ADAPTIVE HEALTH INTELLIGENCE

EVIDENCE IN ACTION

Designing Efficient Clinical Trials

Michael Dymock

8th August 2023

TELETHON KIDS INSTITUTE
Discover. Prevent. Cure.

WESFARMERS CENTRE OF VACCINES & INFECTIOUS DISEASES

THE UNIVERSITY OF SYDNEY

THE UNIVERSITY OF WESTERN AUSTRALIA

About me…

- 2018:
  - BSc (Hons) Mathematics & Statistics
  - Berwin Turlach and Kevin Murray
- 2019-2020:
  - UWA Centre for Applied Statistics
- 2020 onwards:
  - Biostatistician at Telethon Kids, et. al.
- 2023 onwards
  - PhD with Kevin Murray, Julie Marsh and Tom Snelling
  - Council member of SSA WA Branch
  - Member of IBS-AR

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Overview

- Why design clinical trials?
  - Type I error, power, sample size
- Fixed designs
  - Simulations
- (Bayesian) adaptive designs
  - Simulations
- Some examples
- Questions for the future

pic of
us in the
ıte.
ı
cts Tom
's logic
ın
h your a
ıe didn't
ıey had

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Why design clinical trials?

- Better science
- Efficient use of resources
- Ethical reasons
- To keep biostatisticians employed?

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Questions to consider

- What is the research question?
- What are we wanting to measure/observe?
- What is the (primary) endpoint/outcome?
- What is the hypothesis?
- What is our desired type I error and power?
- What sample size do we require?

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

A simple example (infectious diseases and vaccines)

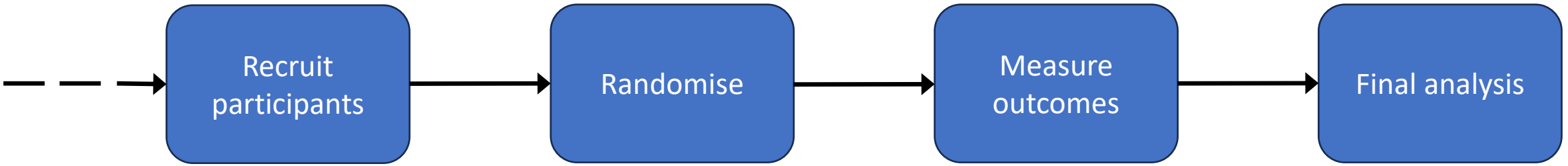
- What is the research question?
  - **Which vaccine (A or B) will offer the greatest protection against the disease?**
- What are we wanting to measure/observe?
  - **Immune response to vaccination**
- What is the (primary) endpoint/outcome?
  - **Log10 antibody concentration at 28 days after vaccination**
- What is the hypothesis?
  - **Vaccine B produces a greater antibody response than Vaccine A**
- What is our desired type I error and power?
  - **Type I error = 0.07; Power = 0.85**
- What sample size do we require?

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Fixed trial designs

- The trial design is **fixed** prior to trial commencement
  - Number of trial arms, randomisation probabilities, sample size, etc. does not change as the trial progresses
- “Easy” to design the trial
  - Choose the design to obtain desired *operating characteristics*
- “Easy” to implement the trial
  - Understandable for participants, analysts, scientists
- Opportunity costs?
  - Information gathered during the trial cannot be used

Fixed trial designs



Sample size calculations

- "With a sample size of **N**, the study is powered at 85% to detect a clinically important difference of **X** units whilst maintaining the type I error below 7%."
- How do we determine **N** and **X**?
- Clinically important difference
  - Stephen Senn: That which is used to justify the sample size but will be claimed to have been used to find it.

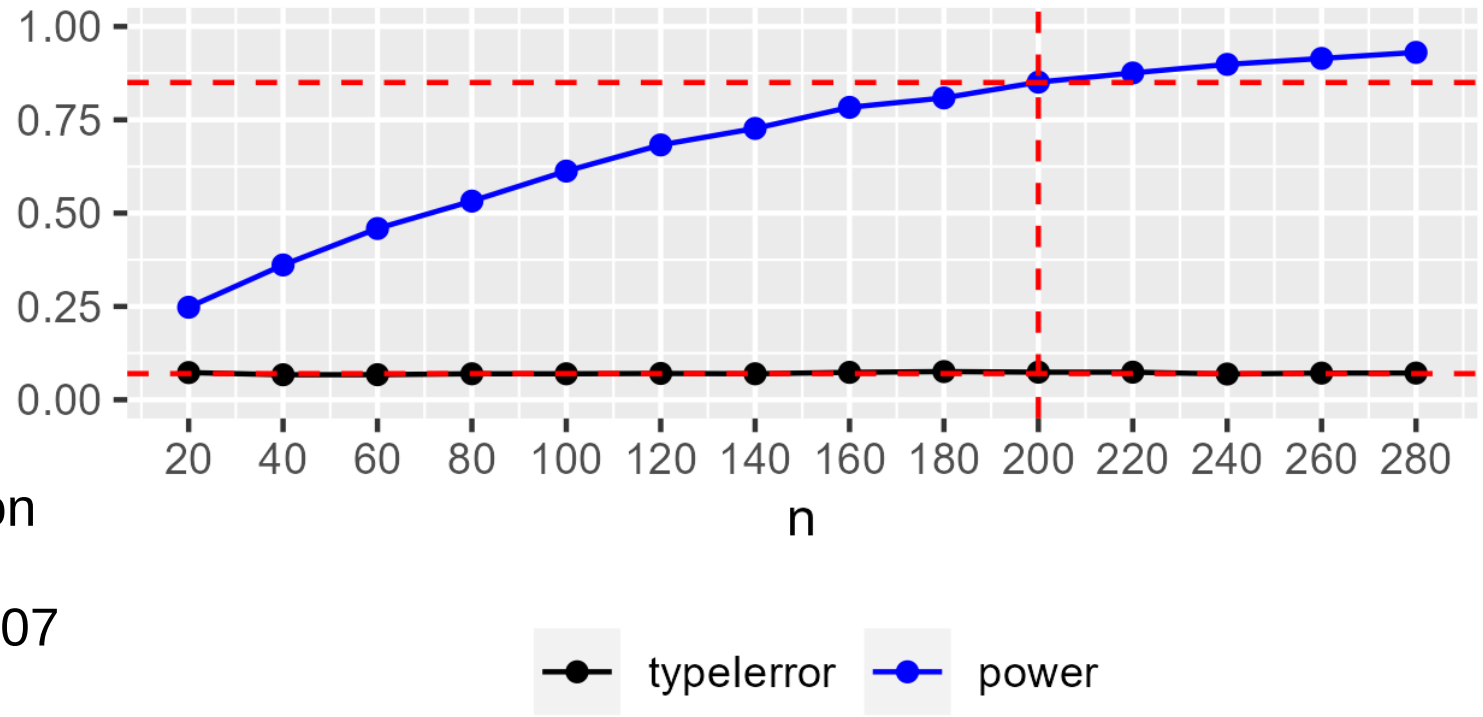
Back to the vaccine example

- What sample size do we need if the clinically important difference is 0.5 units on the log10 scale?
- Suppose we were doing a simple z-test to compare mean log10 antibody concentration assuming known equal variance of 4 units
  - $H_0: \mu_B - \mu_A = 0$
  - $H_1: \mu_B - \mu_A > 0$
- Use a formula!
- $N = 2\left(\frac{(z_\alpha + z_\beta)\sigma}{\delta}\right)^2 = 2\left(\frac{(1.48+1.04)\times 2}{0.5}\right)^2 \approx \mathbf{202}$

