

# Introduction to Adaptive Trial Designs

Jay Park

November 7, 2022

# Learning Objectives

1. Discuss the value of adaptive trial designs and contrast them to conventional trial designs
2. Identify the concept and key principles of adaptive trial designs
3. Discuss the basics of simulation-guided design

# Key intended takeaway

- **The foremost step to good clinical trial research is to ask important research questions and answer them reliably**
- Adaptive trial designs are merely a tool for clinical investigation
  - Tools should not be forced to shape and fit the research question
- Research questions should determine what tool we use

# Conventional approach to clinical trial research

# Conventional trial designs (aka fixed sample trial designs)

- General steps follow: **Design**, **Conduct**, and then **Analysis**
- At the **Design** stage, we perform a power or a sample size calculation to determine or to justify a target sample size (or number of events)
- During the **Conduct** stage, we enroll and follow up patients according to protocol
- The **Analysis** stage comes once we finish with the last patient follow-up.
  - In conventional trial designs, we conduct 1 single analysis at the end

# Sample size and statistical power

- Regardless of the design, determination of sample size and statistical power is a fundamental step to clinical trial research
  - **Informally**, *statistical power refers to the probability of detecting an effect, if there is a true effect*
    - Power = 1 - type II error
- Trials should be planned with sufficiently large sample size that allows high statistical power to detect clinically important treatment effect

# Overview of sample size and power calculation

To determine sample size requirement or power, we need to specify

1. False positive (type I) and false negative (type II) error rates: What error rates are we willing to accept?
  - Conventionally, we pick 5% type I error rate (two-sided) and 10-20% type II error rate (80-90% statistical power)
  - All else being equal, sample size requirements will increase when we want less errors
    - Less type I error and less type II error (higher power) mean larger sample size
2. Event rates / variability and effect size
  - We generally require less sample size as control event rate (CER) increases
    - The same when the effect size increases
3. Dropout rates and etc

# Quick planning exercise: A 2-arm trial for mortality

When planning an RCT with a dichotomous endpoint, the sample size calculation requires pre-specification of:

Parameters	Details
Statistical power & type I error rate	We generally need 80% power and 5% type I error rate to stay competitive for grant
Control event rate	What % of control patients do we expect will die by day 28?
Desired (or expected) treatment effects	How effective do we want the treatment to be?
Drop out rate	How many patients will drop out of the study? <i>For today, let's assume this is 0</i>



# How do we determine the control event rate for our trial?

- It is usually estimated from previous literature, such as previously trial on similar population
- Even if an estimate exists from a similar clinical trial, large uncertainty on CER still remains
  - Different sites, time, recruitment, and etc.
  - We always recruit patients based on convenience sampling, never random sampling

# How do we determine the target effect size?

- Ideally, we want to design our trial such that we have high power (e.g., 80-90%) to detect minimal clinically important difference (MCID)
  - MCID represents the smallest improvement that is considered worthwhile
- We want the target effect size to be small, but this increases our sample size requirement
  - What do we do if it is not possible to recruit such large sample size?

# Sample size calculation exercise

- Let's assume control event rate of 0.3
- Say we want 80% power and 5% type I error rate
- A treatment that can reduce the relative risk of dying at day 28
  - Reduction in relative risk should be at least 0.25 to be worth while

# Sample size calculation exercise - Continued

Sample size required for 80% power at 5% type I error rate

CER	RRR	Event rate for Trt	Total N	N per arm
0.3	0.25	0.225	1078	539

# Sample size calculation exercise - Continued

- Assuming 0.3 CER, we need about 1078 patients to detect at least 0.25 RRR with 80% power at 5% type I error rate
- **But what if we were wrong about the event rate?**

# Misspecification of the event rate

- Very common
- We made the assumption of 0.3 for CER
- But what if CER was 10% lower (0.27)?
  - or even 20% lower (0.24)?

# Effects of misspecification of the event rate on sample size

Sample size required for 80% power at 5% type I error rate

CER	RRR	Event rate for Trt	Total N	N per arm
0.30	0.25	0.2250	1078	539
0.27	0.25	0.2025	1241	620
0.24	0.25	0.1800	1444	722

- When CER is lower, we won't have 80% power if our N was 1078
  - If CER was actually higher, we are now “over-powered” with the current target

# Main challenge with the conventional approach

- There are many unknowns. It is extremely difficult to **guess** right or at least uncomfortable making the guesses
- In conventional trials, we only get one guess
  - If you can predict the future, no problem with the conventional approach
- But how do we plan clinical trials when we don't know much about what we are studying? (e.g., COVID-19 at the start of 2020)



# Anticipated regret

- If a study just barely missed its objective but still had a clinically important effect, in retrospect, it would have likely succeeded if the sample size had been slightly larger
- This might suggest that one should have planned for flexible sample size designs (adaptive trial designs) that can “react” to the accumulating trial data
  - Instead of having to wait when the trial is already finished
- The ability to anticipate what one might regret and then planning the trial designs around these areas can be effective in increasing the likelihood of study success

# Adaptive trial designs

# What are adaptive trial designs?

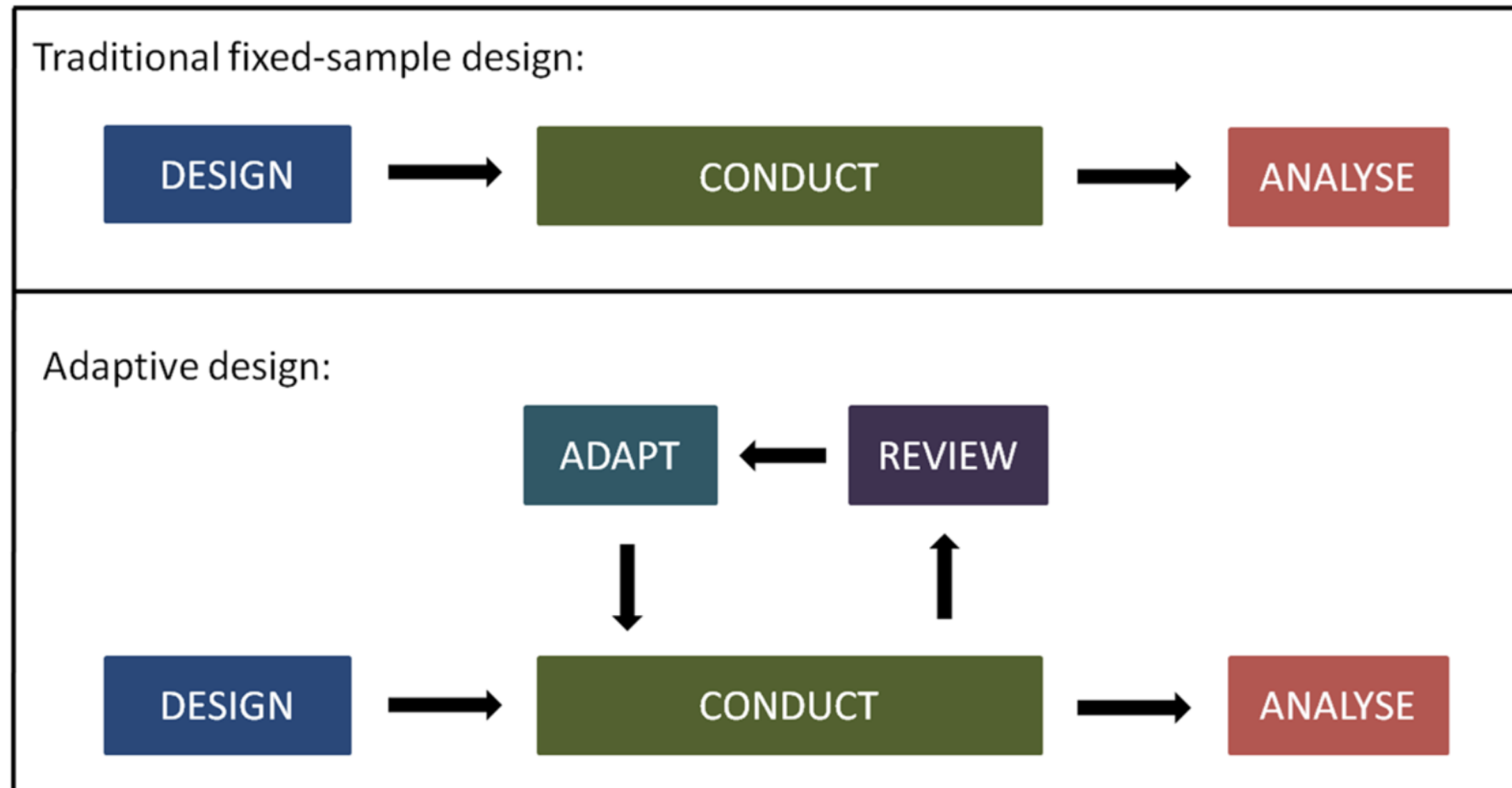
- The term, **adaptive trial designs** is an umbrella term that refers to a group of clinical trial designs that offer pre-planned opportunity to modify aspects of an ongoing trial based on accumulating trial data
- **The unifying property of adaptive trial designs:**
  - Use of accumulating interim data based on **pre-specified plans** that are developed and outlined *a priori*

