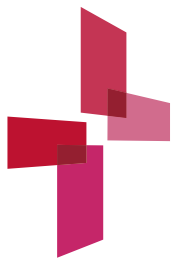


Designing a Platform Trial

Dr Julie Marsh

7 November 2022



Australian
Clinical
Trials
Alliance

www.clinicaltrialsalliance.org.au

ACTA gratefully acknowledges operational funding from the Australian Government's Medical Research Future Fund

Outline

Adaptive designs: common features

Platform definitions

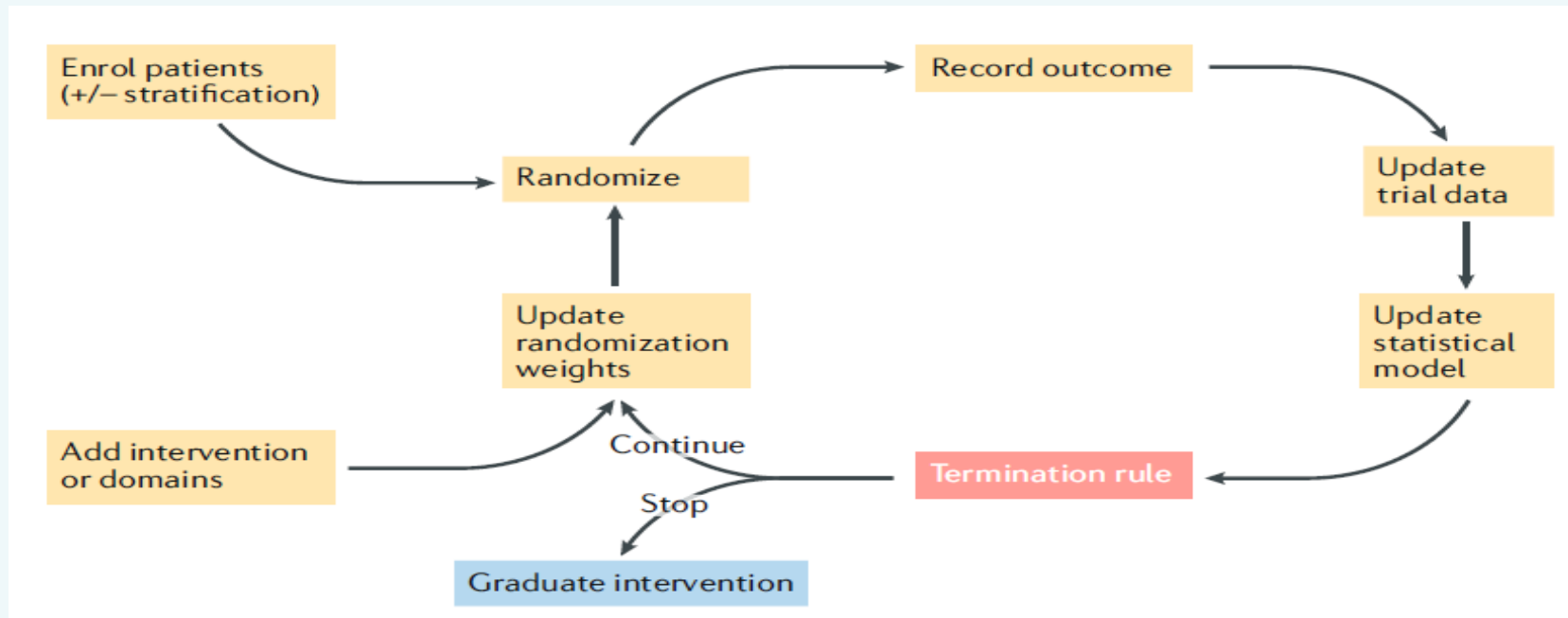
Decision criteria for adaptations

Role of simulation in design

Platform design criticisms

Adaptive designs: common features

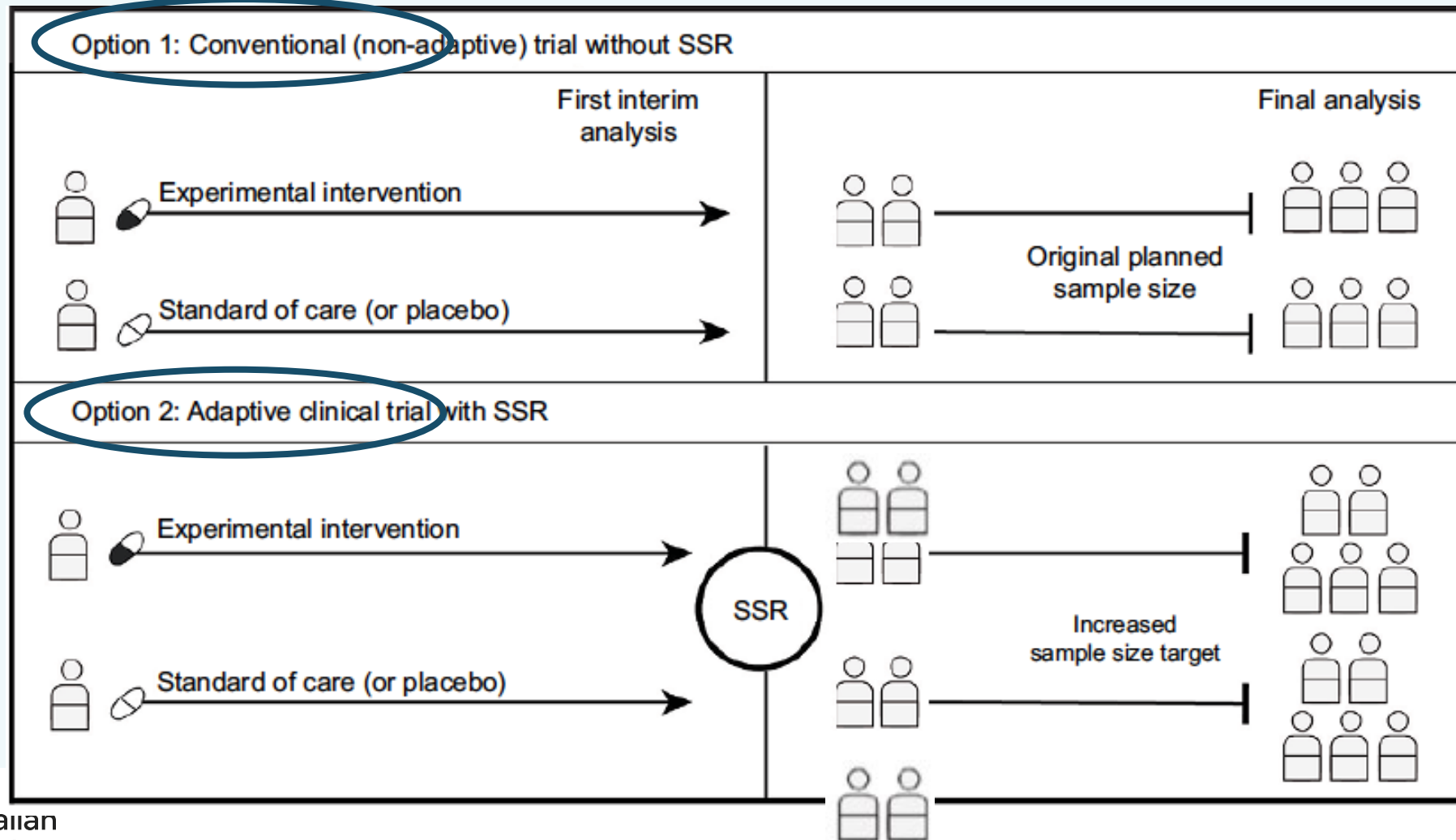
Outcomes are repeatedly assessed on accumulating data
Study design may be modified based on pre-specified rules