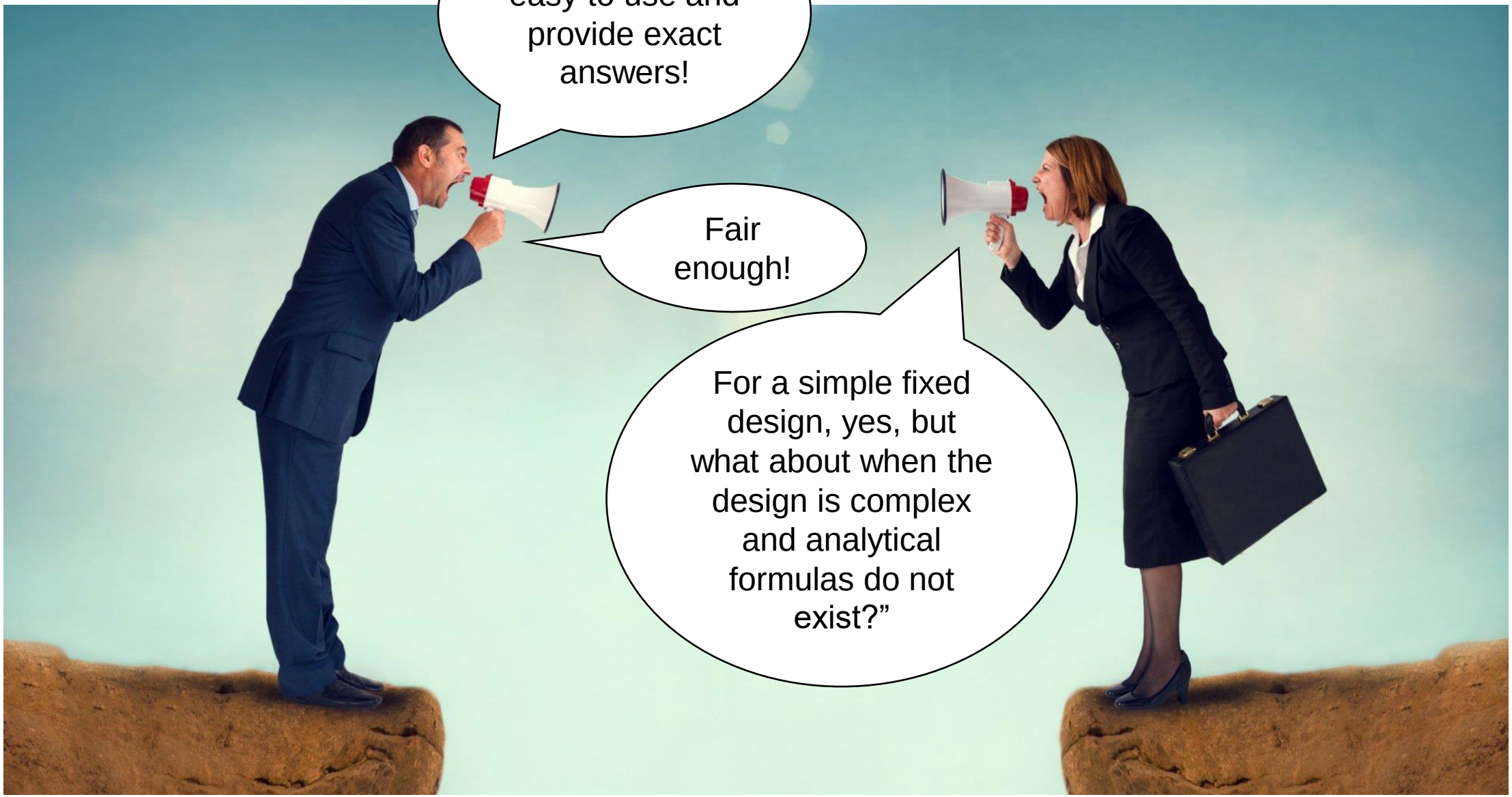
ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Is there another way?

- Let's pretend we did not know the formula (or have internet access to Google it)
- We can use simulation instead!
  - Consider $n \in \{20,40, \ldots, 280\}$
  - "Null" & "effect" scenarios
  - Simulate 10,000 trials for each n
  - Compute p-value for each simulation
  - Power = proportion of p-values < 0.07
  - Type I error maintained at 0.07
  - Approximately 0.85 power at n = 200



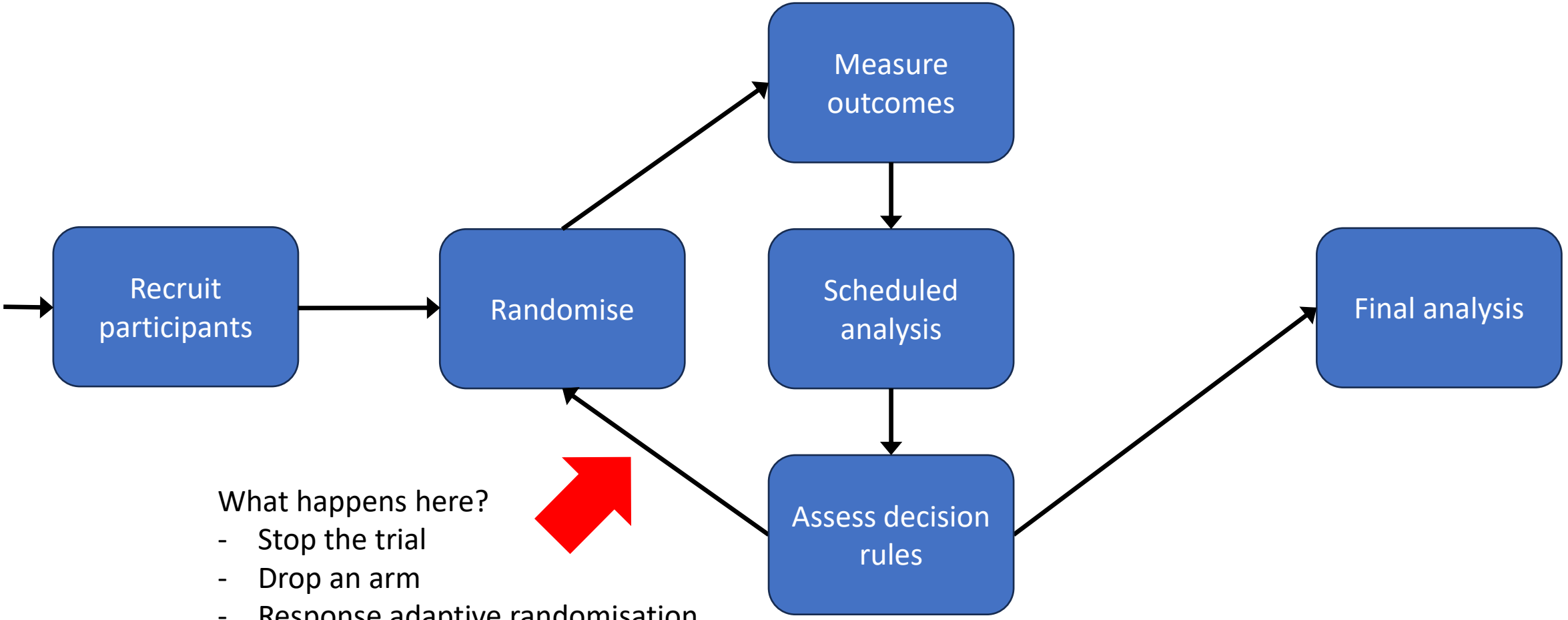
But why?


Formulas are easy to use and provide exact answers!
Fair enough!
For a simple fixed design, yes, but what about when the design is complex and analytical formulas do not exist?”

Adaptive trial designs

- The trial design **adapts** in response to the accrued data
  - Number of trial arms, randomisation probabilities, sample size, etc. **may** change as the trial progresses
- “Hard” to design the trial
  - Requires simulations, many levers to pull
- “Hard” to implement the trial
  - More complicated for participants, analysts, scientists
- Opportunity gains?
  - Information gathered during the trial can be used to increase design efficiency

Adaptive trial designs



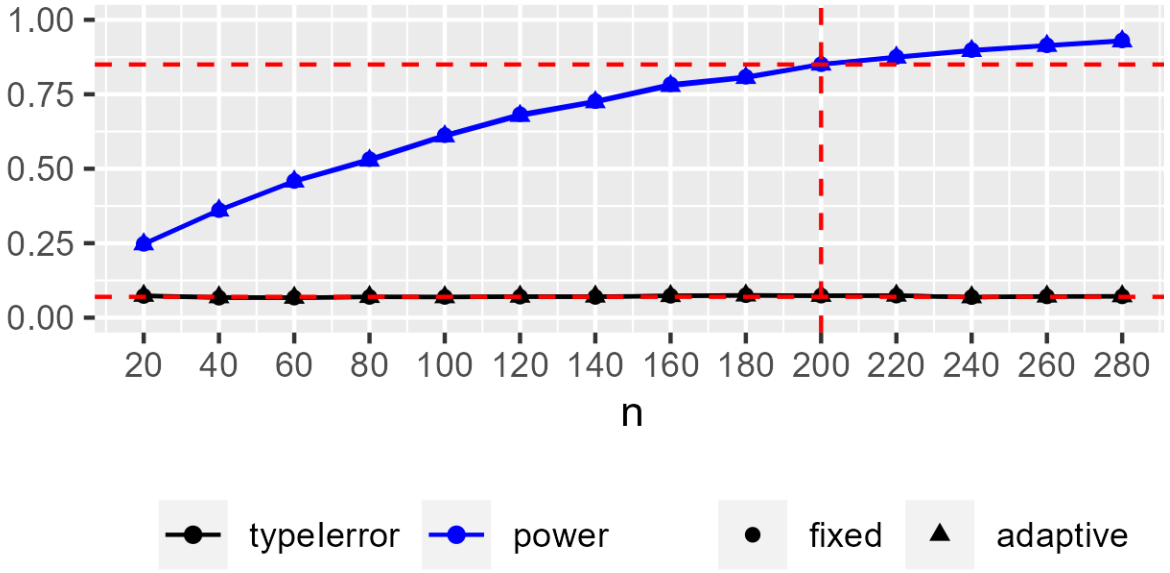
What happens here?