# Difference between conventional and adaptive trial designs

Recall in conventional trials, we do not use the interim data

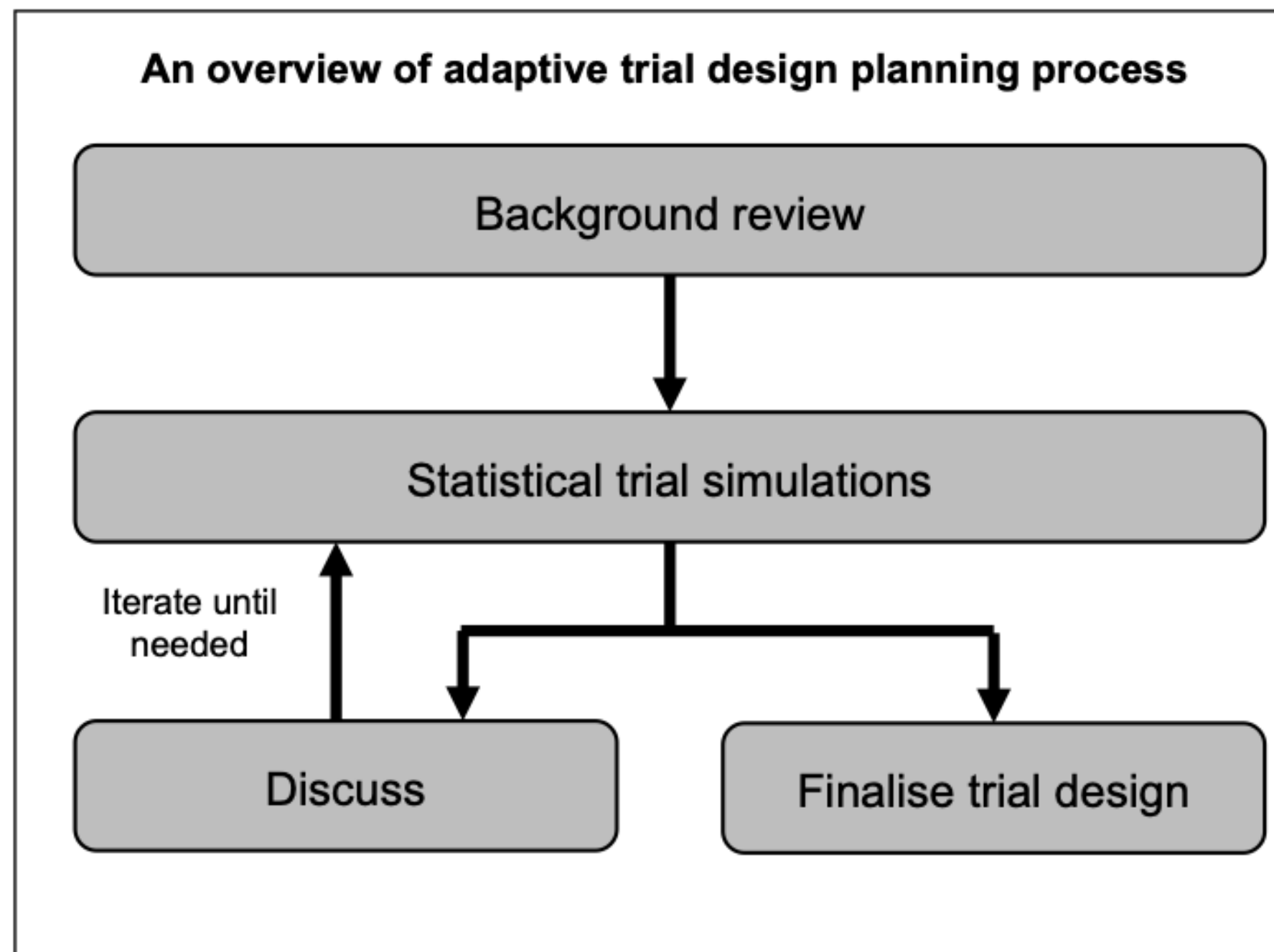
- **Conventional designs:** A fixed sample size and a single analysis at the end
- **Adaptive trial designs:** An umbrella term for various designs where **pre-planned opportunity to modify** the trial designs are permitted based on interim trial data
- **The main motivation for adaptive trial designs is to learn from the data as they are collected during the trial and act accordingly**

# Comparison vs conventional trial designs



- We conduct one or more of the planned interim analyses according to the plan developed during the design stage

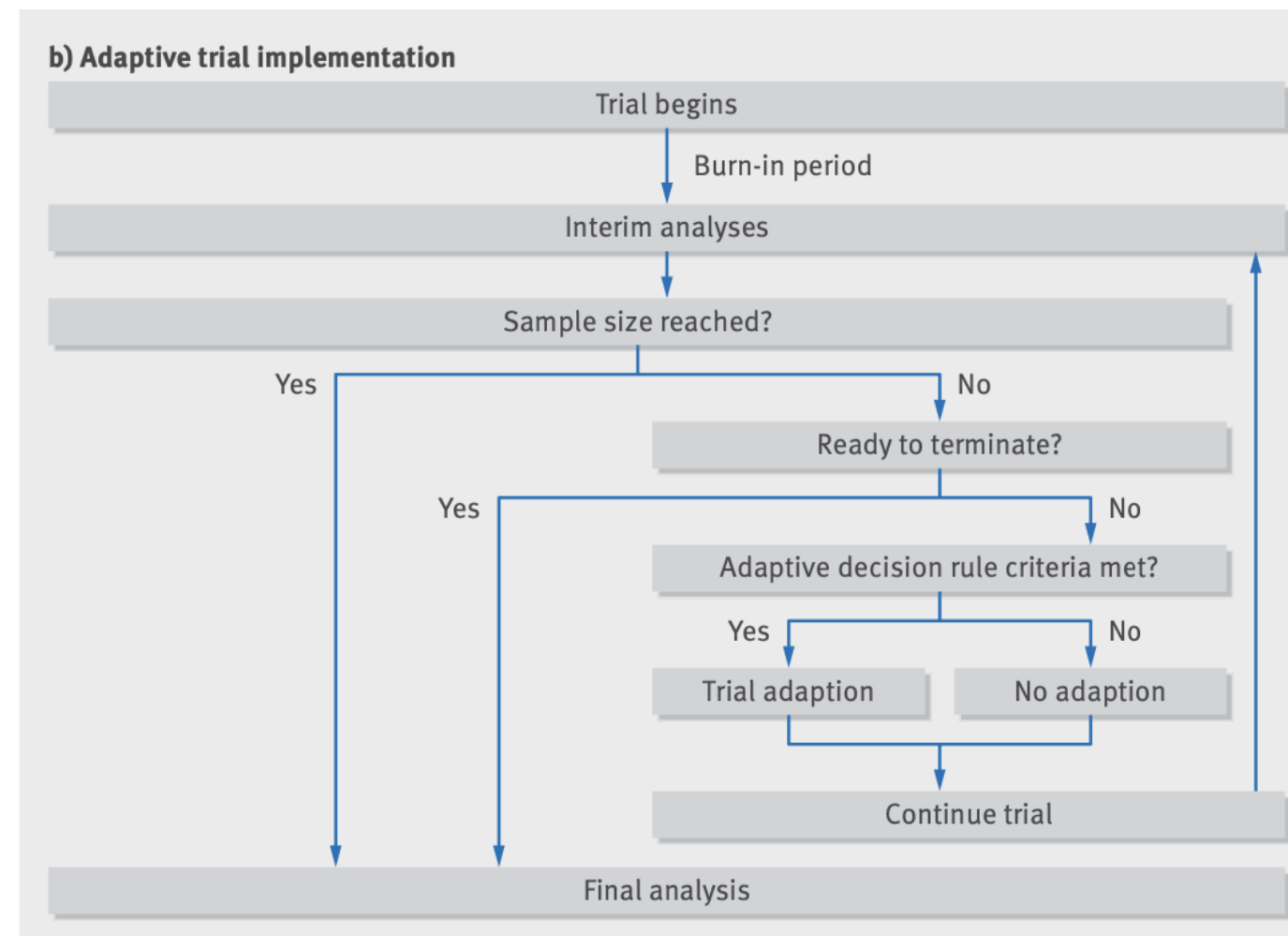
# Adaptive trial designs planning



# Adaptive trial conduct

- This comes after we conduct simulations to come up with the trial designs
- Our plans on interim analyses include specifications of:
  - When will the first interim analysis occur?
    - *Burn-in period*
  - How many interim analyses will we conduct? And how frequently?
  - What adaptations will be allowed? What are the decision criteria?
- In addition to these statistical rules, we specify plans to prevent operational biases
  - Who will conduct the analyses? Who will be blinded and who will not be?

