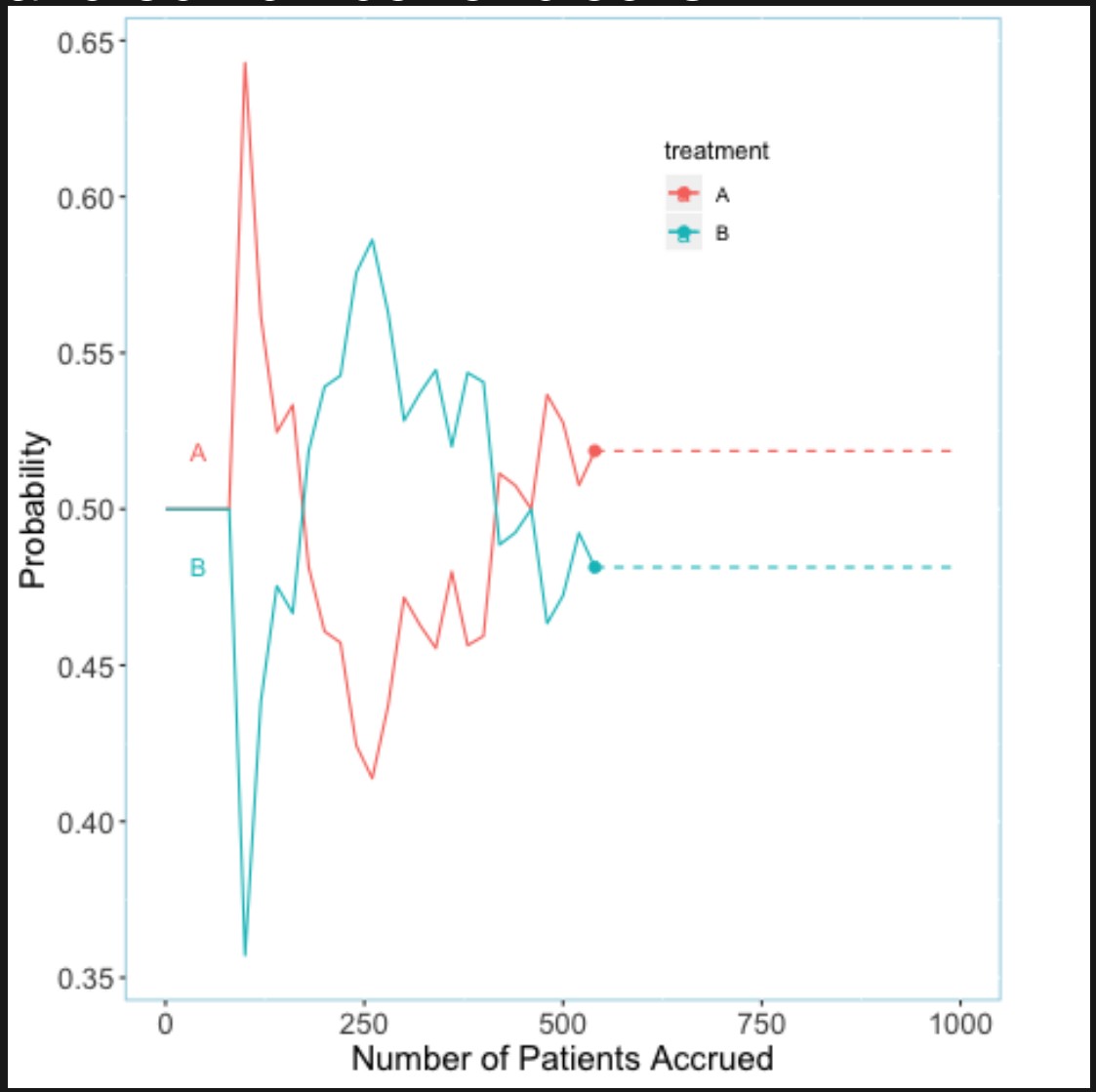
Example RAR

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



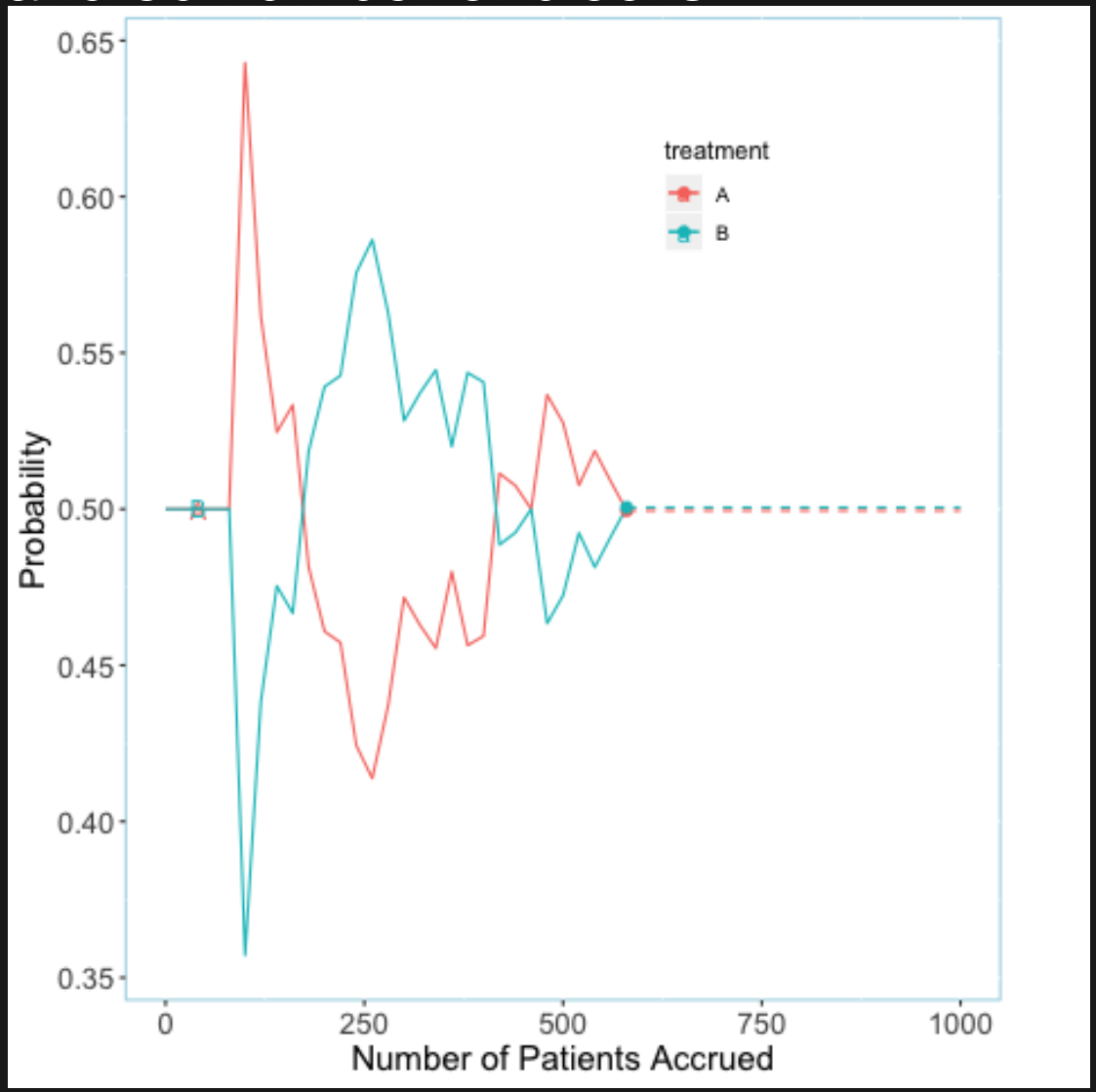
Example RAR

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