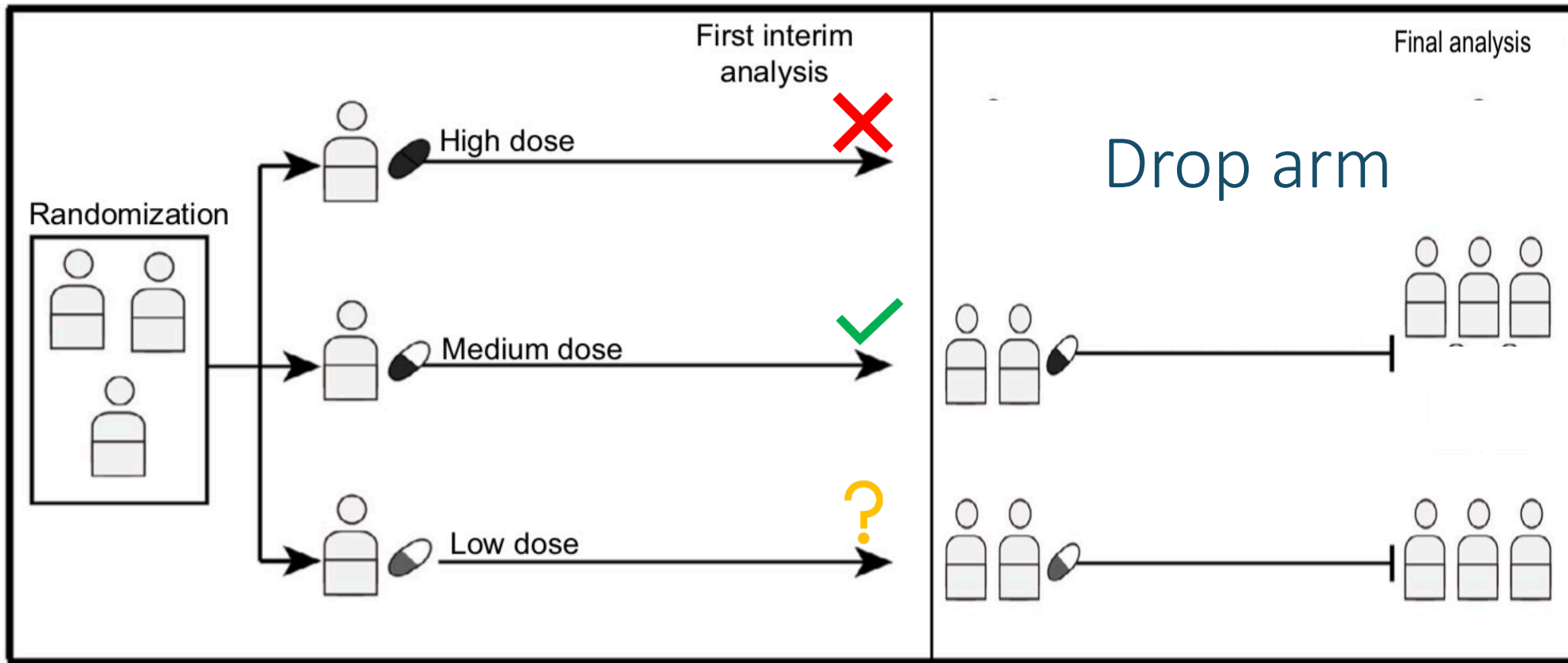
Adaptive designs: common features

- Sample size reassessment: *insurance policy*
- Treatment selection (arm dropping): *promising candidates*
- Seamless (combined data over stages/phases): *economical/time saving*
- Response adaptive randomisation: *patient-centric/lower exposure to ineffective treatments*
- Enrichment (population dropping): *promising populations*
- *Platform trial: multiple treatments & populations evaluated simultaneously/efficiency*