# General steps to adaptive trial conduct



- After the burn-in, we conduct an interim analysis
  - If the criteria for adaptation(s) are met, we adapt.
    - Otherwise, we continue on to the next analysis
- **We only adapt if the pre-specified decision criteria are met**

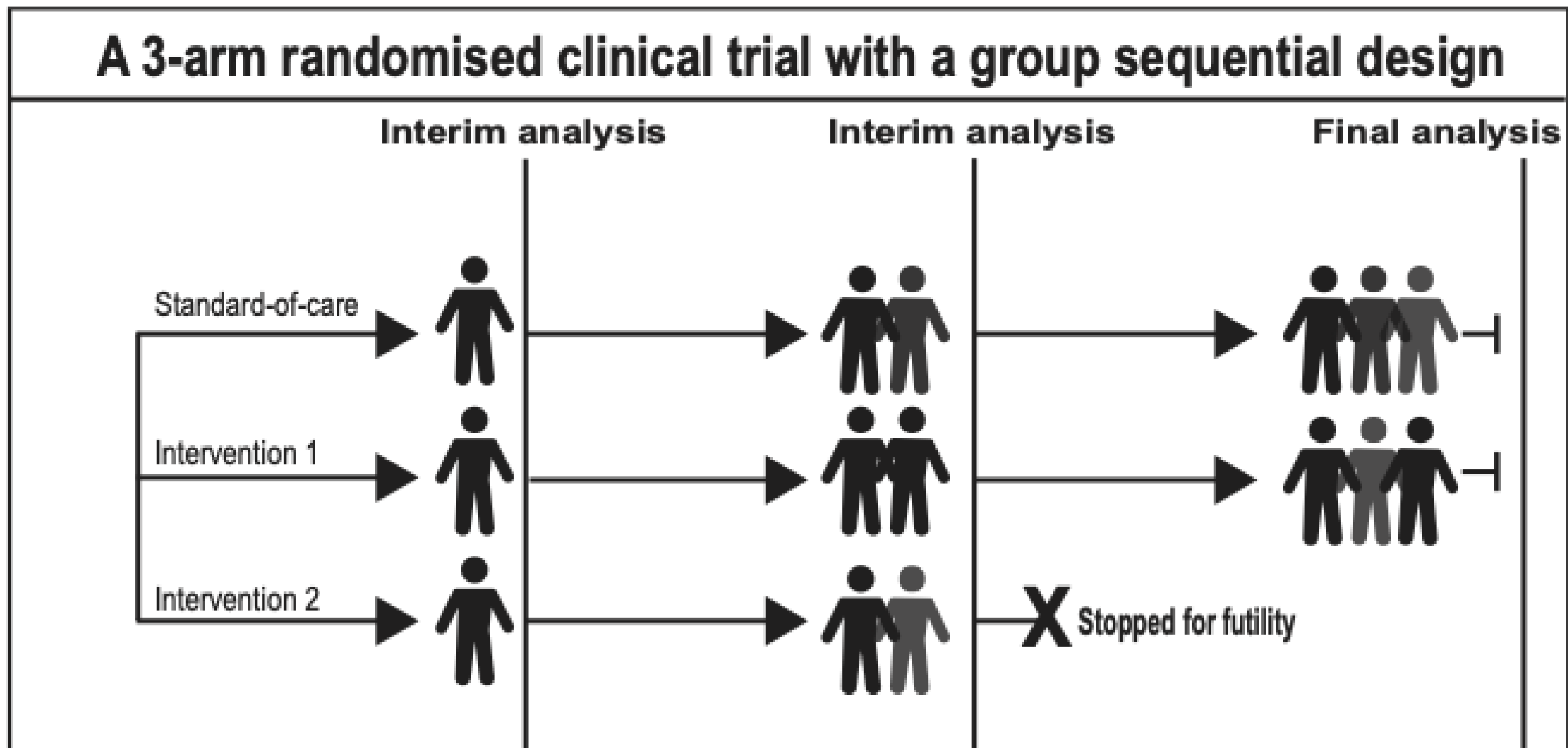


# Common types of adaptive trial designs

# Common types

- Today, we will review the following:
  - Sequential designs
  - Sample size re-assessment
  - Response adaptive randomization

# Sequential designs example



# Sequential designs

This is the most common type of adaptive trial designs

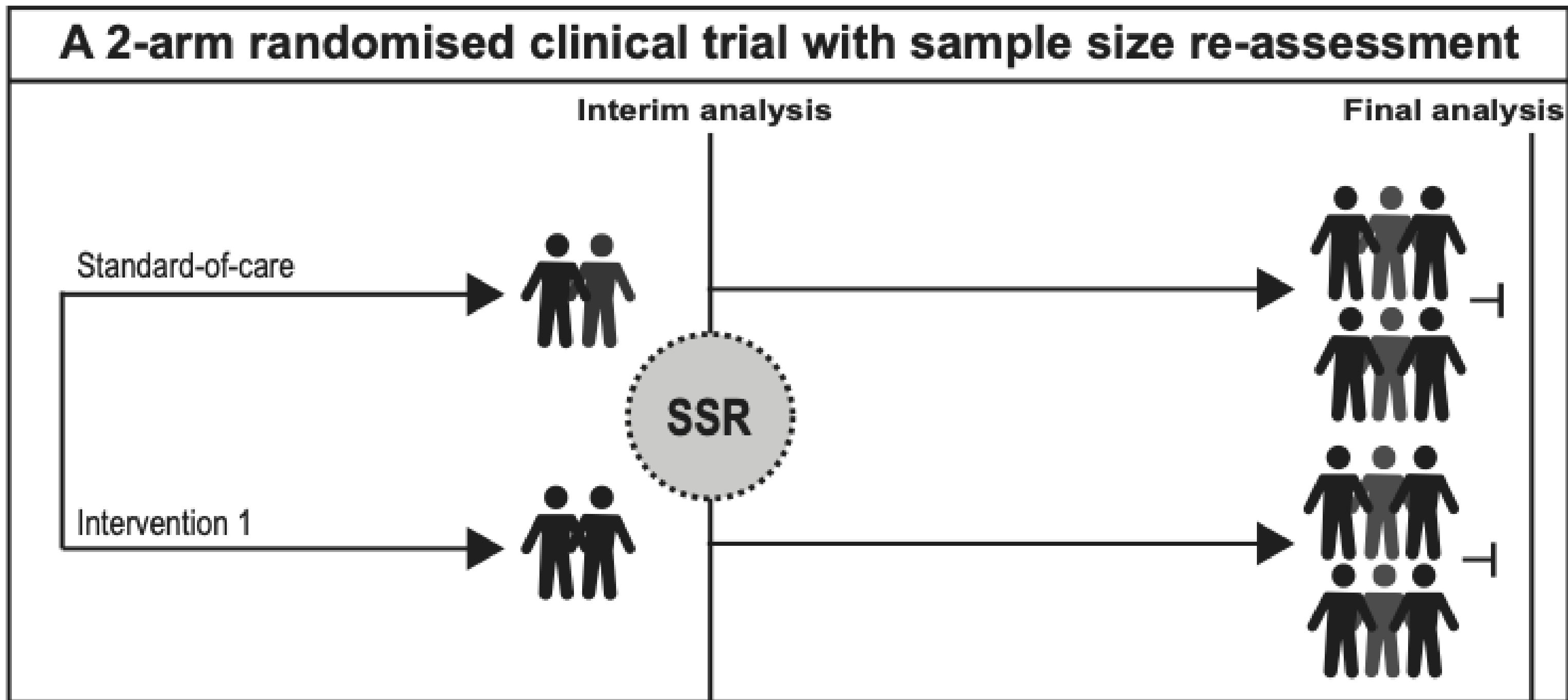
- Refer to a type of trial designs that allow you to stop enrollment early
- You can decide to allow for early stopping based on:
  - **Superiority**: There is overwhelming evidence that the treatment works
  - **Futility**: There is underwhelming evidence for treatment
- You can allow for both superiority and futility, or one of them only

# Motivation for sequential designs

## Fail faster, succeed faster

- In case of overwhelming evidence of efficacy, is completing the full trial necessary?
- If the treatment is really underwhelming (ineffective), is completing the full trial necessary?