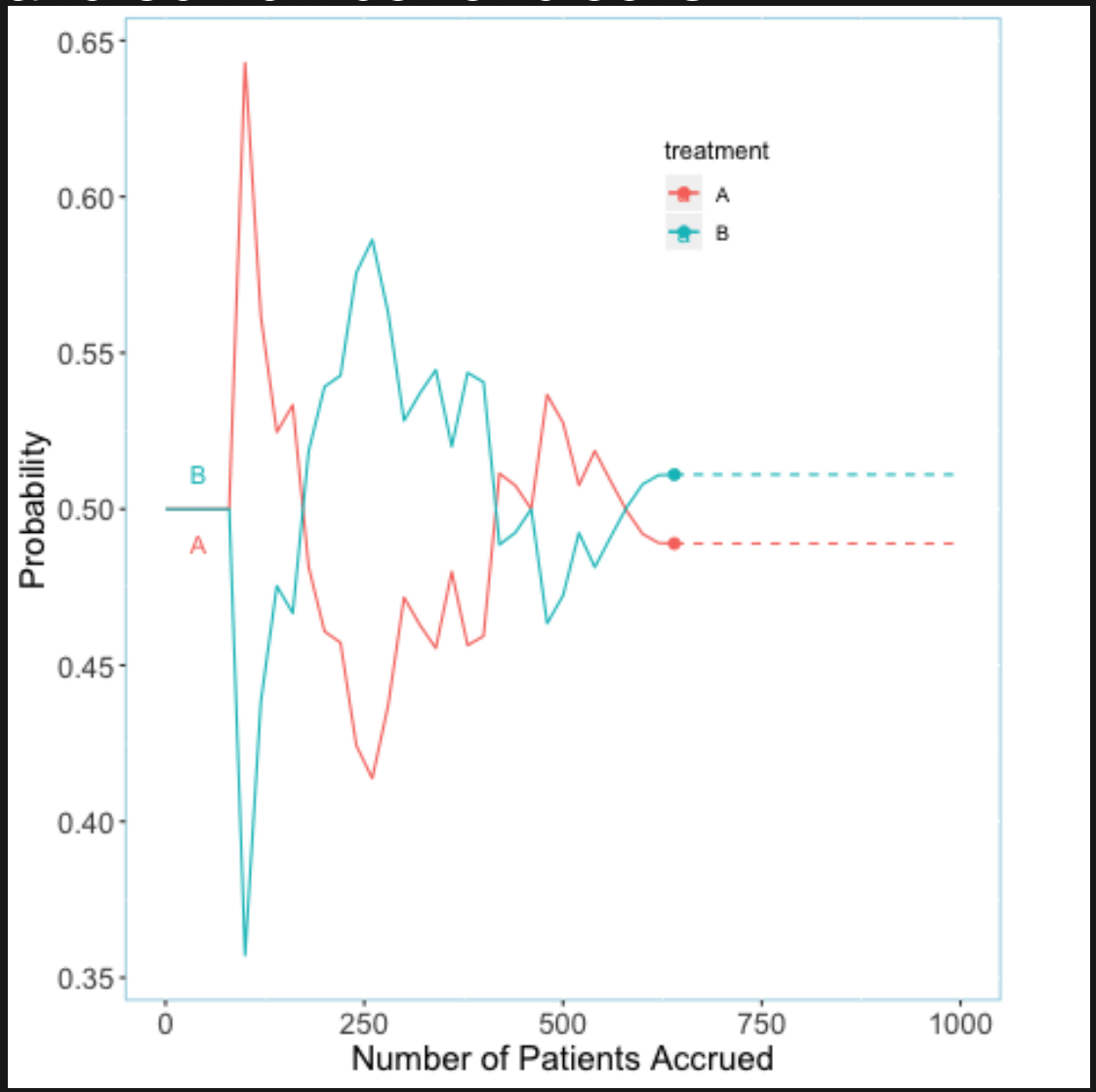
Example RAR

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



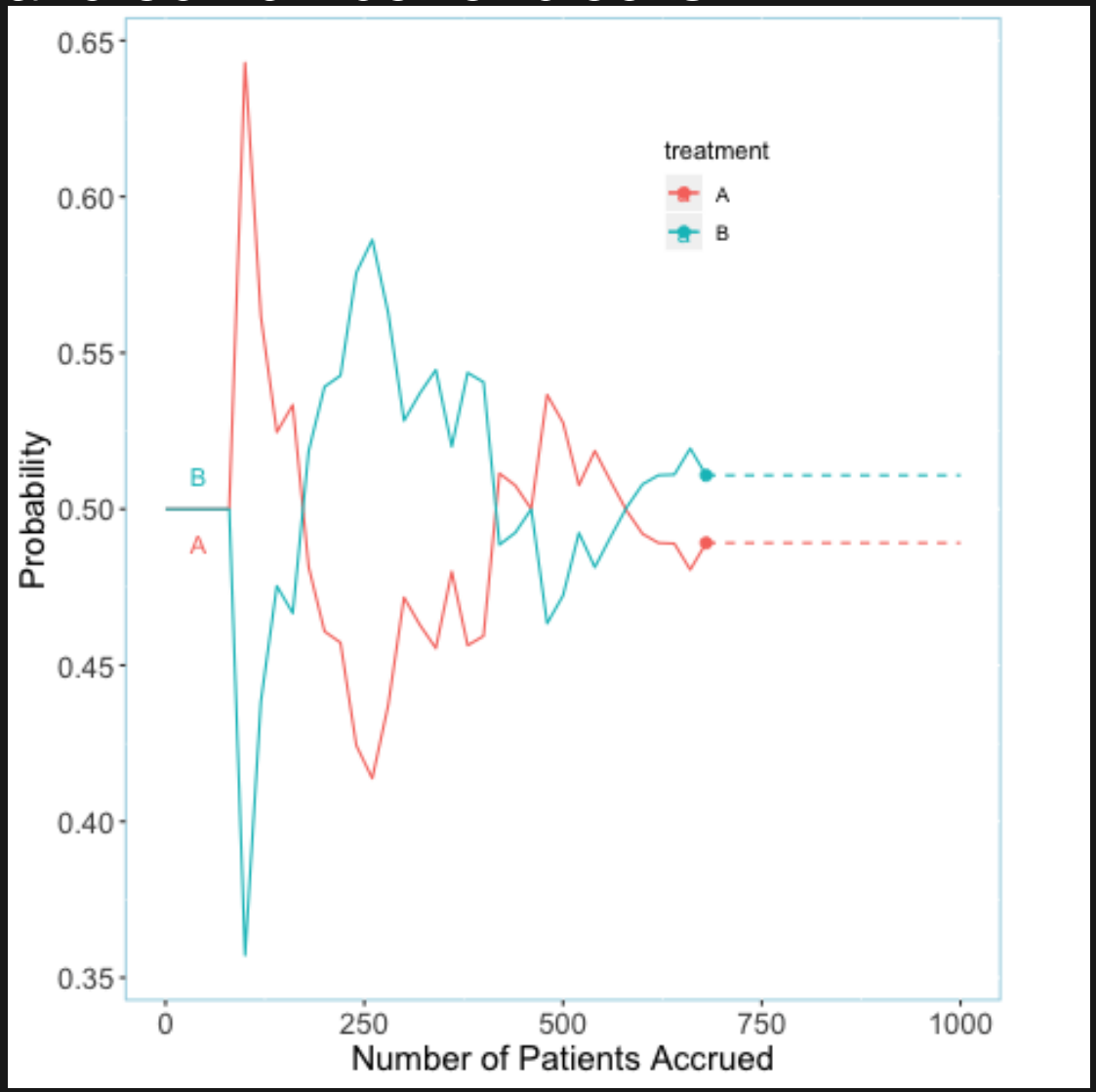
Example RAR

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