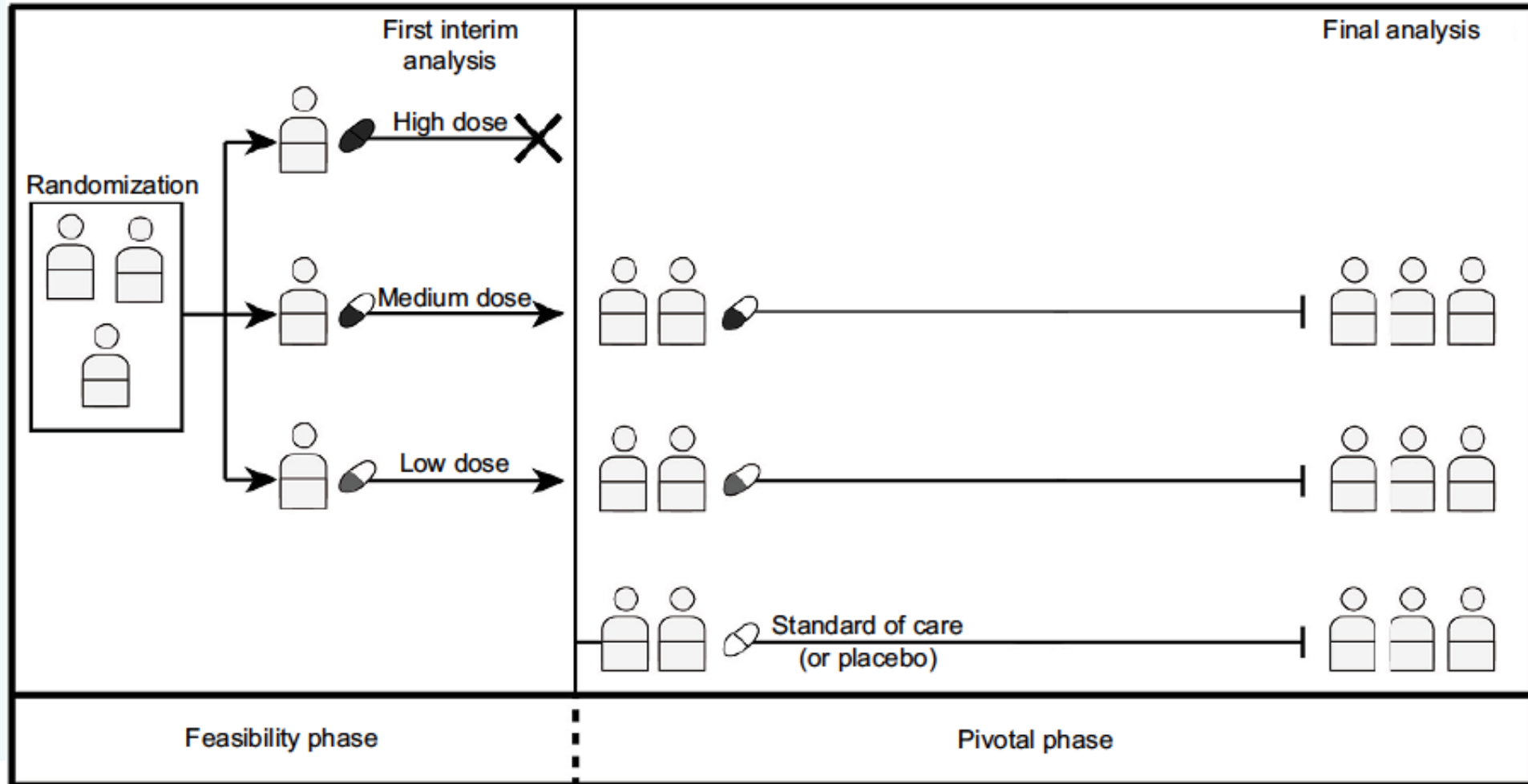
Sample size reassessment (SSR)