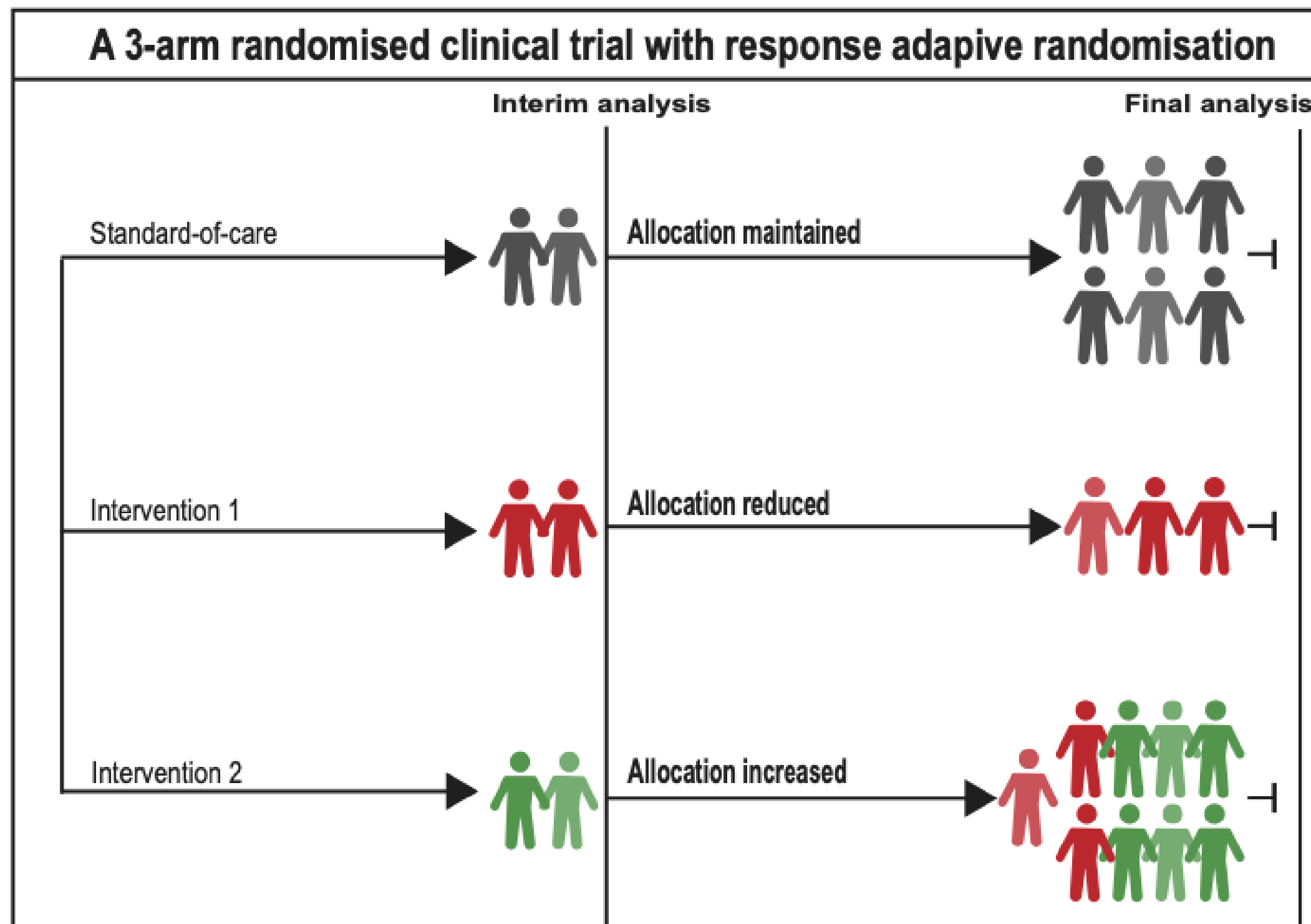
# Sample size re-assessment example



# Sample size re-assessment

- Refer to designs where you can re-assess the sample size during the trial
  - You can do it blinded or unblinded
- With **blinded SSR**, we do not look at the efficacy data at the level of the study arm, only the overall events and etc
  - For binary endpoint, this can involve looking at the overall number of events observed in the trial thus far
- **Unblinded SSR** is commonly combined with sequential designs
  - This one you do look at the data at the study arm level

# Response adaptive randomization example





# Cautions with response adaptive randomization

- Refer to designs where you preferentially adapt the allocation ratio during the trial to the study arm(s) that are performing better
  - Theoretically very appealing
- Challenging to design and implement from both statistical and operational point of view
- For example, decreasing the allocation to the control does have a trade-off for reduced power
  - In a 2-arm trial, 1:1 equal allocation has the highest power

# Simulation-guided design for adaptive trial design planning

# Simulation-guided design planning

- Simulations are necessary to plan clinical trials with complex adaptations
  - Not necessary for conventional trial design planning, but it can make your trial better
- Highly iterative process with dedicated statisticians performing simulations
- Clinical team ask questions and provide feedback for multiple rounds of simulations

# What do we mean by simulations?

- In statistics, simulation generally refers to repeated analyses of randomly generated datasets with known properties
  - Datasets are repeatedly drawn using computer software (e.g., R) under assumptions made on the data-generating mechanism and design specifications
- In the context of trials planning, clinical trial simulations refer to a large number of computer-generated runs performed under various assumptions to evaluate the trial's operating characteristics
  - Type I error rate, power, expected sample size, and etc.

# Benefits of clinical trial simulations

- Useful for planning since they allow evaluation of multiple potential scenarios and candidate designs
  - We can use simulations to compare a fixed trial design option to different adaptive trial designs
- With many unknowns and assumptions that need to be made at the trial planning stage, clinical trial simulations can help to avoid trial design decisions that trial investigators would regret later after the trial shows negative findings (areas of **anticipated regret**)

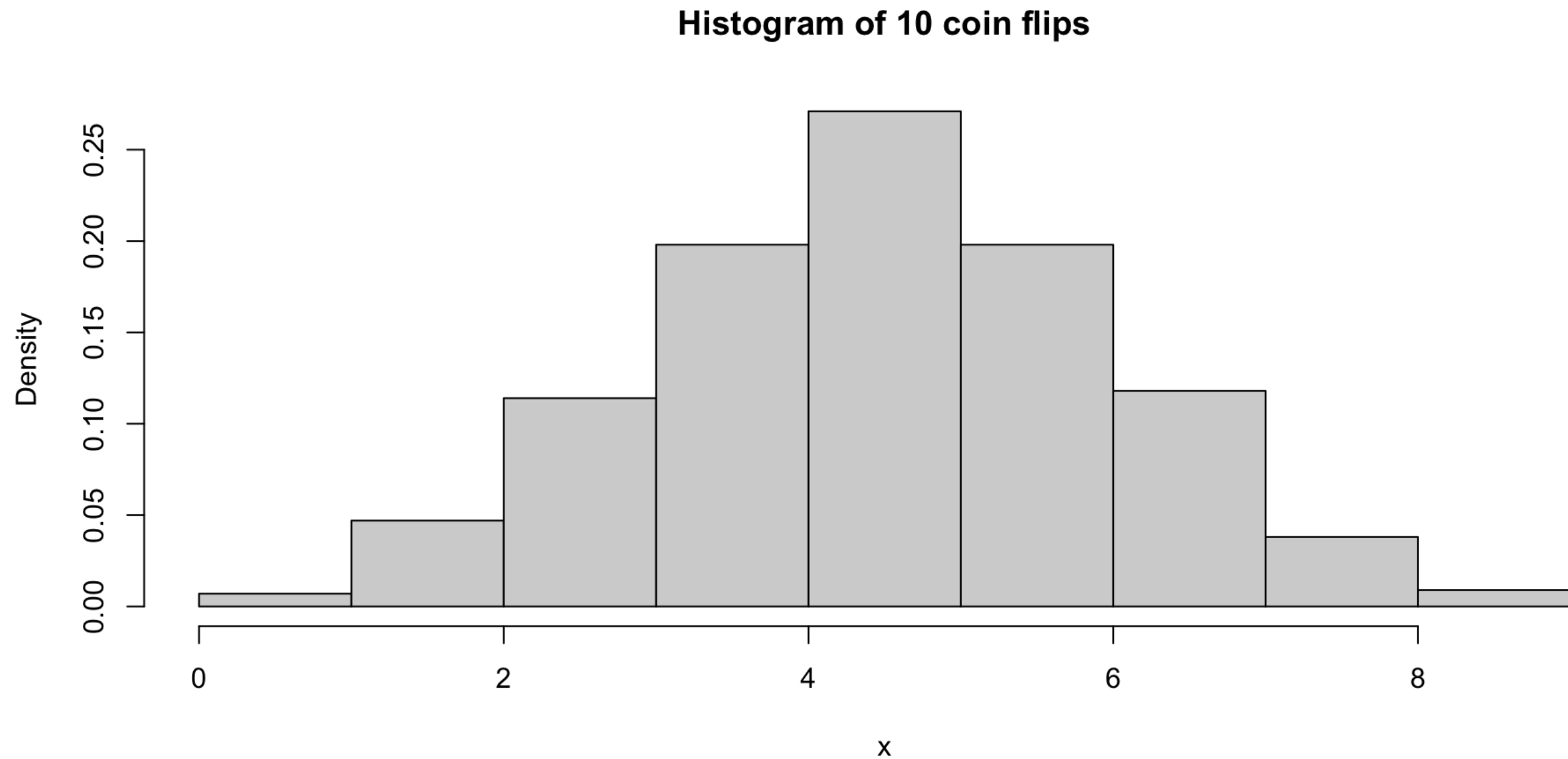
# Simple simulation example

A question for the class:

- Given a fair coin, what is the probability of getting 5 heads out of 10 coin tosses?
- If we manually flipped the coin 10 times, we might get 3 heads and 7 tails
  - If we flipped it another time, we might get a different results
- Given the law of large numbers, we need to repeat the 10 coin tosses many different times (say 1,000) to obtain a reliable estimate
  - Instead of manual tosses, we can use a program like R with a few simple lines of code