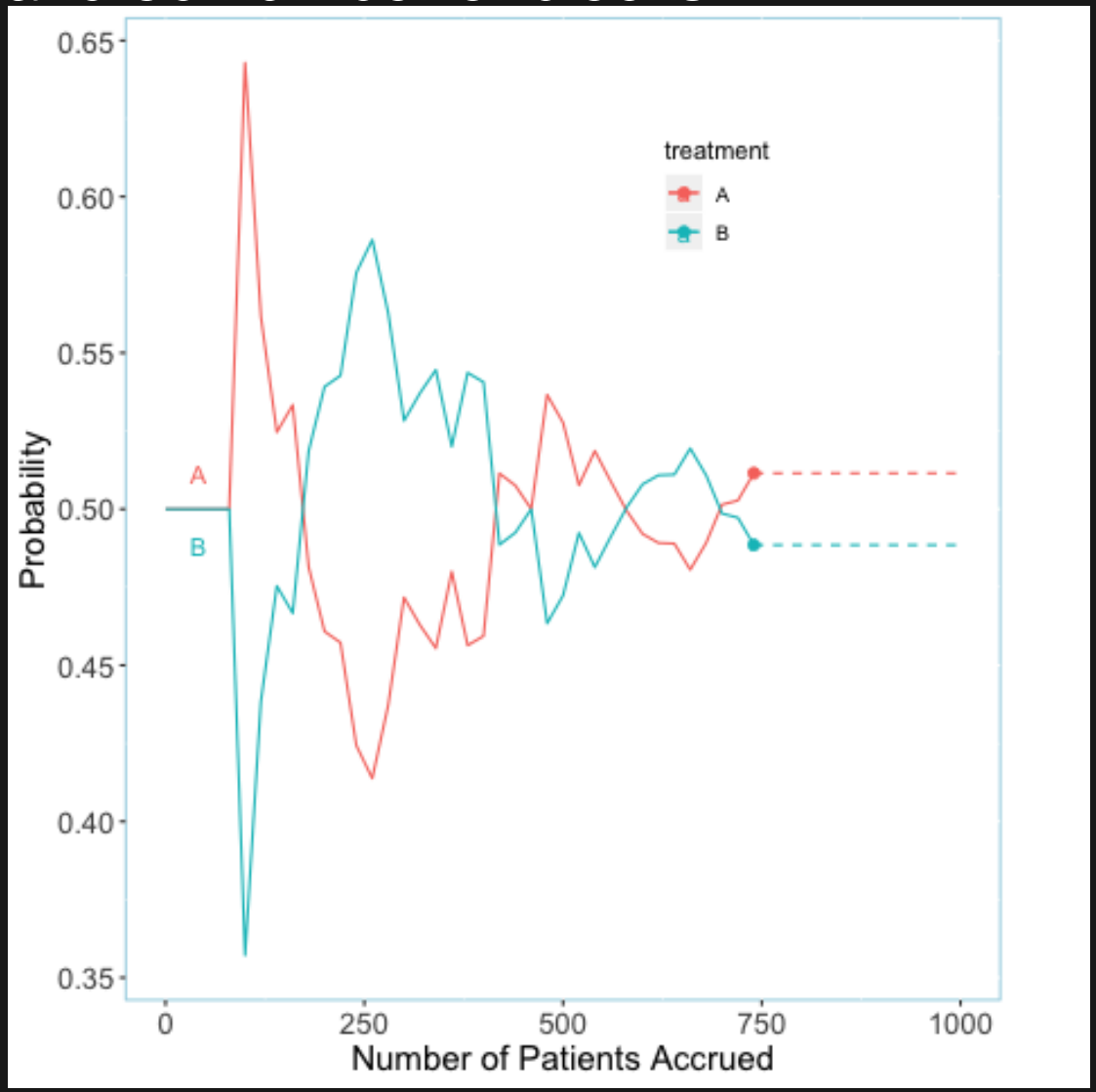
Example RAR

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



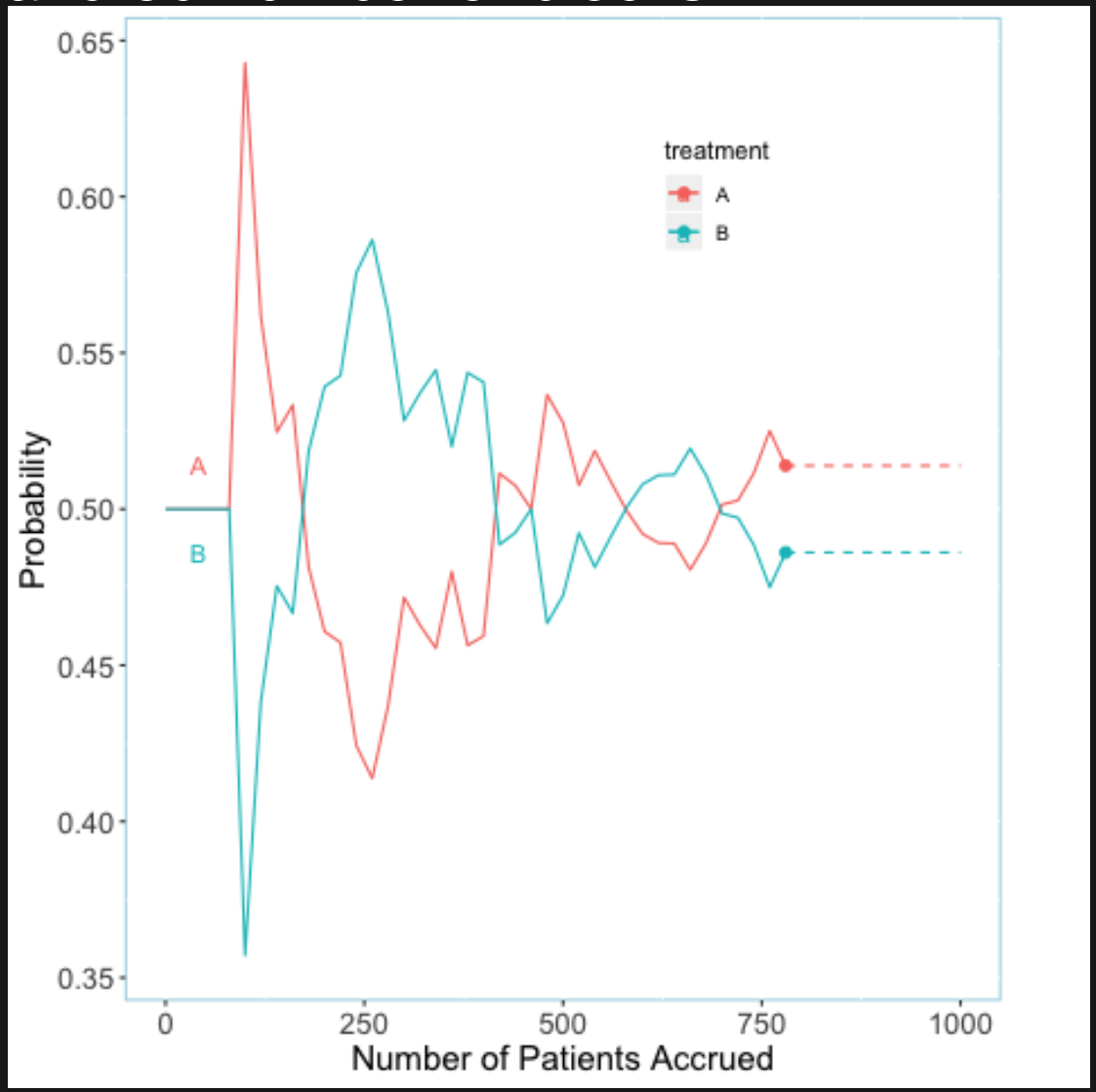
Example RAR

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