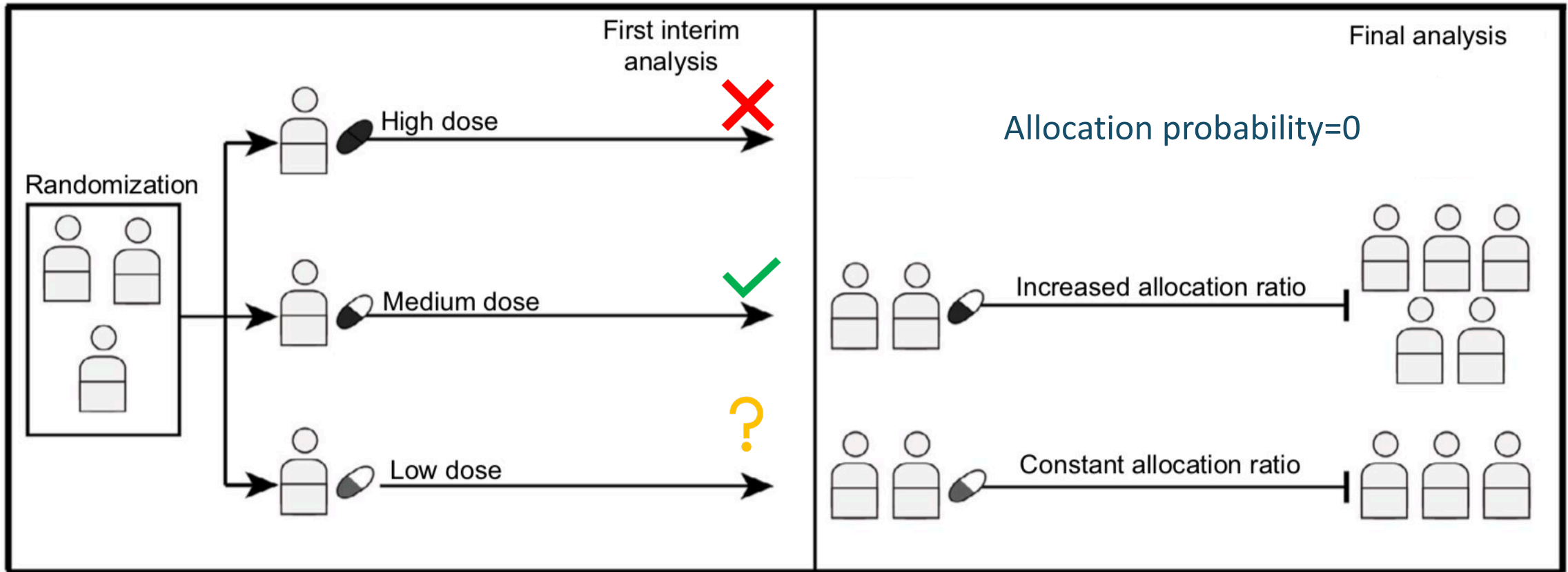
Treatment selection (arm dropping)



Seamless design



Response adaptive randomisation



Response-Adaptive Randomisation

RAR requires a short time to outcome relative to the overall recruitment period

- Skews intervention allocation probabilities away from equal allocation over time
- Favours the better performing interventions based on accrued data, *without undermining the validity and integrity of the trial*
- Allocation probabilities adapted by pre-specified rules