# Coin toss example

```
1 x = rbinom(n = 1000, size = 10, p = 0.5)
2 hist(x, probability = TRUE, main = "Histogram of 10 coin flips")
```



# Overview of simulations

1. Choose data generating mechanism / model
2. Specify numerical values for assumptions and analysis
3. Simulate and record the results for each replication
4. Display the results



# Overview of simulations - Continued

General steps	Applied to the coin toss example
Choose data generating mechanism / model	Binary events (heads or tails) generated from a binomial distribution
<b>Specify numerical values:</b>	
-Simulation parameters	Probability of heads = 0.50
-Sample size	10 coin tosses
-Analysis / test statistic	Number of heads out of the 10 tosses

# Overview of simulations - Continued

## For each replication:

A. Generate data according to the model and parameters	Generate 10 random coin tosses
B. Run analysis	Count the number of heads out of 10 tosses
C. Keep track of performance	Keep track of step B and repeat 1000 times
<b>Display the results</b>	E.g., The histogram in the previous slide

# Why do we need simulations?

- Techniques for estimating sample size or power for conventional trials are well established
  - There are mathematical equations for this
- For adaptive trial designs, there are no mathematical expressions (closed form solutions)
  - So we use statistical simulations to evaluate their statistical properties

# Simulation case study

# Simulation case study

- 3 candidate designs for a 2-arm RCT
- Each have a maximum sample size of 200
- Burn in period of 30 patients
  - Interim analyses conducted every 10 patients thereafter
  - 30 (minimum), 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 (maximum) subjects.

# Candidate designs for a 2-arm RCT

1. Sequential designs with equal allocation
2. Response adaptive randomization design with no minimum or maximum allocation
3. Response adaptive randomization design with 10% minimum / 90% maximum allocation

# Assumptions for different scenarios

1. Null effect scenario: Null effect scenario Response rates of control and treatment are both at 30% (30% vs 30%)
2. Alternative effect scenario 1: 10% higher response rate in treatment (**40%** vs 30%)
3. Alternative effect scenario 2: 20% higher response rate in treatment (**50%** vs 30%)

# Simulation overview for the case study

- To compare the design options, 10,000 virtual trials were simulated for each scenario
- Stop early for superiority
  - If the probability of superiority was 99.5% or greater
- Stop early for futility
  - If the probability of superiority was less than 0.5%



# Comments on simulations

- These decision criteria were calibrated to have expected type I error rate that was below 3.3%
- Type I error rate is estimated using the null effect scenario
  - If the treatment had the same response rate as the control, declaring that the treatment was more effective than the control here is a false positive finding
- We can estimate the power using alternative effect scenarios
  - Since we specified that the treatment was better than the control,
    - We want our designs to declare that the treatment is in fact more effective here

# Simulation metrics

Design	% superiority for Exp	N for Trt	N for control	P20
--------	-----------------------	-----------	---------------	-----

---

- The % of superiority for Exp column: % of virtual trials where the Treatment was superior (0.995) over the control
- The N for Exp and the N for control: Average sample size w/ 2.5th and 97.5th percentile in bracket
- The P20 column: % of virtual trials that enrolled at least 20 more patients on the control arm over the Experimental treatment arm.
  - *Imbalance of 20 patients represents a significance difference that is beyond random chance*

# Simulation results for null effect scenario

Design	% superiority for Exp	N for Exp	N for control	P20
Fixed alloc.	0.032	96 (27,114)	97 (27,113)	0.07
RAR no min	0.030	96 (18,177)	96 (18,178)	<b>0.42</b>
RAR w/ 10% min	0.033	95 (22,167)	97 (23,168)	<b>0.42</b>

# Comments on the null effect scenario

- Even though our type I error rate is under the desired rate, imbalance of patient allocation is concerning for the RAR designs
- In both RAR designs, there is 42% probability that more patients will be allocated by to the treatment arm
  - Again **P20** represents imbalance that is beyond random chance

# Simulation results for alternative effect scenario 1 (10% higher response rate for Trt)

Design	% superiority for Exp	N for Exp	N for control	P20
Fixed alloc.	0.24	89 (21,113)	90 (21,114)	0.06
RAR no min	0.15	128 (24,182)	56 (14,155)	0.10
RAR w/ 10% min	0.19	124 (26,171)	58 (16,149)	0.10

# Comments on alternative effect scenario 1

- The expected power is generally poor
- The power is highest for fixed allocation design

# Simulation results for alternative effect scenario 2 (20% higher response rate for Trt)

Design	% superiority for Exp	N for Exp	N for control	P20
Fixed alloc.	<b>0.72</b>	68 (18 - 110)	68 (18 - 110)	0.04
RAR no min	<b>0.43</b>	123 (23 - 183)	32 (12 - 90)	0.01
RAR w/ 10% min	<b>0.56</b>	111 (22 - 172)	36 (14 - 88)	0.01

# Comments on alternative effect scenario 2

- The expected power improved, but we see the same trend of power being greater in fixed allocation design
- The trade-off on statistical power should be noticed