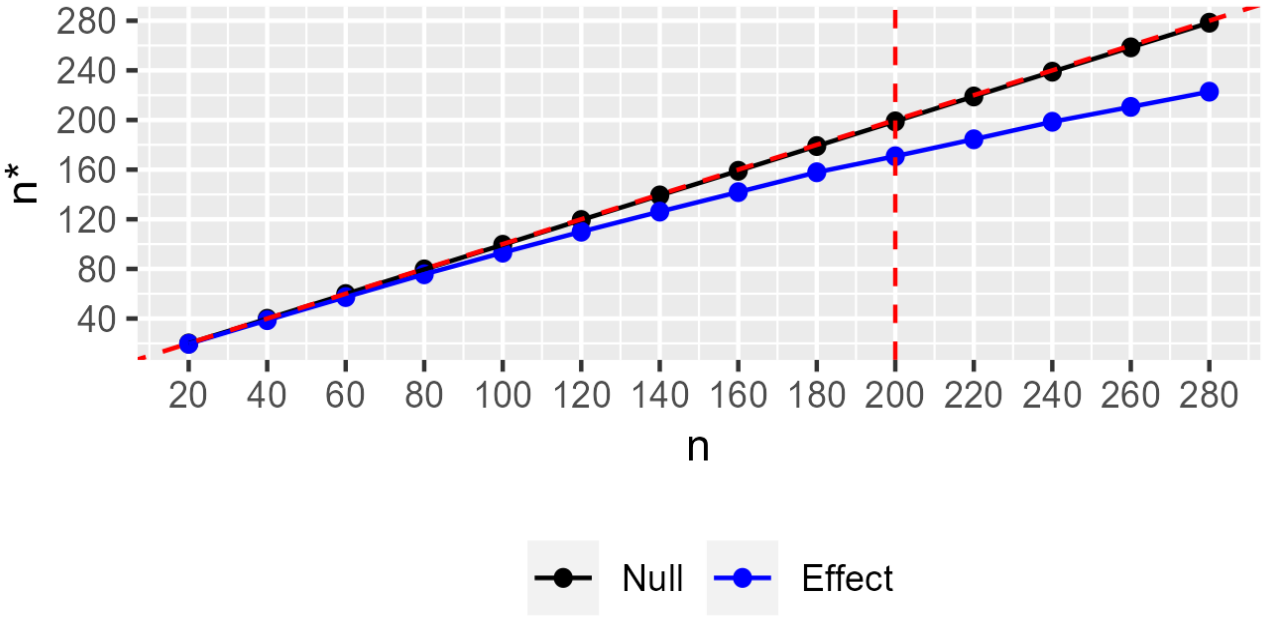
- Stop the trial
- Drop an arm
- Response adaptive randomisation
- Stop recruitment to a stratum

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Simple example

- What if we just looked at the data earlier?
  - At the halfway point check if the p-value is significant
  - Stop if it is, otherwise continue recruitment
  - Will need to “spend” alpha wisely





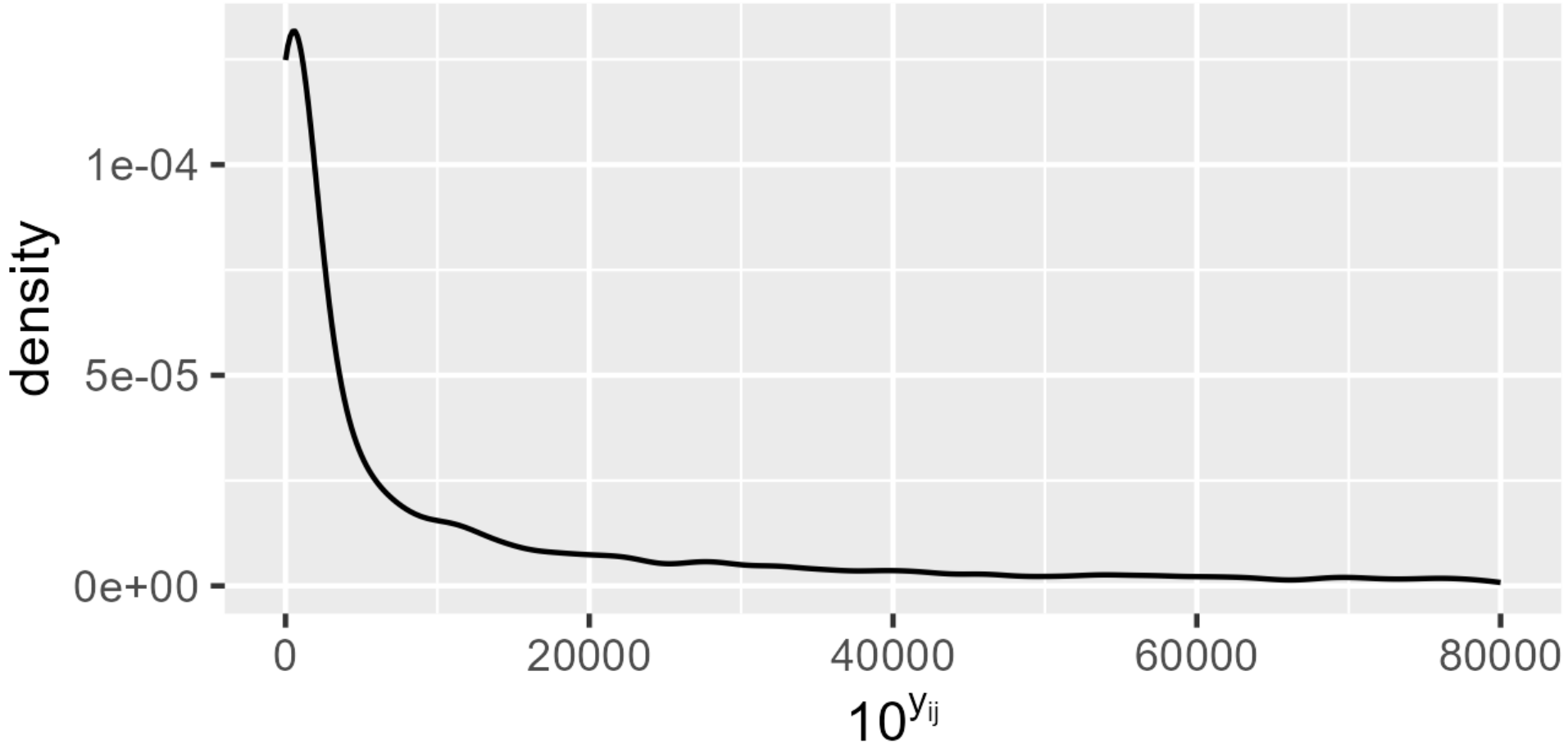
Is this cherry picking?