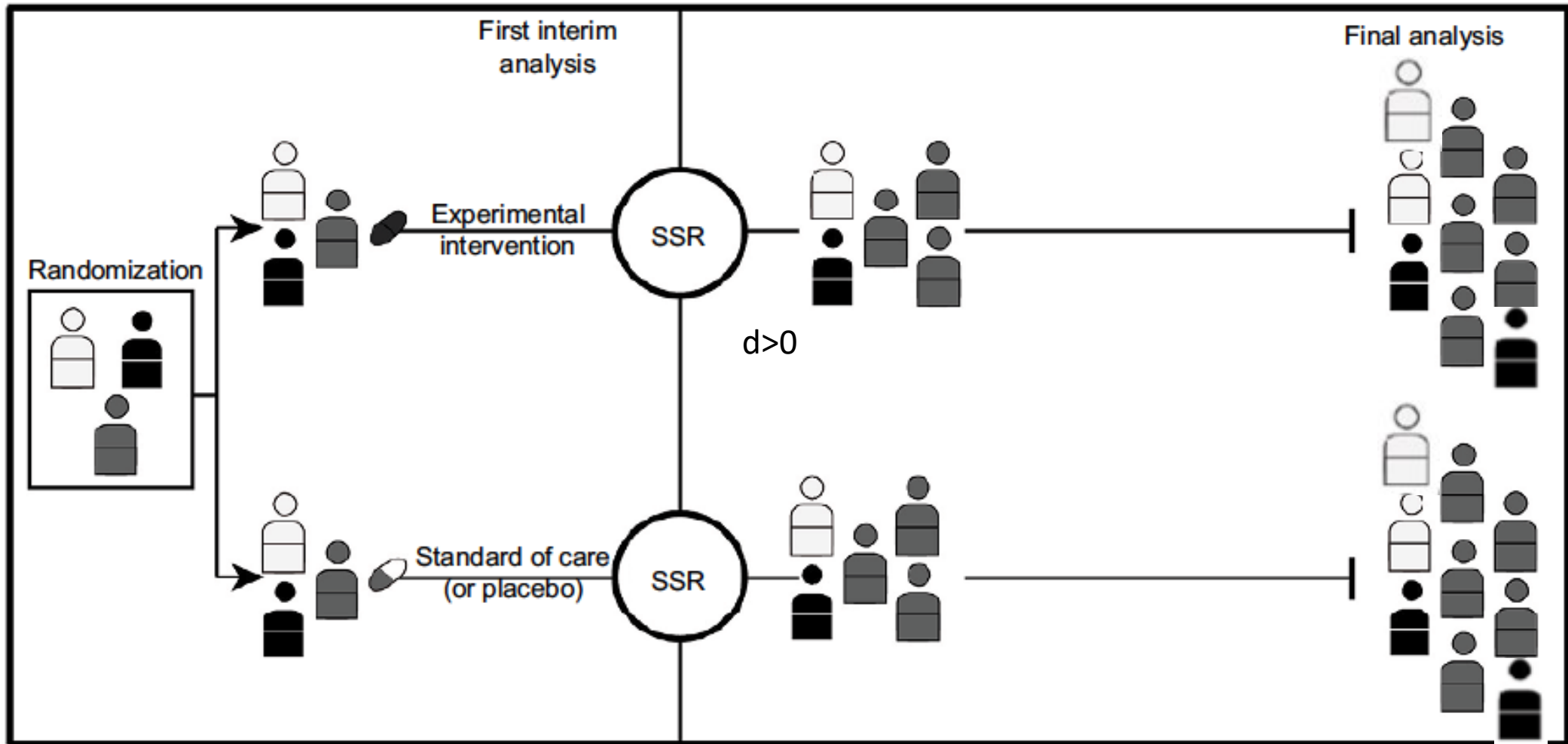
Response-Adaptive Randomisation

If by chance, **higher B** cure rate than true value: increased allocation to treatment B, therefore, **faster return to true value**

If by chance, **lower A** cure rate than true value: reduced allocation to treatment A, therefore, **slower return to true value**

In simulations, RAR algorithms tend to slightly under-estimate true treatment values (on average)