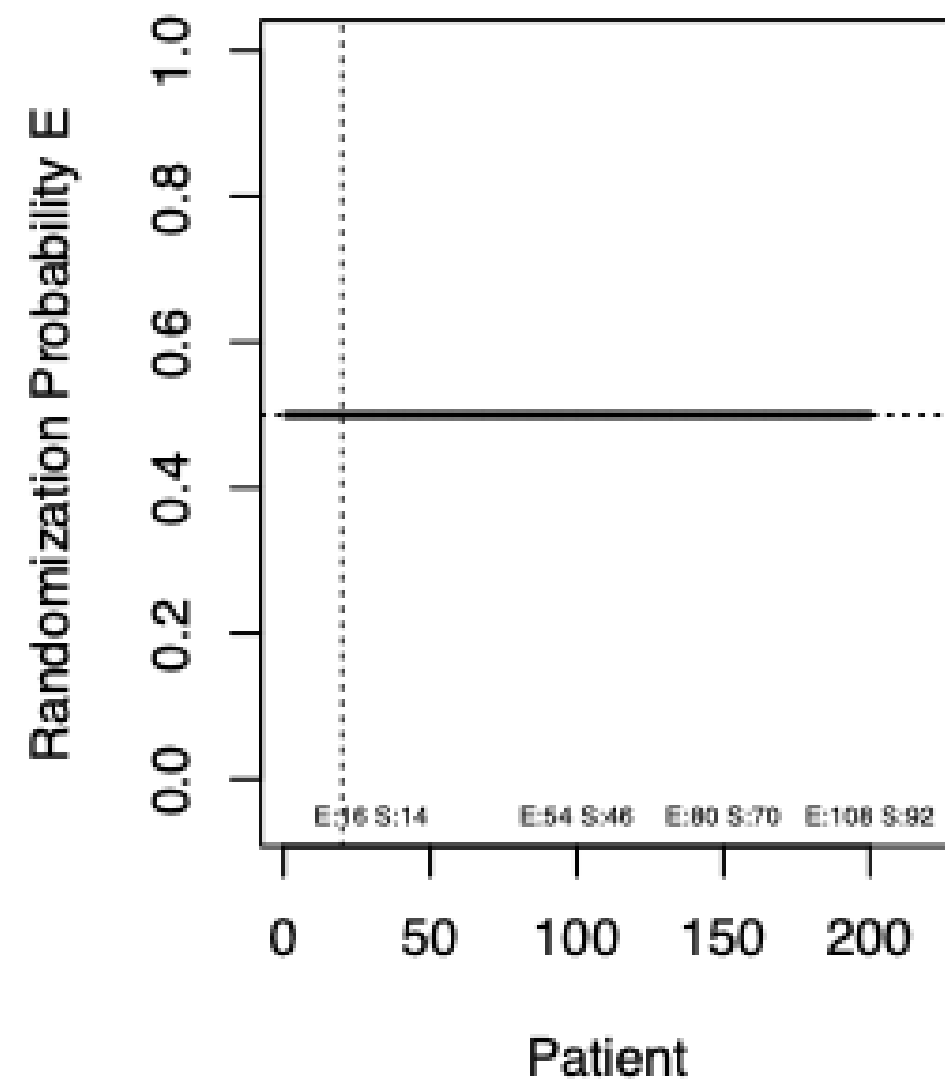


# Examination of single virtual trials

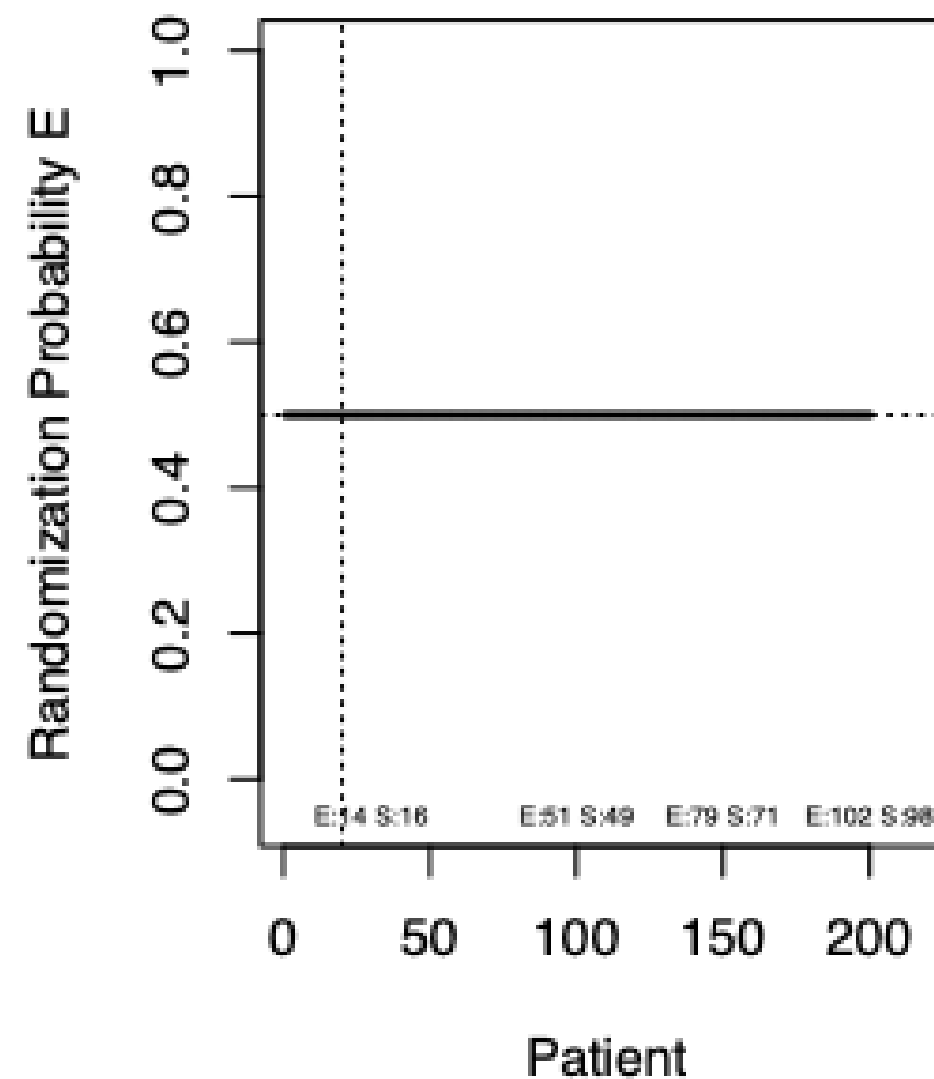
- We usually repeat simulations many times to adhere to the law of large numbers and to avoid Monte Carlo errors
  - However, it is important to note that a trial in reality will be conducted only once
- It is critical to look at examples of single virtual trials to understand how the trial may progress to illuminate many potential issues that may arise
  - Beyond average operating characteristics
- The specific details of what needs to be looked at depends on the nature of the trial designs and should be tailored to investigate any of the planned adaptations