- For an adaptive design we **prespecify** the decision rules
  - Although decisions are made conditional on the data, the rules for decision-making were agreed beforehand
- Impact of decision rules is explored prior to trial implementation (simulations)
- We have “freedom” with decision rules but typically use “superiority” and “futility” rules

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Modelling the vaccine example

- Let's … the smuq
- Let, … eiving vacci
- Intere…



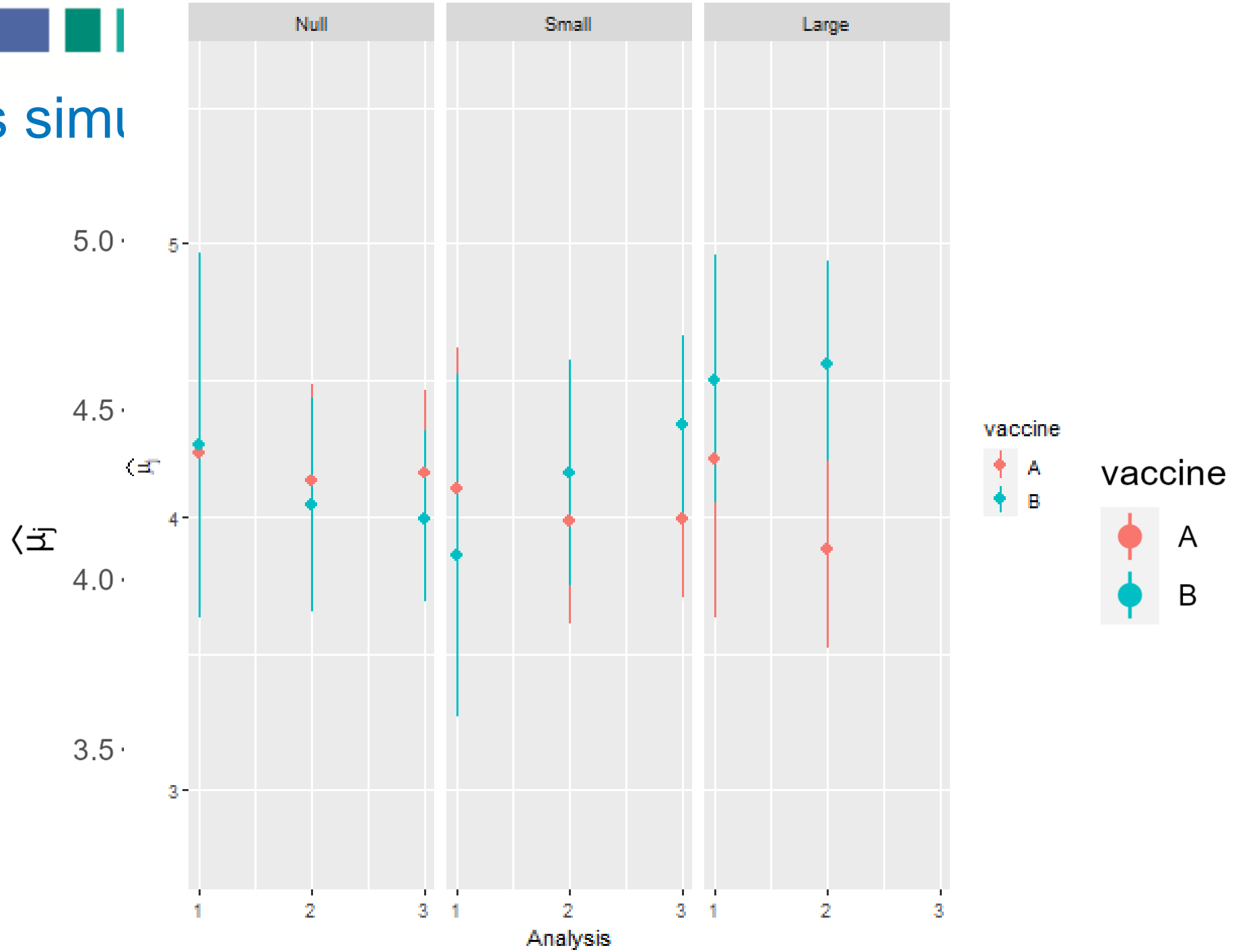
What does the trial look like?

- Suppose we plan to uniformly recruit up to 300 participants **over 1 year**
  - Randomise 1:1 to vaccine A or B
  - Schedule an analysis after every 100 participants
- Define treatment comparison (contrast) as $\theta = \mu_B - \mu_A$
- Three scenarios: Null ($\theta = 0$), Small Effect ($\theta = 0.2$) and Large Effect ($\theta = 0.5$)
- Define decision rules for each analysis based on posterior distribution of θ:
  - Stop and declare vaccine B superior if: $P(\theta > 0) > 0.95$ (or $P(\theta > 0) > 0.975$)
  - Stop and declare trial futile if: $P(\theta > 0) < 0.05$

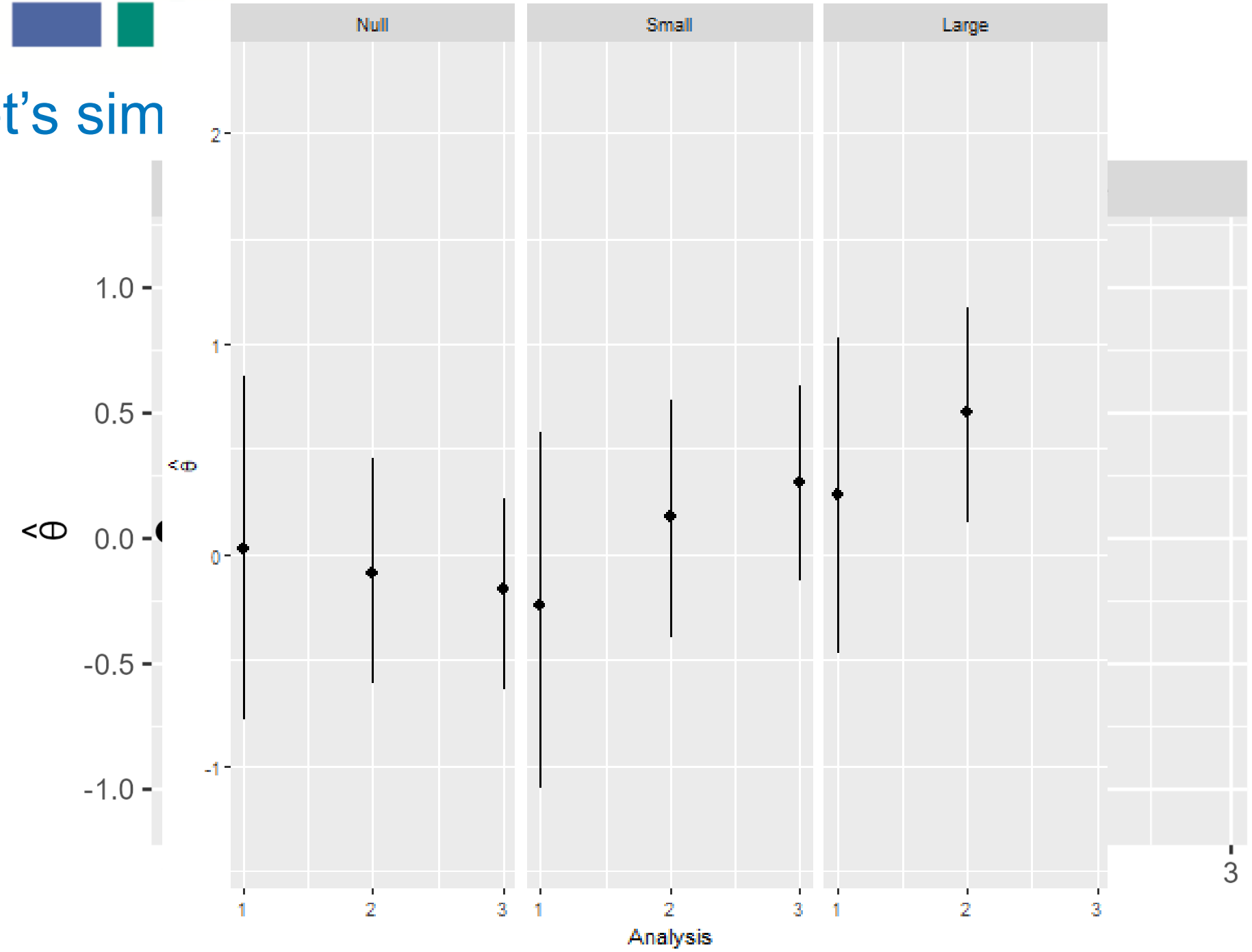
Decision rules galore

- In theory, these decision rules can be anything that we can compute
- We could use a different threshold at each analysis (similar to alpha spending)
  - $P(\theta > 0) > 0.995$ at early analyses
  - $P(\theta > 0) > 0.94$ at the final analysis
- We could directly compare to a “clinically important difference"
  - $P(\theta > 0.5) > 0.9$
- We could define a decision rule based on another quantity entirely
  - $P\left(\frac{\mu_B - \mu_A}{\mu_A} > 0\right) > 0.95$