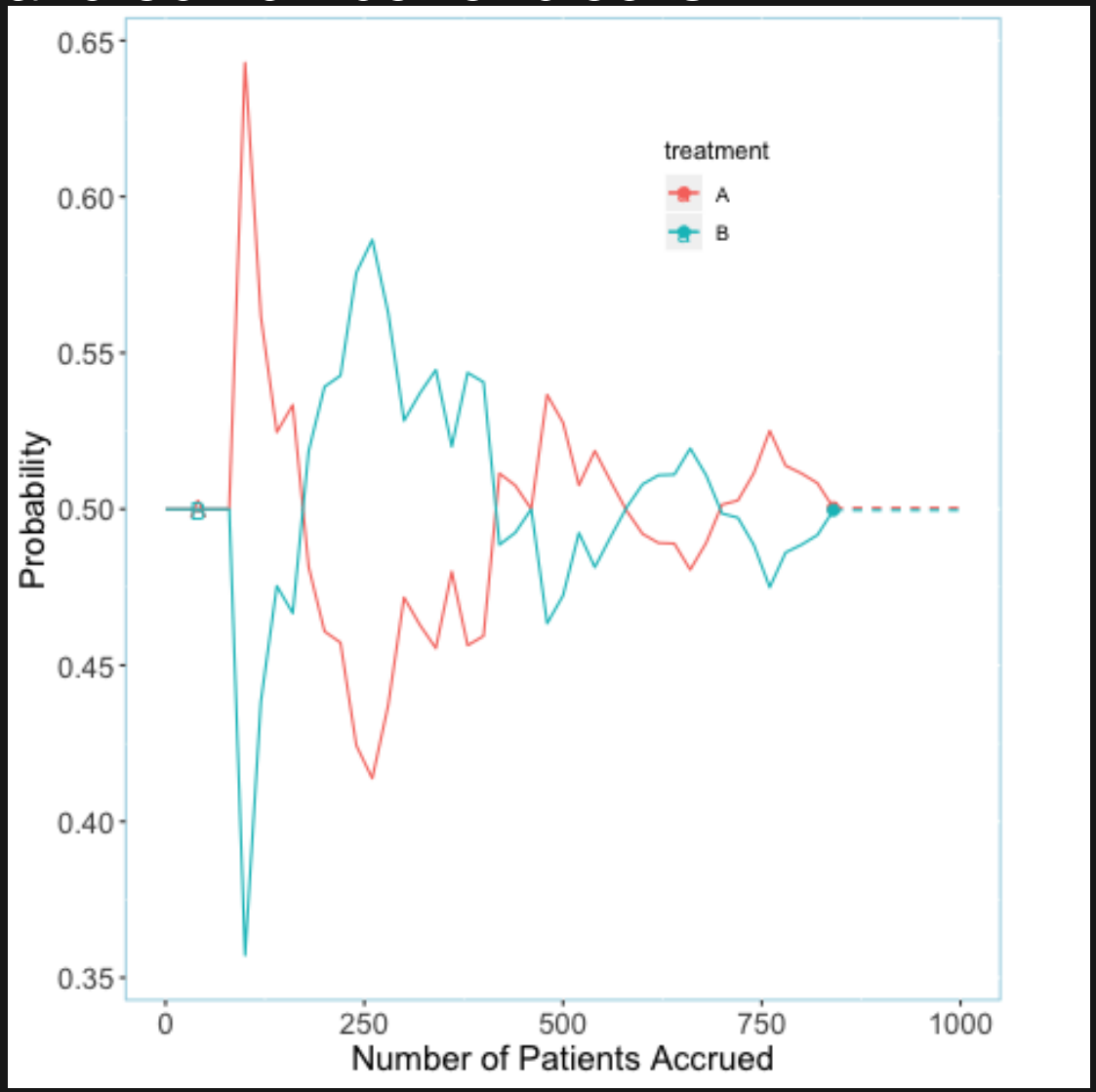
Example RAR

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



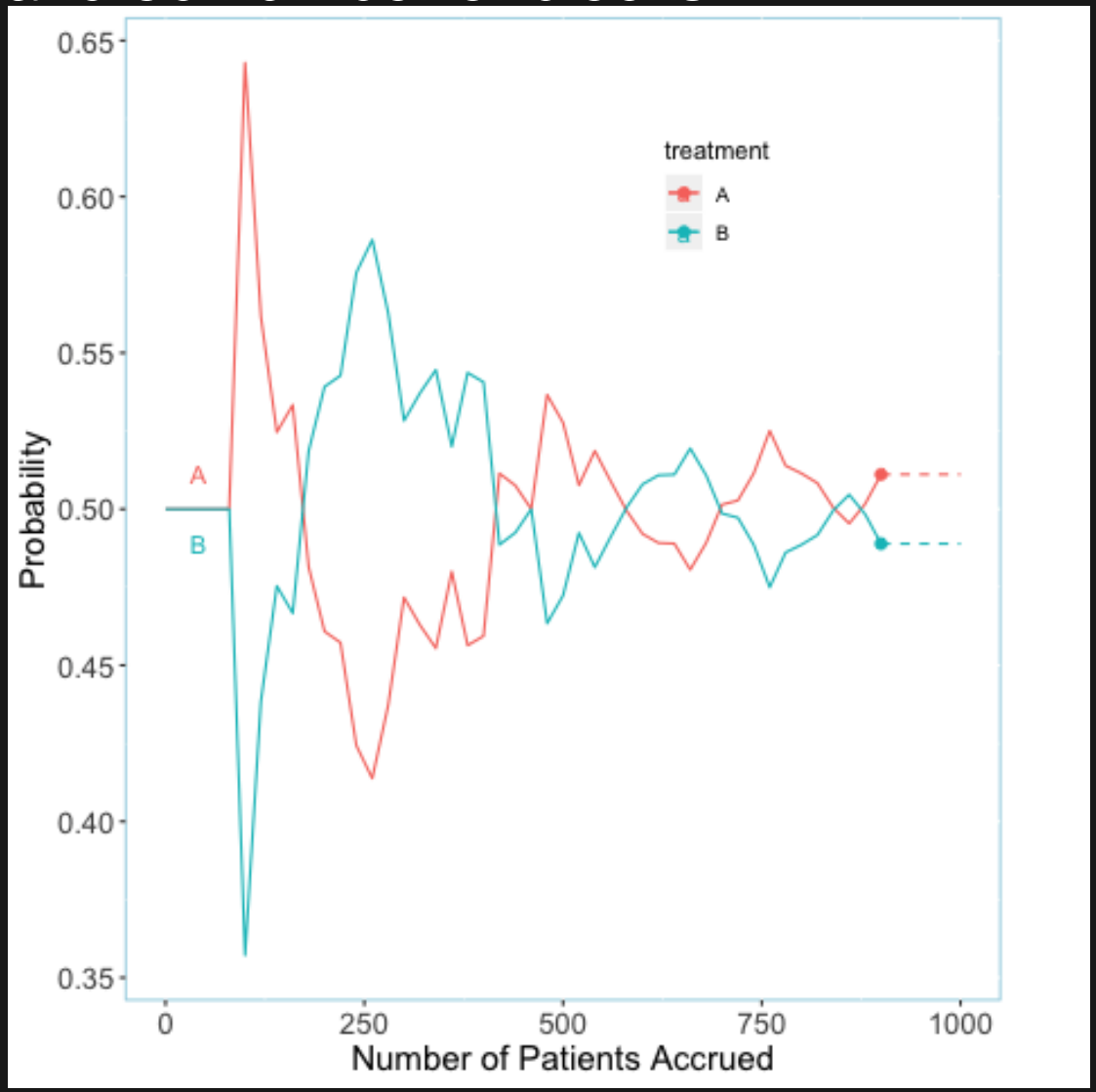
Example RAR