# Virtual trials for fixed allocation design

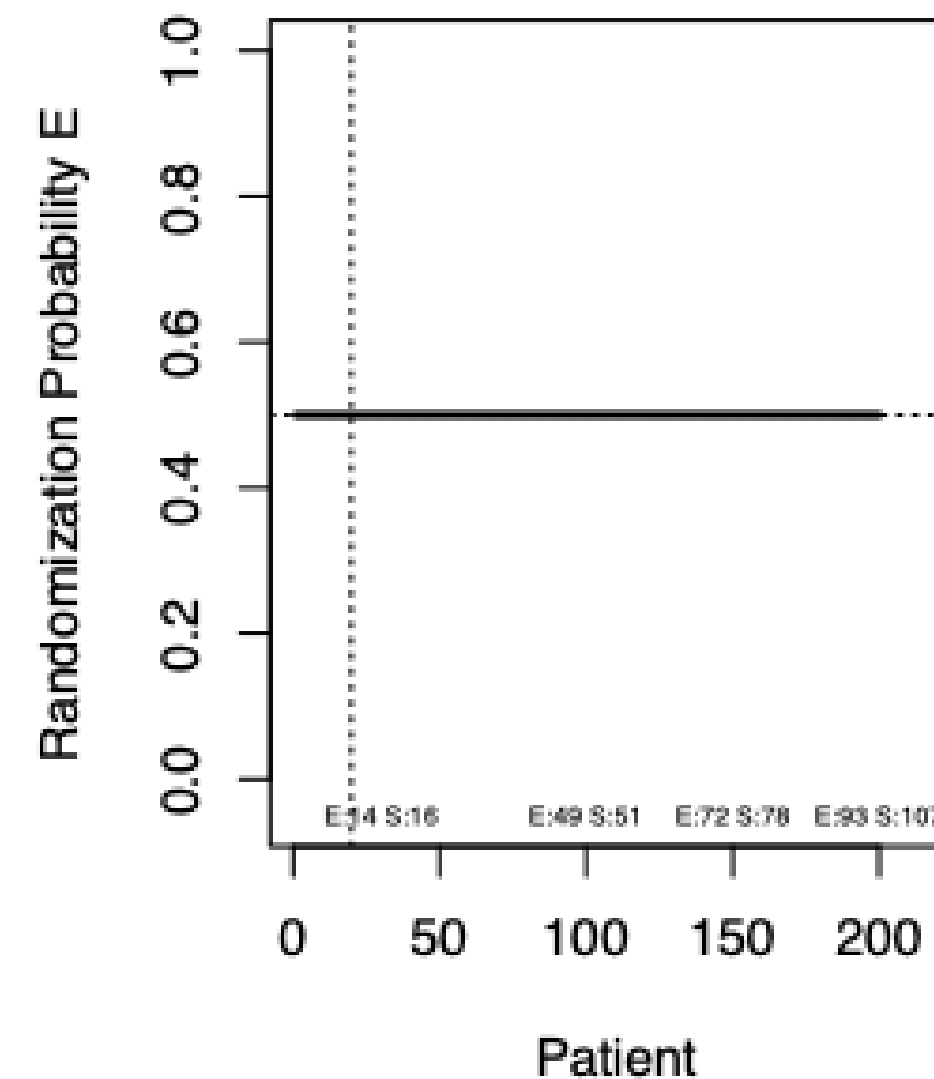
Virtual Trial 1



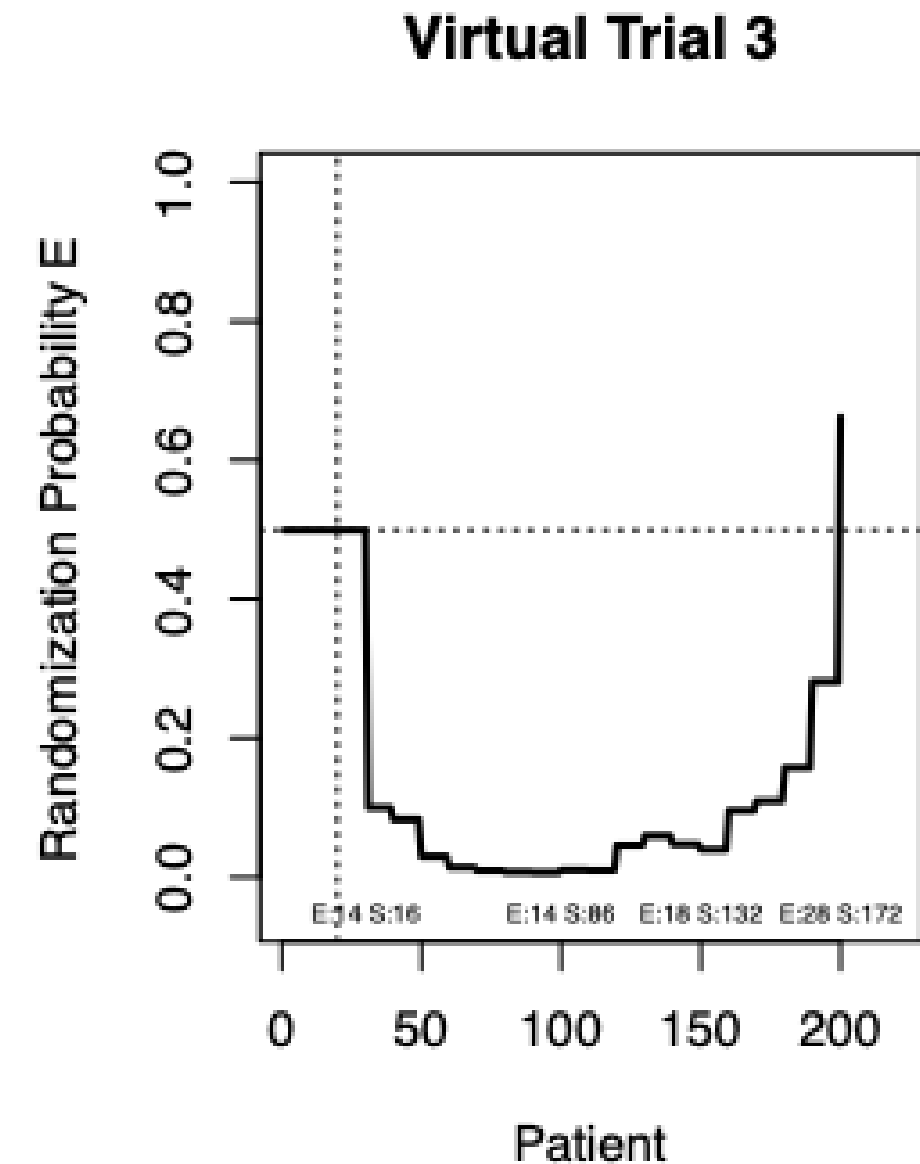
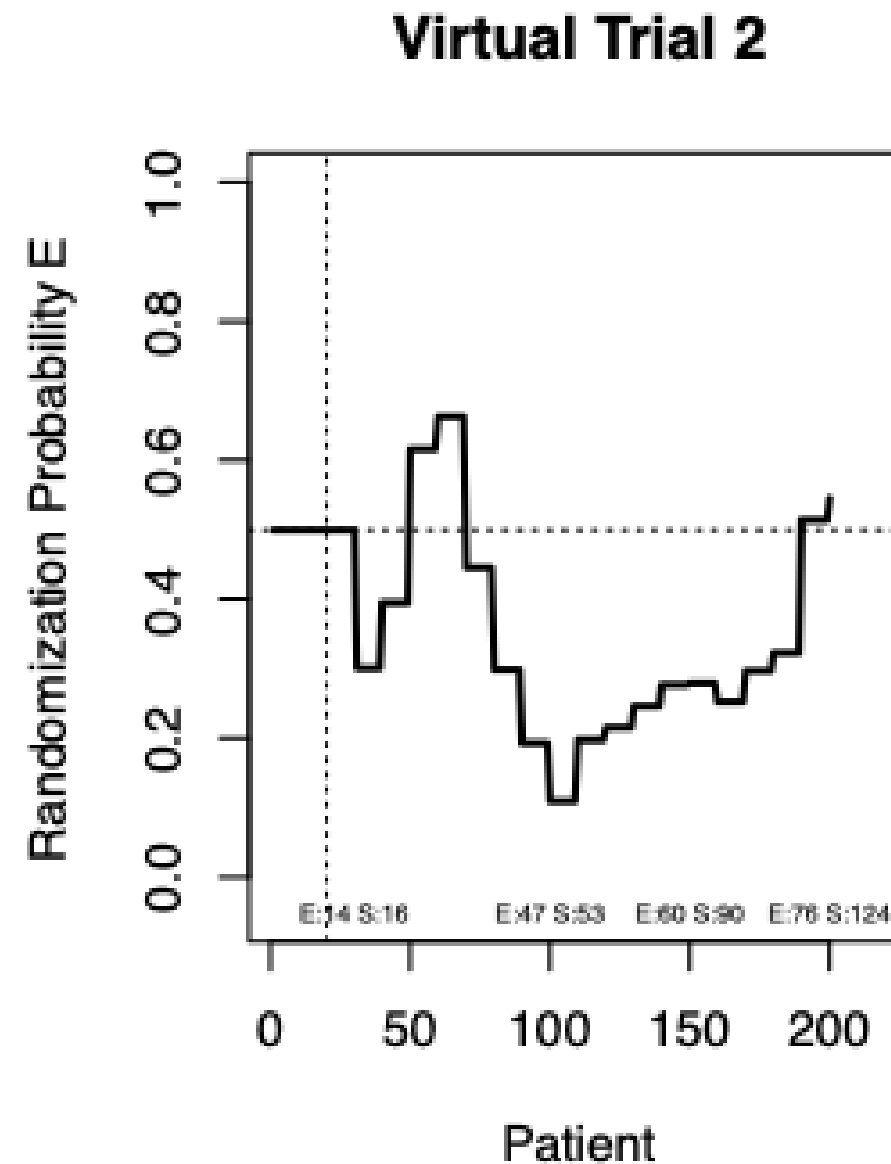
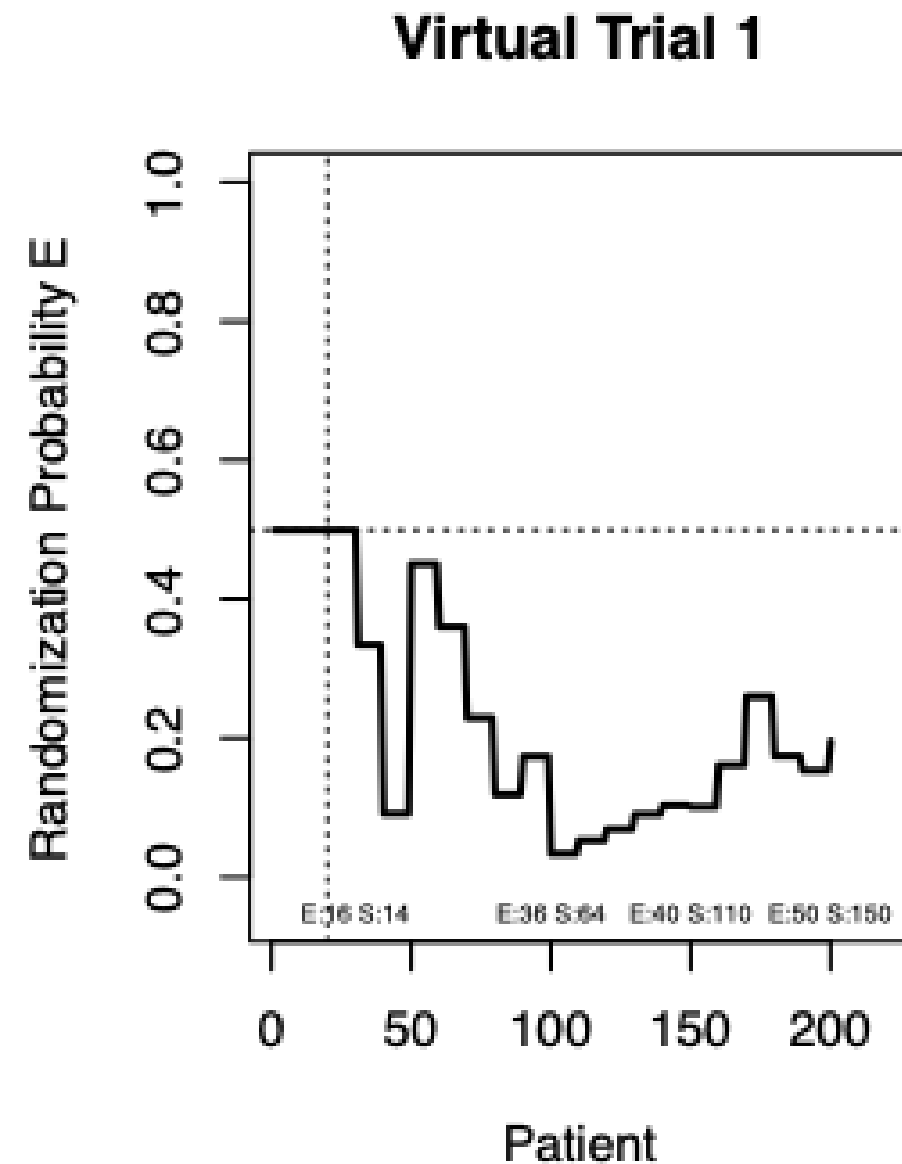
Virtual Trial 2