Let's simu



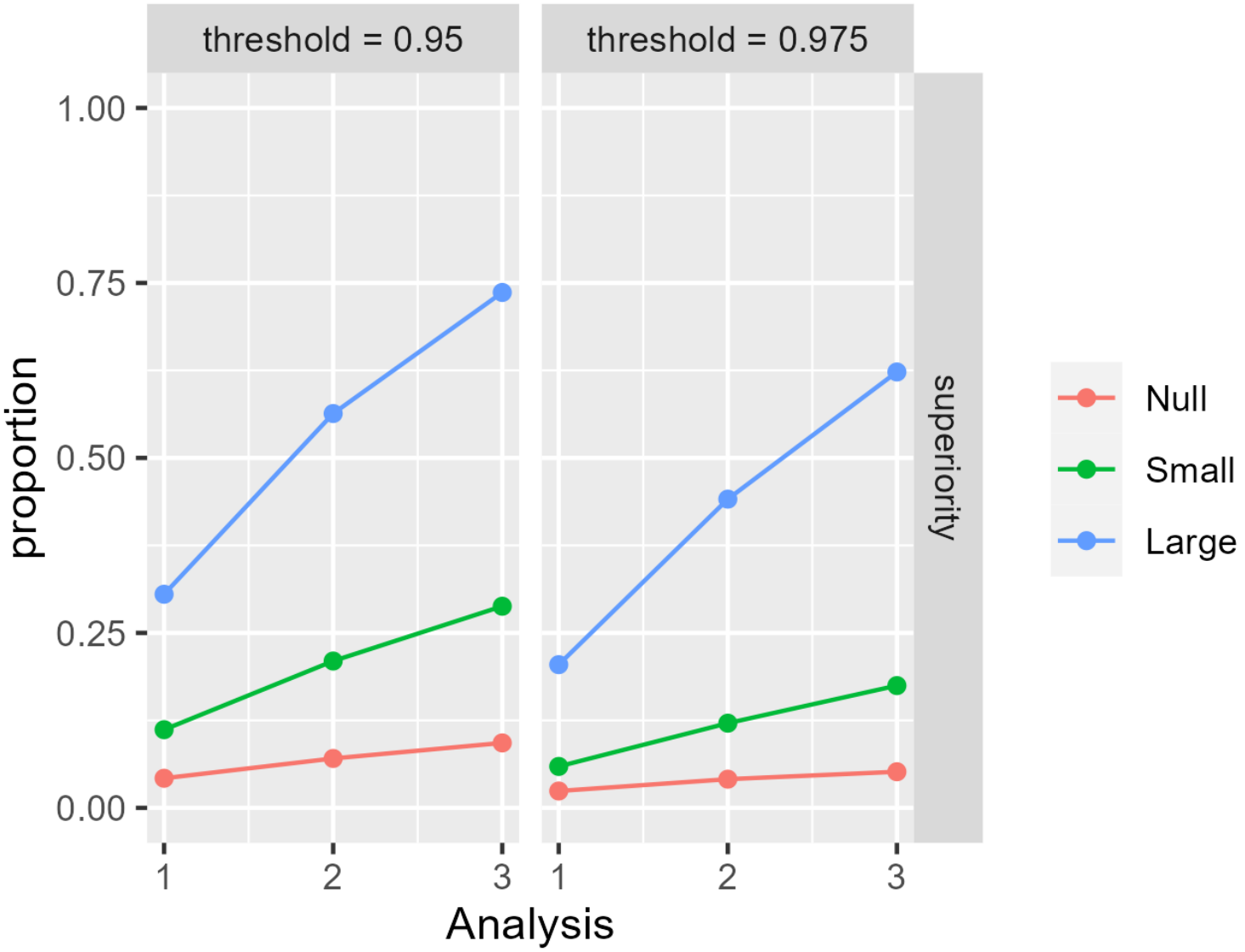
Let's sim



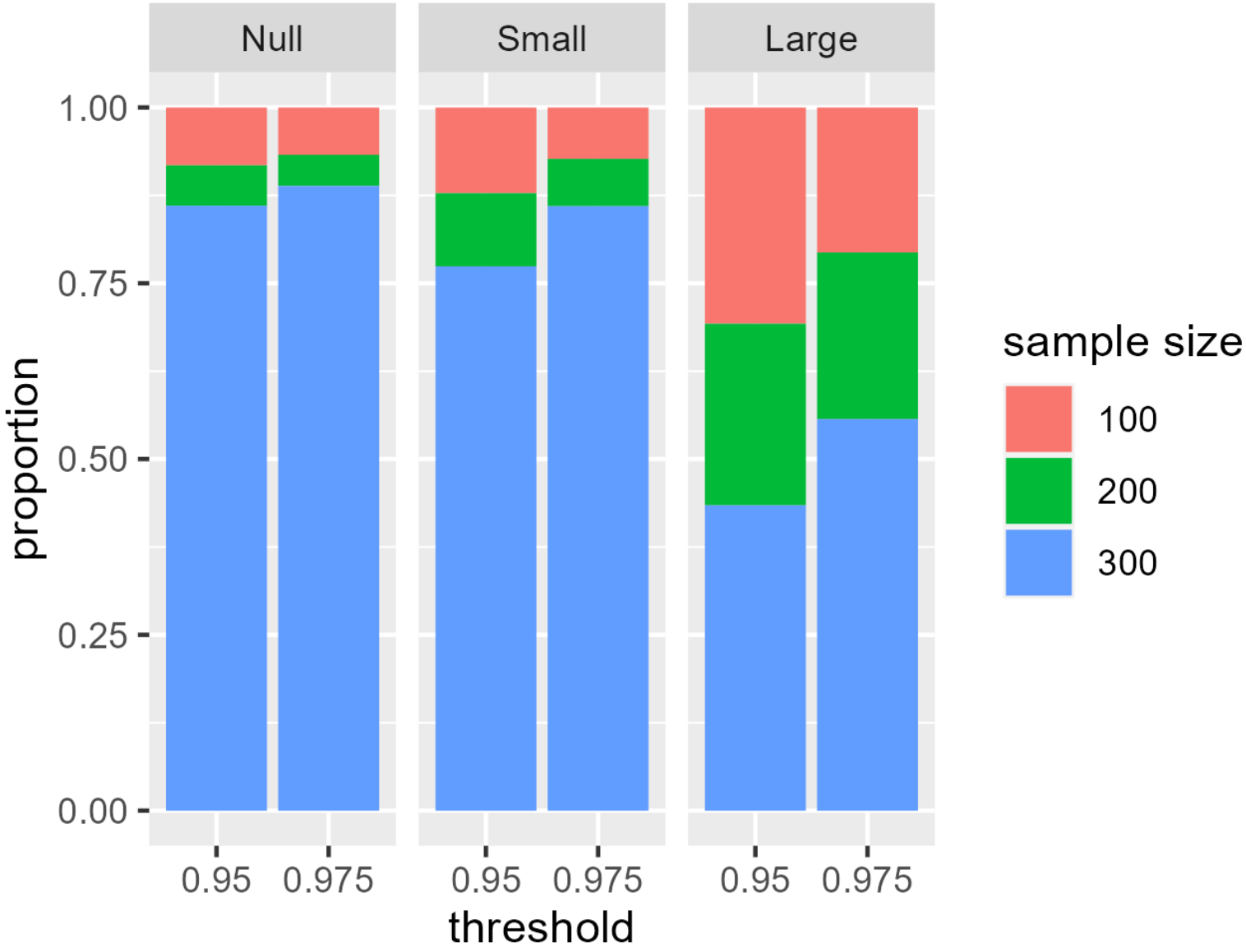
Let's simulate some trials

Scenario	Threshold	Proportion of trials declaring superiority	Mean Sample Size
Null	0.95	0.093	278
	0.975	0.052	282
Small Effect	0.95	0.288	265
	0.975	0.175	279
Large Effect	0.95	0.736	213
	0.975	0.623	235