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



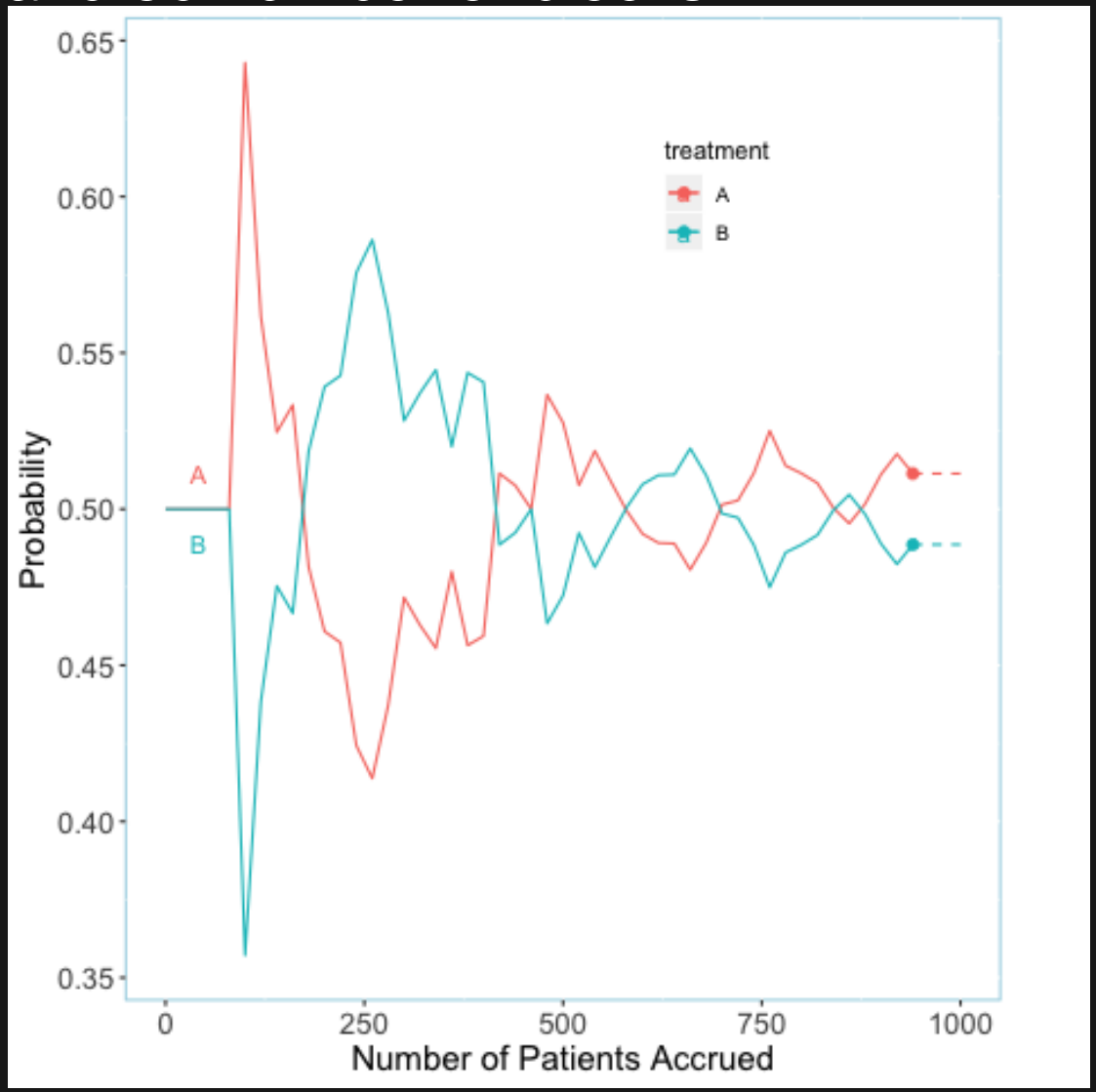
Example RAR

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



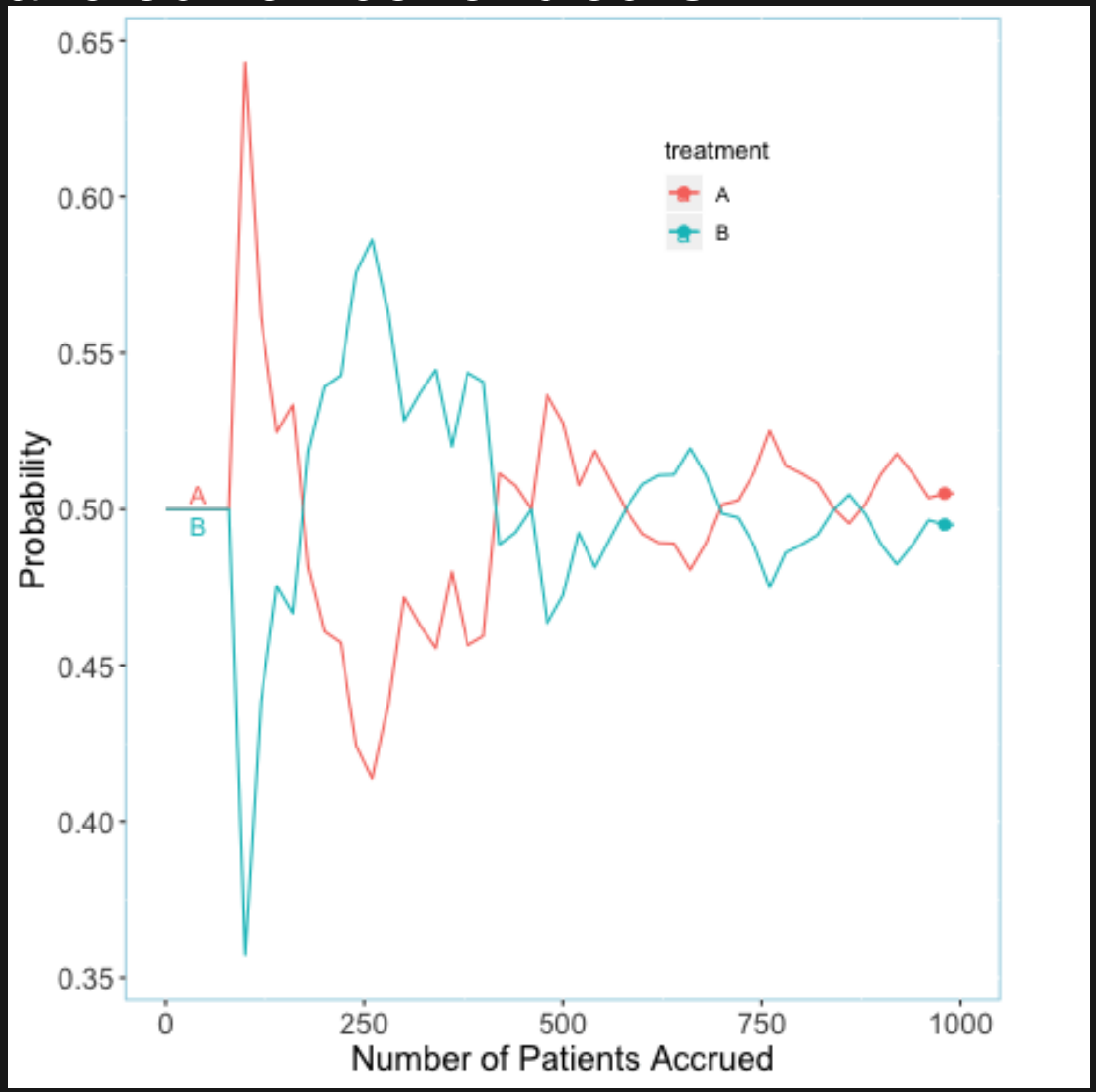
Example RAR

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



Example RAR

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



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**

How did we address the issues with RAR?