Enrichment design



More complex adaptive designs

Umbrella trial

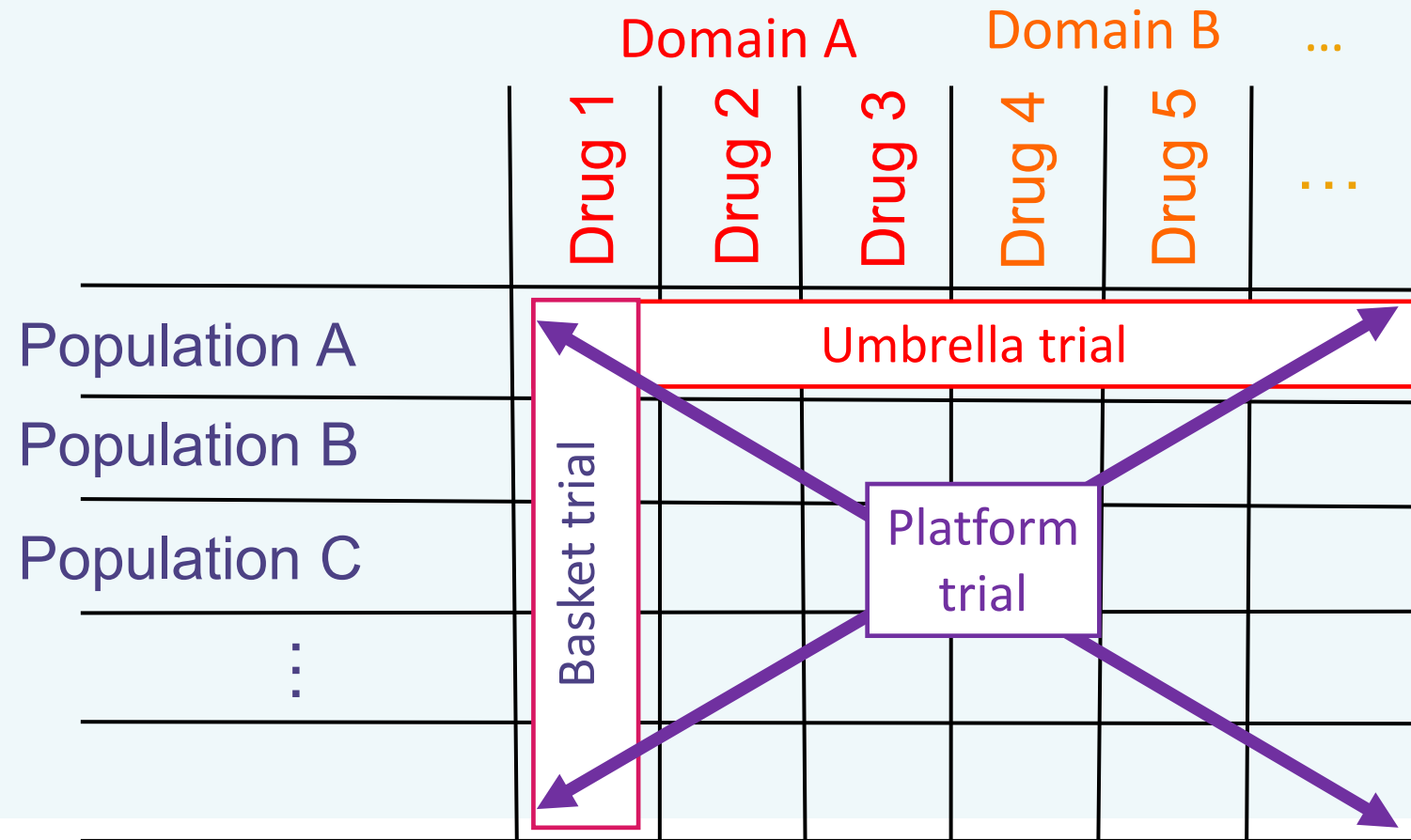
One population, many drugs

Basket trial

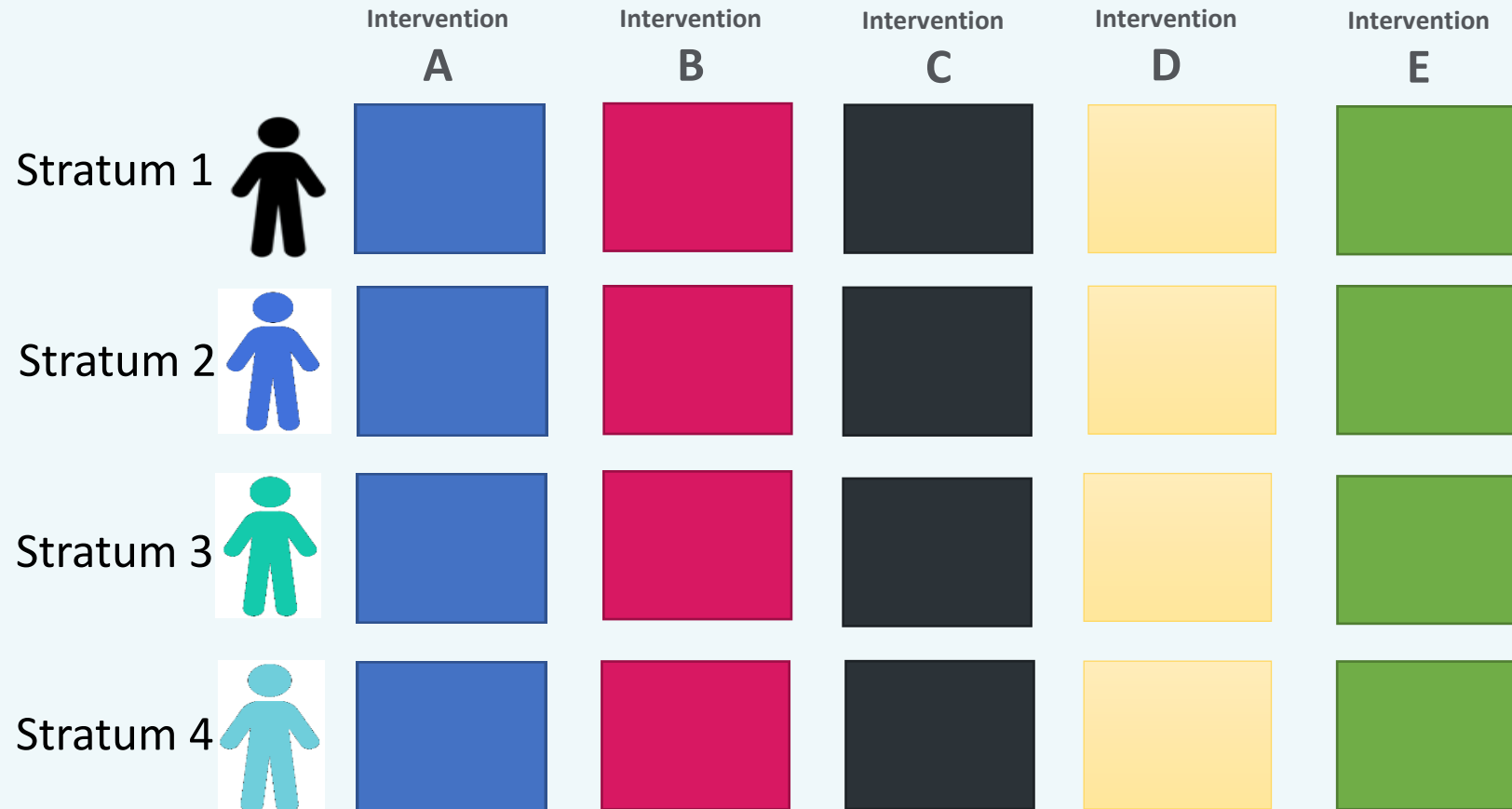
Many populations, one drug

Platform trial