Let’s simulate some trials



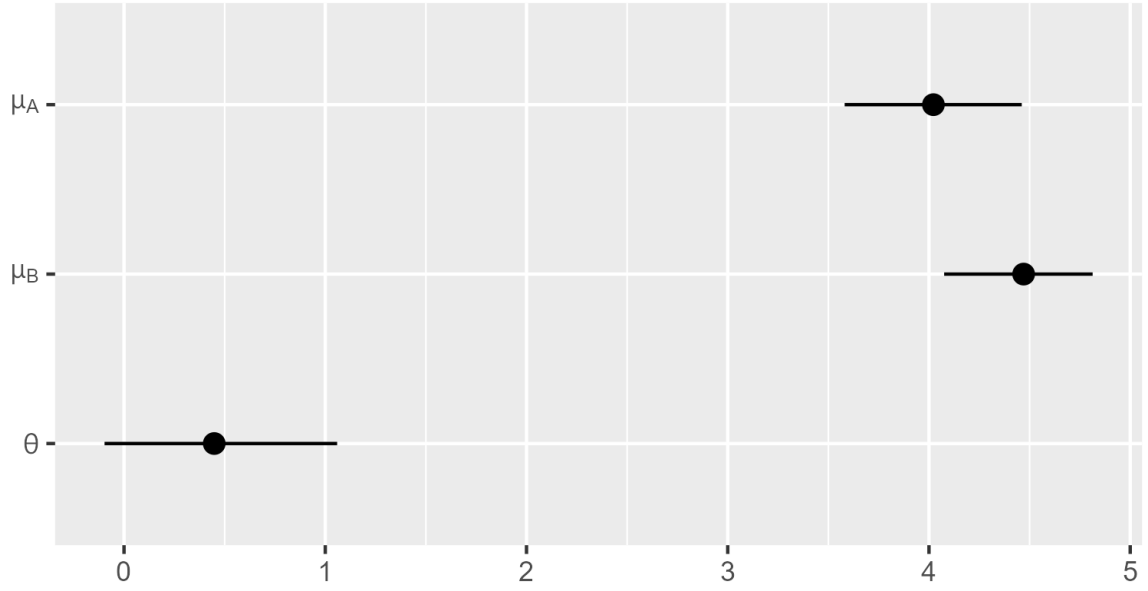
Let’s simulate some trials



How should we report our estimates?

- Look at one analysis with 200 participants ($\mu_A = 4, \mu_B = 4.5, \sigma_A = \sigma_B = 2$)

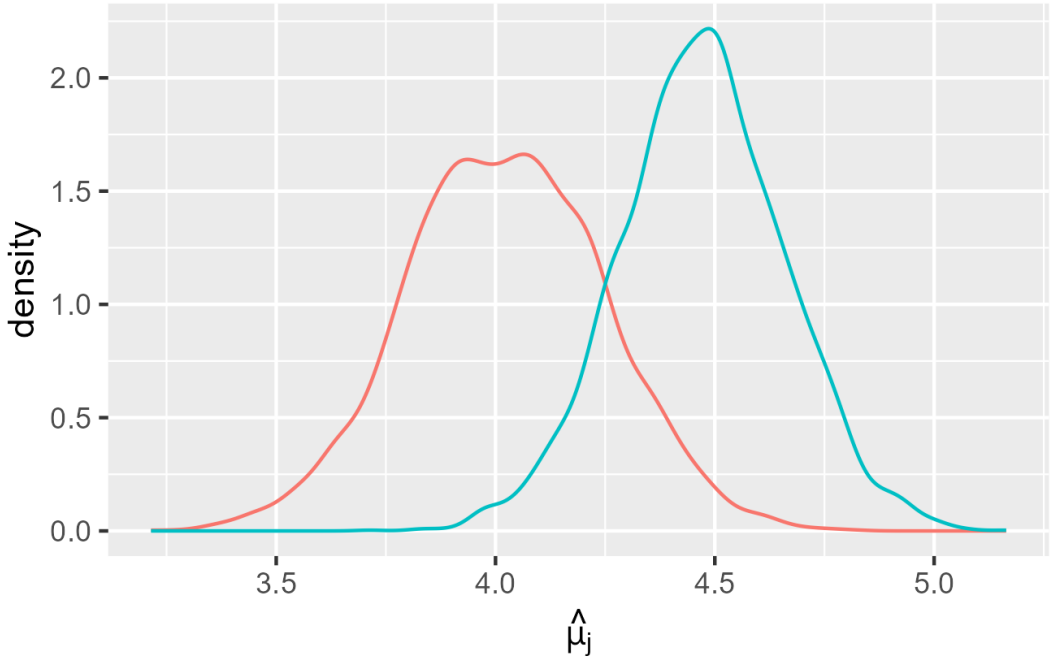
Parameter	Mean	95% HDI Interval
μ_A	4.02	(3.58, 4.46)
μ_B	4.47	(4.08, 4.81)
θ	0.45	(-0.10, 1.06)

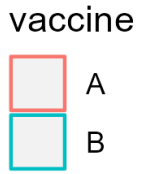


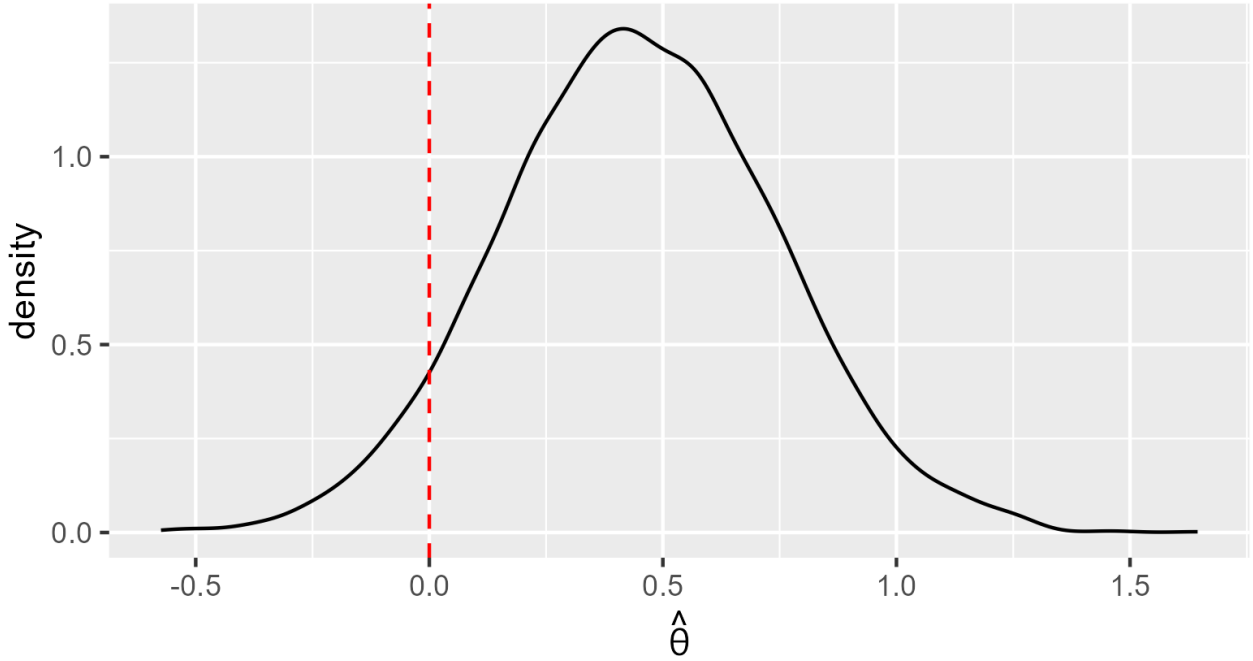
- Decision: $P(\theta > 0) = 0.94$
  - Do not stop for superiority or futility

How should we report our estimates?

- How about visualising the full densities?