Time-trend issues in RAR.

Blocking!

How did we address the issues with RAR?

Time-trend issues in RAR.

Blocking!

Thall et al. (2015): "... AR may behave pathologically in that it carries a nontrivial risk of creating a large sample size imbalance in favor of the inferior treatment".

Run-in period.

How did we address the issues with RAR?

Time-trend issues in RAR.

Blocking!

Thall et al. (2015): "... AR may behave pathologically in that it carries a nontrivial risk of creating a large sample size imbalance in favor of the inferior treatment".

Run-in period.

Poor/extreme schemes to alter the allocation ratios.

Bound the probability of allocation.

Historical Data

Historical data from previous studies can be used to lesson sample size, reducing time and expense, as well as decreasing or eliminating patient exposure to sub-par treatments.

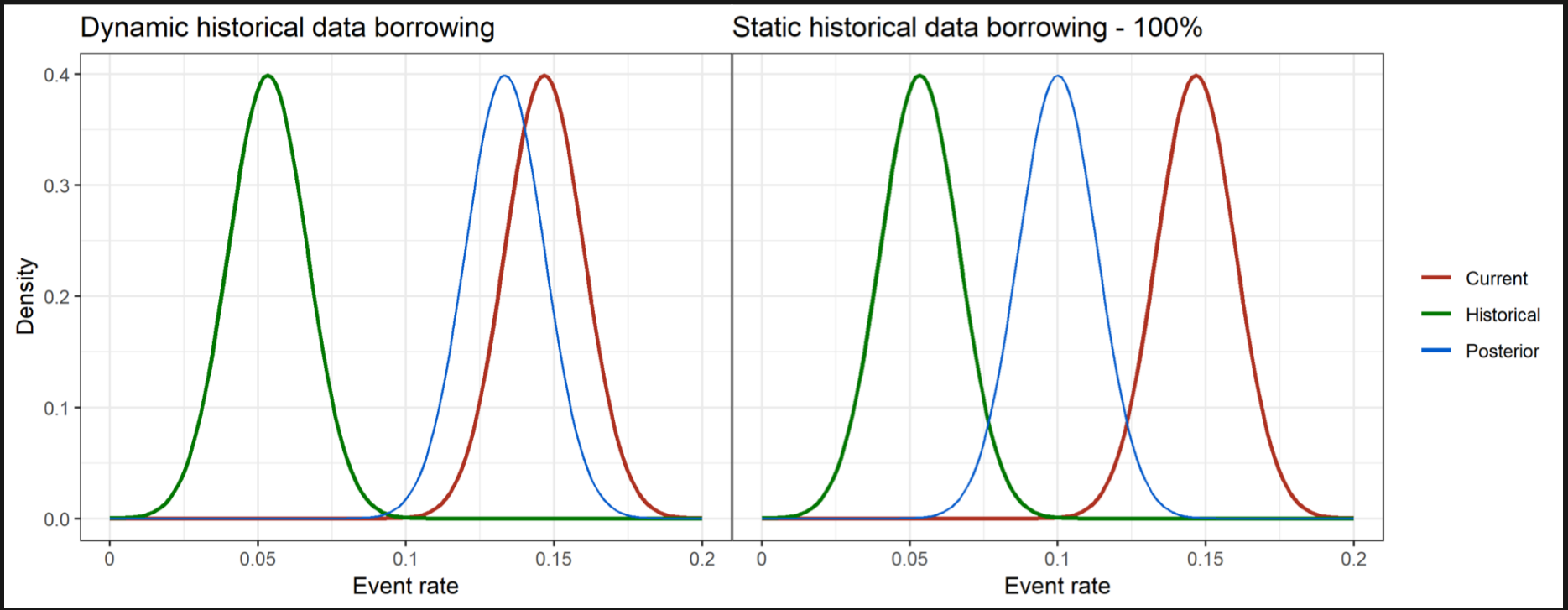
Historical Data

Historical data from previous studies can be used to lesson sample size, reducing time and expense, as well as decreasing or eliminating patient exposure to sub-par treatments.

Examples:

- Previous clinical trials, especially historical control data. Need to ensure same “standard treatment” and inclusion criteria. Published data/reports.
- Note: historical data already used to determine clinically relevant effect size, guide sample size determination, recruitment rates, etc.
- Electronic health records

Historical Borrowing



Binomial Outcome Example

Single-arm binomial count endpoint with incorporation of historical data

Similar event rates between current and historical data

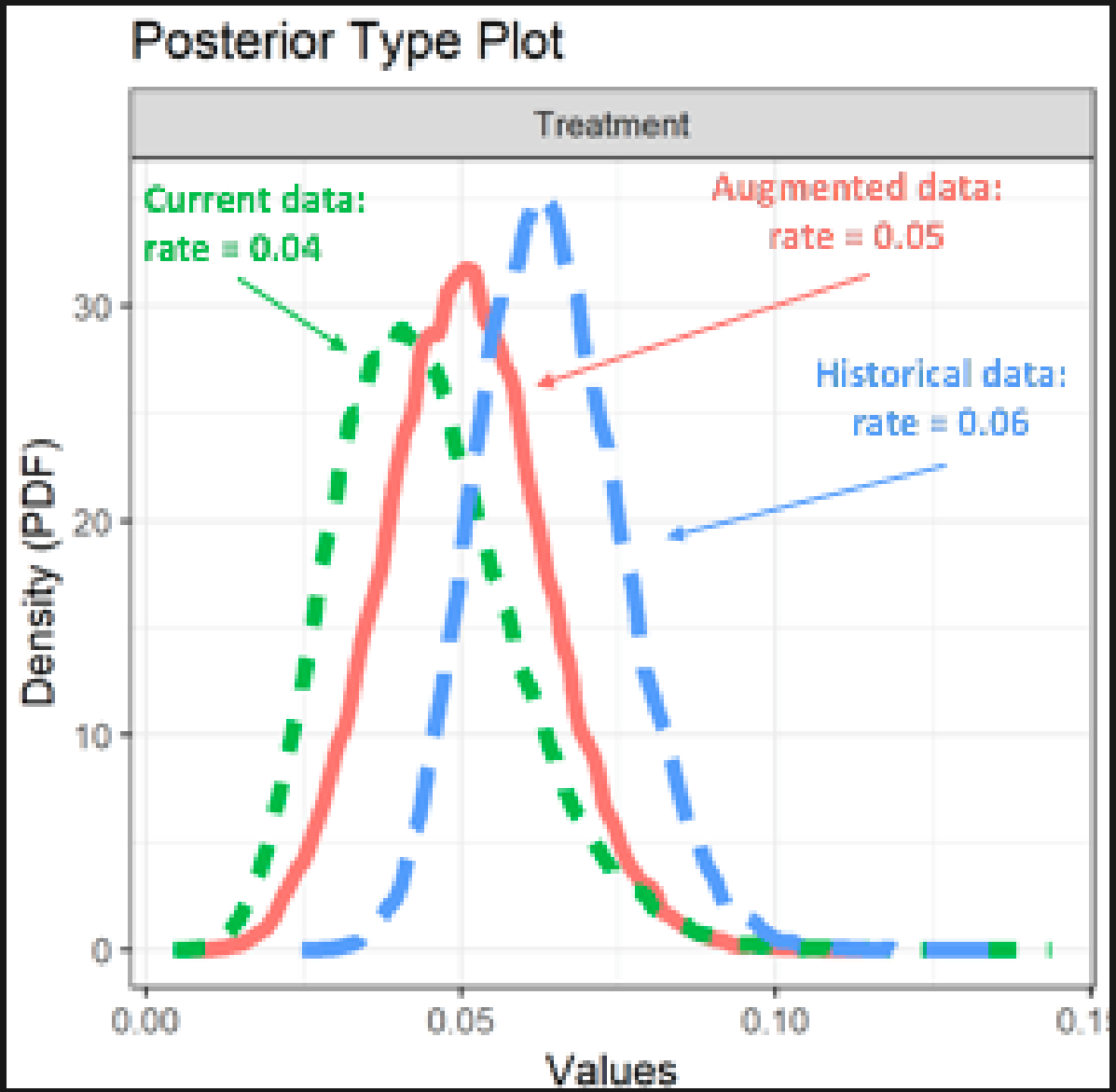
1) Historical data: 28 events in 450 patients