Virtual Trial 3

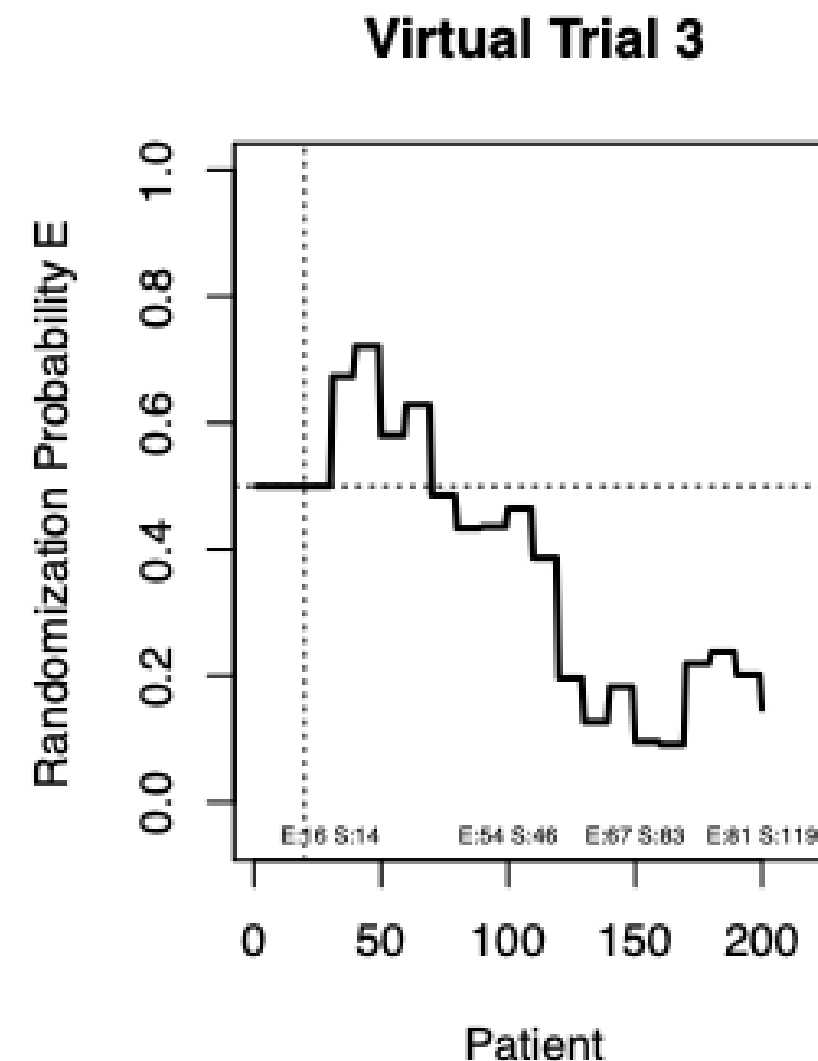
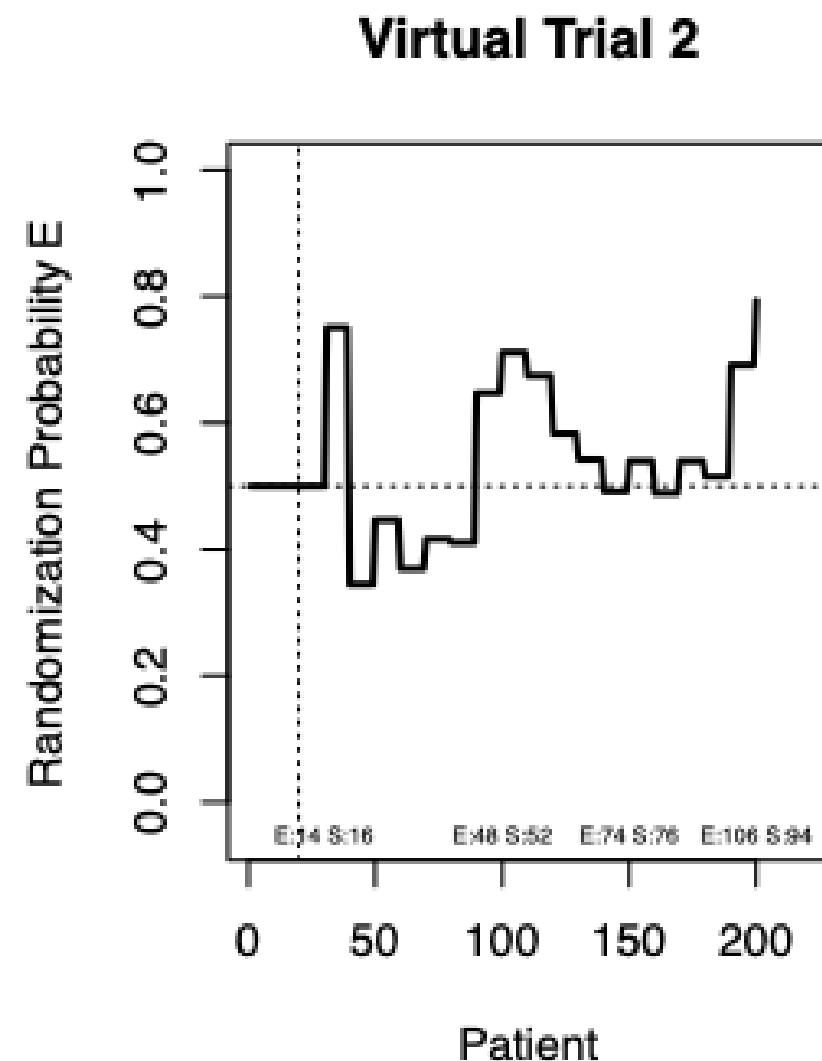
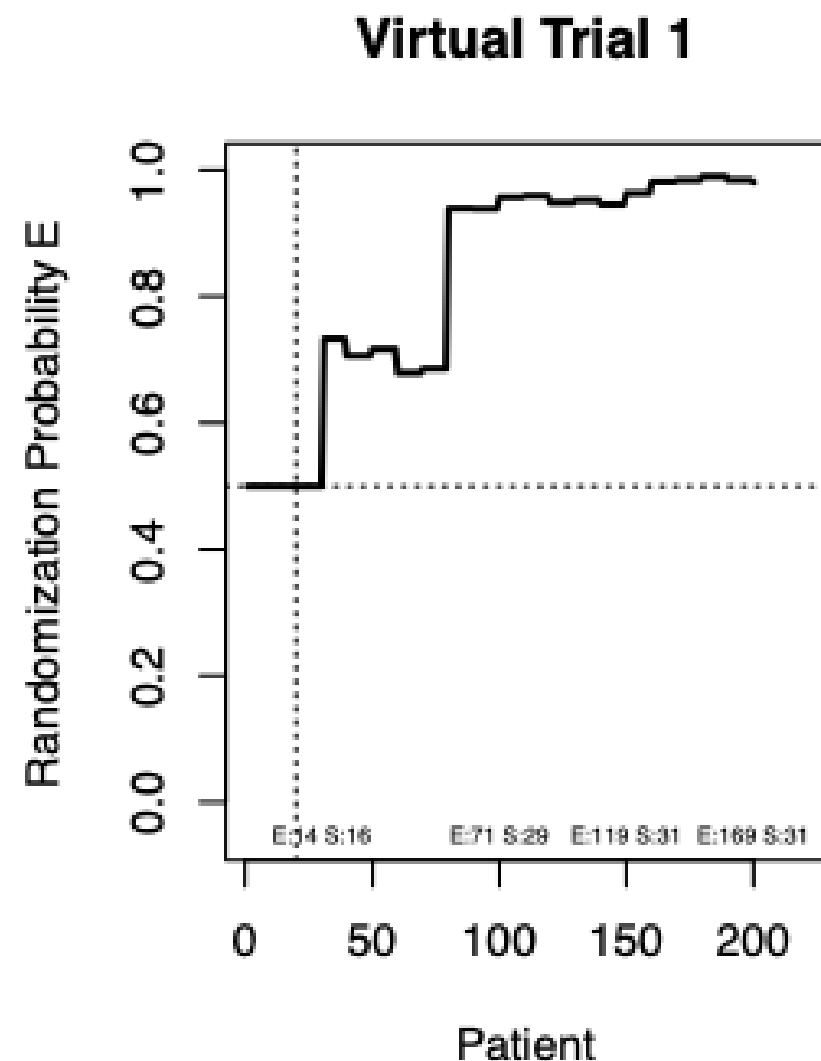


# Virtual trials for RAR design under the null scenario



If your trial looked like one of these trials, would you be happy?

# Virtual trials for RAR design under the alternative scenario 1 (10% higher response rate for treatment)



What about now?

# Intended takeaways from the case study

- This is not to illustrate with just one example that all response adaptive randomization procedures are flawed
- When planning adaptive trial designs, theoretical appeals should be evaluated and confirmed with simulations
- We should not draw a broad general conclusion. Trade-offs should be carefully considered.
  - There are merits to each design option

# Final comments

# Final comments

- Whether they are adaptive or not, no designs should be chosen by default
  - Each question requires careful considerations
- Adaptive trial designs can be more challenging and complex to plan and execute than conventional trials
- It generally does take longer time and more efforts to plan, but the efforts are generally well worth it
- Using simulation-guided trial designs can improve our ability to conduct conventional trial designs as well

# References

1. Dimairo M, Pallmann P, Wason J, Todd S, Jaki T, Julious SA, Mander AP, Weir CJ, Koenig F, Walton MK, Nicholl JP. The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ* 2020 Jun 17;369
2. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, Holmes J, Mander AP, Odondi LO, Sydes MR, Villar SS. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC medicine*. 2018 Dec;16(1):1-5
3. Thorlund K, Haggstrom J, Park JJ, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians. *BMJ*. 2018 Mar 8;360
4. A Practical Adaptive & Novel Designs and Analysis (PANDAS) toolkit



# Book

## **Introduction to Adaptive Trial Designs and Master Protocols**

**Jay J. H. Park**

McMaster University, Ontario

**Edward J. Mills**

McMaster University, Ontario

**J. Kyle Wathen**

Cytel, Cambridge, Massachusetts



- Master protocols are beyond the scope for this week, but watch out for this book starting next year if interested

Upcoming book on adaptive trial designs and master protocols