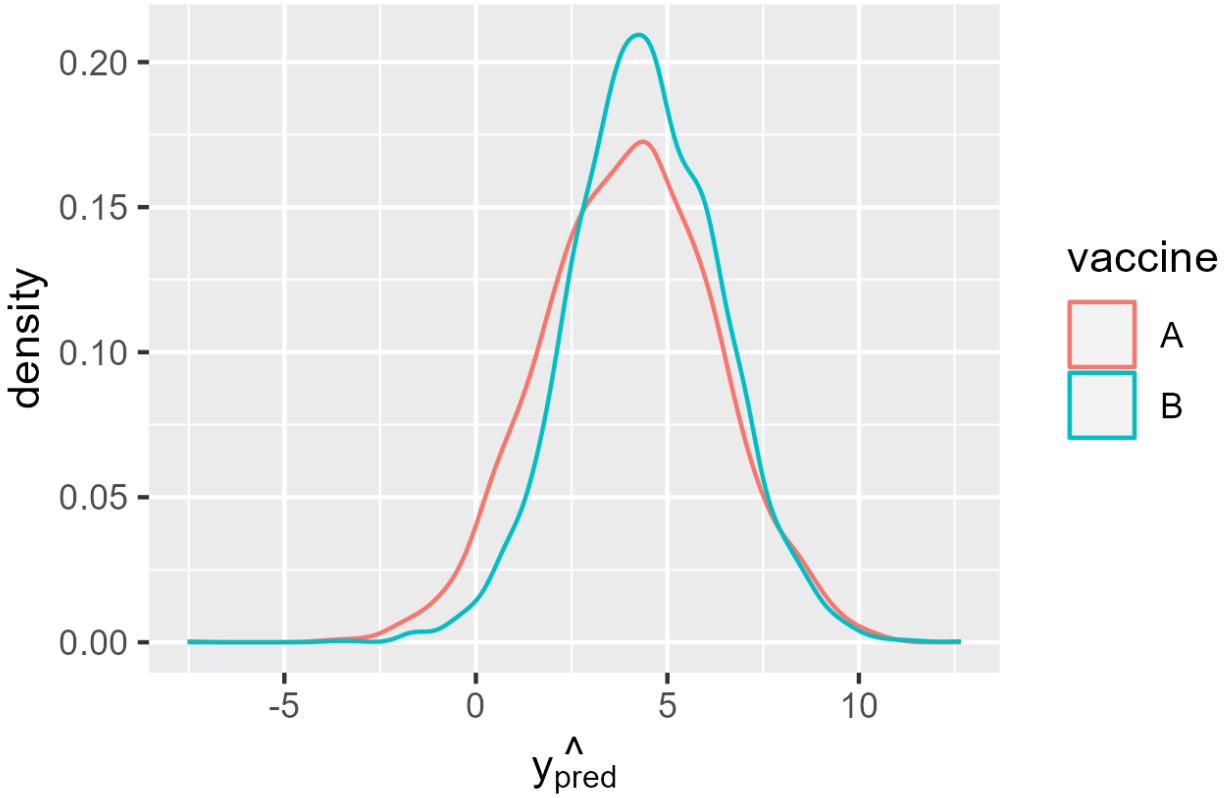


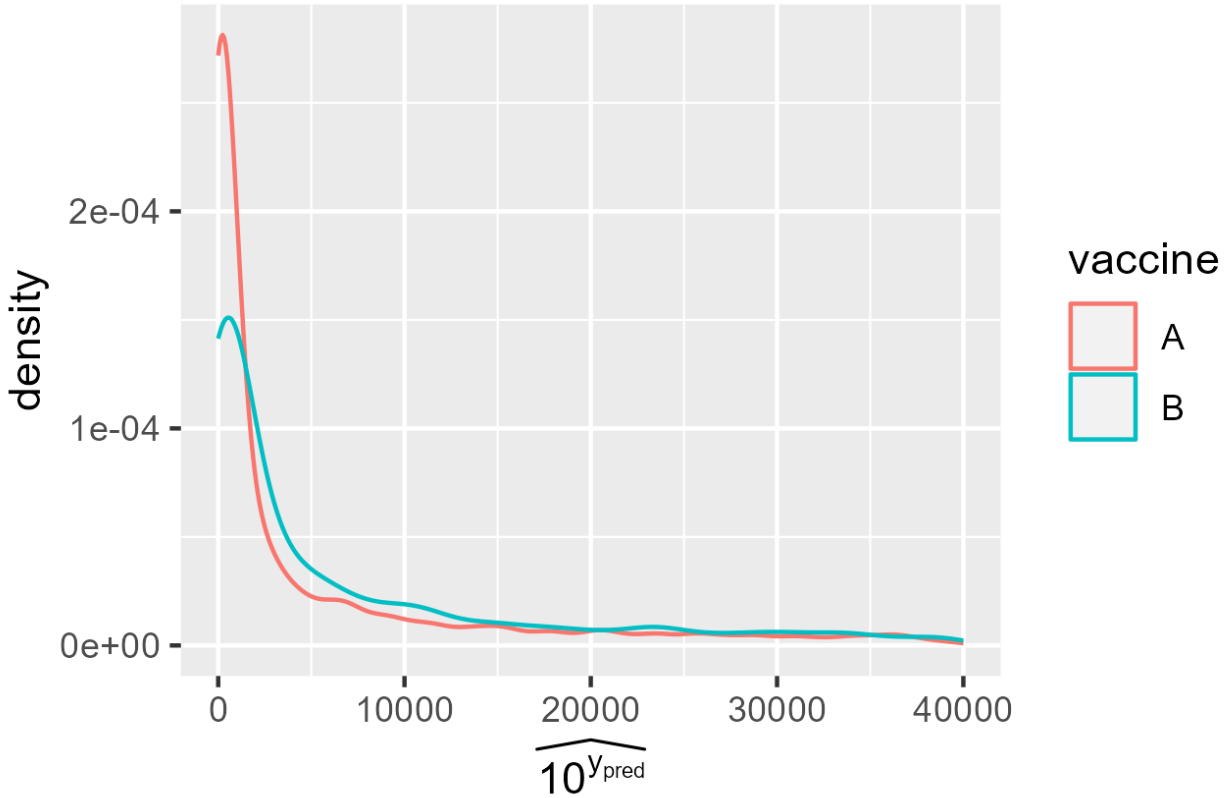




How should we report our estimates?

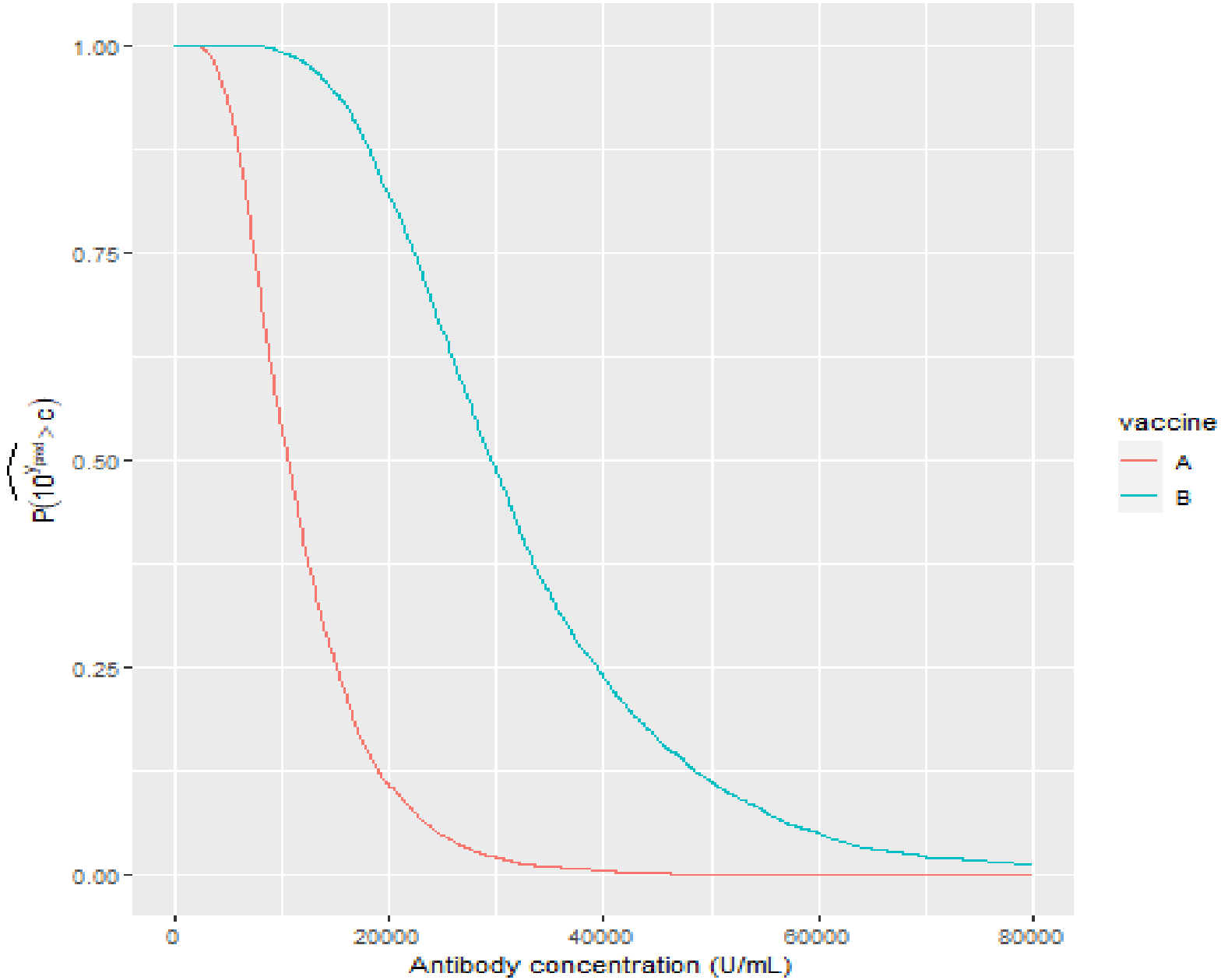
- How about visualising at the individual level?