2) Current data: 8 events in 200 patients

- Carry out estimation via bayesDP:

```
library(bayesDP)
# Estimate model via bayesDP using defaults
fit <- bdpbinomial(y_t  = 8,  N_t  = 200,
                   y0_t = 28, N0_t = 450)
```

Binomial Outcome Example



bayesCT R Package

bayesCT R package available: https://thevaachandereng.github.io/bayesCT/

- CRAN release 0.99.2

Analysis types

- Single-arm: objective performance criteria trials, treatment data only
- Two-arm: treatment + control data

Function

- Incorporation of historical data
- Allow early stopping for futility and expected success
- Pipes for modular input & parallelization for fast computing

Binomial Trial: Single Arm

Attain Stability Quad Study

$$H_0 : \pi_{treatment} \geq 0.08 \qquad H_A : \pi_{treatment} < 0.08$$

- Lead related complications at 50 days post-implant
- Enrollment piecewise poisson functions
- Historical data with 5 failures in 55 samples
- *Beta*$(0.5, 0.5)$ prior

Input Parameter

```
value <- binomial_outcome(p_treatment = 0.08) %>%
  enrollment_rate(lambda = c(0.3, 1),
                  time   = 25) %>%
  study_details(total_sample_size     = 900,
                study_period          = 50,
                interim_look          = c(600, 700, 800),
                prop_loss_to_followup = 0.10) %>%
  hypothesis(delta                  = -0.03,
             futility_prob          = 0.05,
             prob_accept_ha         = 0.95,
             expected_success_prob  = 0.90,
             alternative            = "less") %>%
```

Input Parameter: Continued

```
impute(no_of_impute = 1000,
       number_mcmc  = 10000) %>%
beta_prior(a0 = 0.5,
           b0 = 0.5) %>%
historical_binomial(y0_treatment      = 5,
                    N0_treatment      = 55,
                    discount_function = "identity",
                    y0_control        = NULL,
                    N0_control        = NULL,
                    alpha_max         = 1,
                    fix_alpha         = FALSE,
                    weibull_scale     = 0.135,
                    weibull_shape     = 3) %>%
simulate(no_of_sim = 10000)
```

Results

Number of Interim Looks	Power = 0.80	Power = 0.90
1	(160, 210)	(280, 300)
2	(160, 190, 220)	(280, 300, 310)
3	(160, 180, 210, 230)	(280, 300, 310, 320)

Table: Estimated cumulative sample size vectors that provide 80% and 90% power for binomial trial from the example and prior of Beta(0.5, 0.5)

bayesCT R Package

Benefits:

- ease of developing an entire adaptive trial using this package.
- functions are simple to understand for regulatory agency (FDA).
- employable for trials with most common data types.
- ability for a smaller company with a single statistician to be able to develop an entire adaptive clinical trial code.
- universal language used for describing Bayesian adaptive trials.
- working on regulatory agency approval!

Future Work

Regulatory agencies (FDA) tend to evaluate proposals for adaptive designs with great scrutiny.

- limited exposure with adaptive designs.
- concern on the ability of adaptive designs to control type I error.

Future Work

Regulatory agencies (FDA) tend to evaluate proposals for adaptive designs with great scrutiny.

- limited exposure with adaptive designs.
- concern on the ability of adaptive designs to control type I error.

However, adaptive trials can speed up the drug discovery process rapidly.

Future Work in RAR

My future work in RAR will include

- Study multiple treatments rather than 2 treatments.
- Continuous outcomes.

Future Work

My future research work will include

- developing new adaptive designs that alter the inclusion criteria of patients (adaptive enrichment designs).
- developing phase II-III seamless designs.
- creating high-quality open-source computational tools in adaptive designs.

Acknowledgement

- Rick Chappell (UW-Madison)
- David DeMets (UW-Madison)
- Thomas Cook (UW-Madison)
- Kyungmann Kim (UW-Madison)
- Donald Musgrove (Medtronic Inc.)
- Tarek Haddad (Medtronic Inc.)
- 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.

References

- Rosenberger WF, Stallard N, IvanovaA et al. Optimal adaptive designs for binary response trials. Biometrics 2001; 57(3): 909–913.
- Thall PF and Wathen JK. Practical bayesian adaptive randomisation in clinical trials. European Journal of Cancer 2007; 43(5): 859–866.
- Liu, S., & Lee, J. J. (2015). An overview of the design and conduct of the BATTLE trials. Chinese clinical oncology, 4(3).
- Mike V, Krauss AN and Ross GS. Neonatal extracorporeal membrane oxygenation (ecmo): clinical trials and the ethics of evidence. Journal of medical ethics 1993; 19(4): 212–218.

THANK YOU!