How s

- Is there a



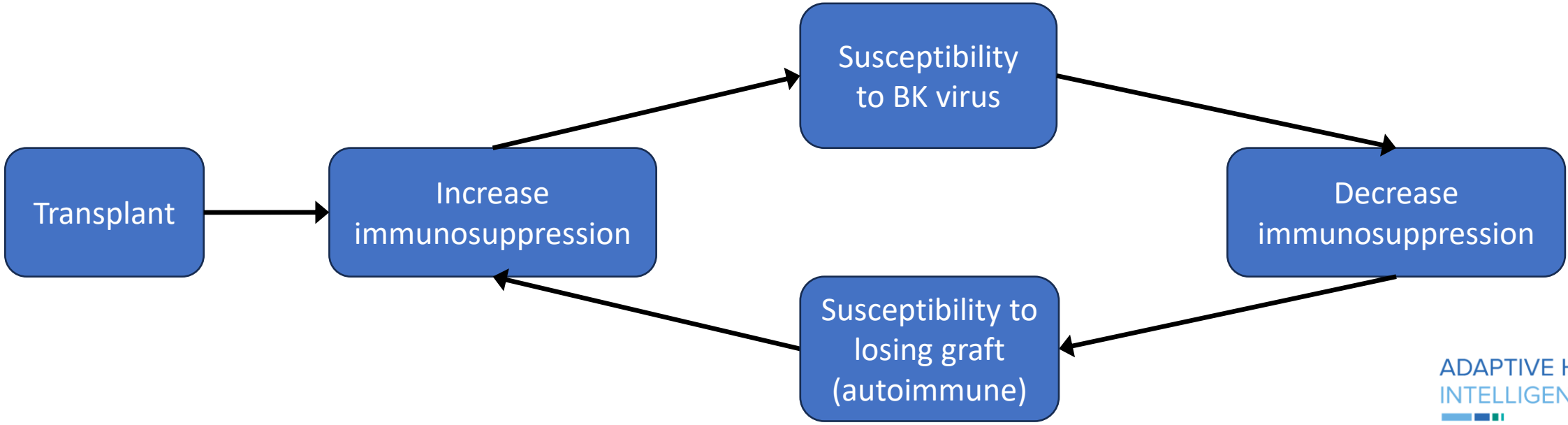
Some real examples – Healthy Ears

- Otitis media (middle ear infections) in children aged 4-7 years
- Interventions: Hygiene advice vs hygiene advice + blow breath cough technique
- Outcome: Resolution of infection at 4 weeks (binary)
- Decision rules: Stop early for superiority or futility

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Some real examples – BEAT-BK

- BK viraemia in kidney and kidney + pancreas transplant recipients
- Interventions: Control vs IVIG (blood product)
- Outcome: Five category rank on condition at 12 weeks (ordinal)
- Decision rules: Stop early for superiority or futility



ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Some real examples – PICOBOO

- COVID-19 vaccinations in immuno**competent** participants (3rd dose – 5th dose)
- Interventions: Pfizer, Moderna, Novavax, …
- Outcome: log10 antibody concentration at 28 days
- Modelling: Hierarchical model (partial pooling) over strata, age etc.
- Decision rules: Stop recruitment to a stratum if adequate precision

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

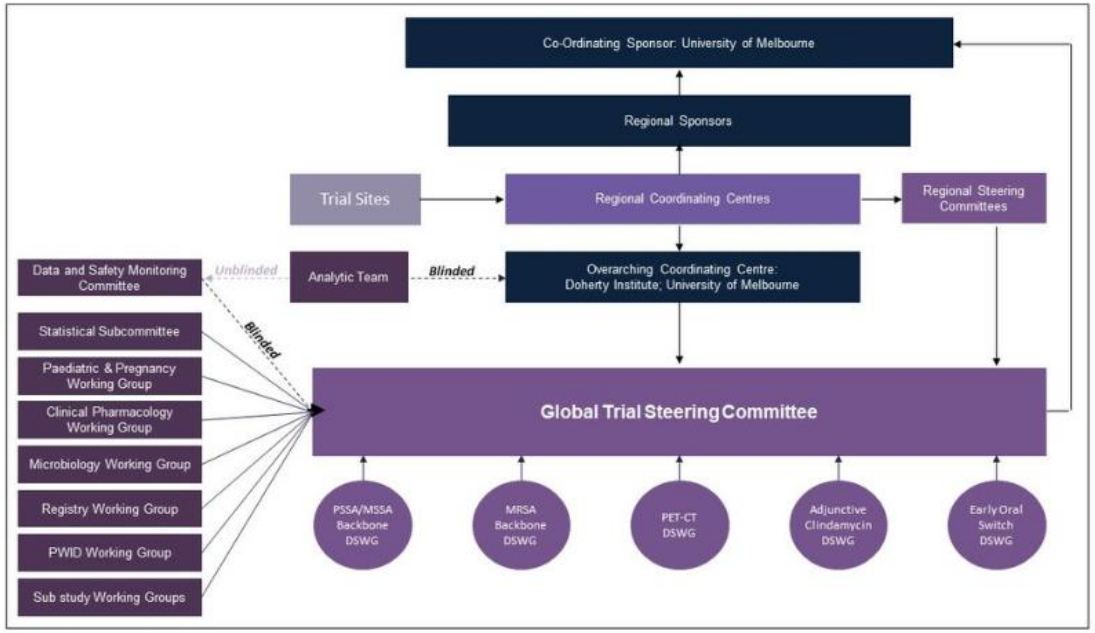
Some real examples – BOOST-IC

- COVID-19 vaccinations in immuno**compromised** participants (4th dose – 8th dose)
- Interventions: Pfizer, Moderna, Novavax, … but one or two doses
- Outcome: log10 antibody concentration at 28 days
- Decision rules: Stop recruitment to a stratum if adequate precision

ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Some real examples – SNAP

- Platform trial for treatments of staphylococcus aureas
- Domains: Backbone, adjunctive, early oral switch
- Silos: Resistance of strain to classes of antibiotics
- Interventions: Different antibiotics, timing of switch
- Outcome: Mortality at 90 days (binary)
- Decision rules: Superiority and non-inferiority dependent on domain x silo



ADAPTIVE HEALTH INTELLIGENCE
EVIDENCE IN ACTION

Some final thoughts

- Adaptive designs allow for the accrued data to inform the design
- There are drawbacks
  - Risk of biased estimates (e.g., stopping at a random high)
  - Harder to design and implement (more resources)
  - Not suitable for all clinical trials (e.g., fast recruitment with slow endpoint)
- But there are advantages
  - Resource efficiency
  - Answers to scientific questions faster (translation to policy)

Some final (open) questions for future research

Publicly funded research is intended to improve the *health* of the (future) population.

- What (ethical) responsibility do we have as publicly funded researchers to design our studies with improving *health* as the objective?
- How can we design clinical studies to inform the decision-making of clinicians and consumers?
- How can we report our results to best inform decision-making?
- How should we handle multiple (possibly competing) endpoints (e.g., efficacy vs safety)?
- How can we implement our research to drive policy? (Instead of back filling the evidence)