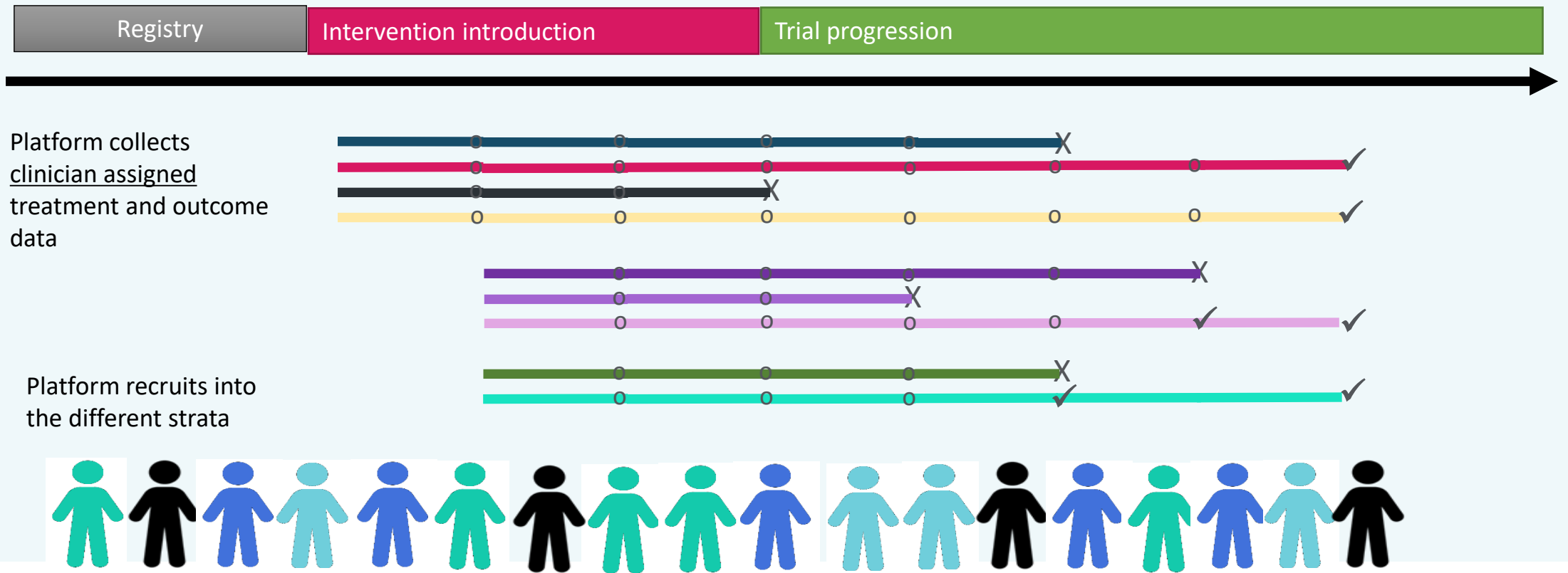
Many populations, many drugs



Platform designs



Platform designs



Platform definitions

- **Strata:** mutually exclusive and exhaustive categories, defined by baseline characteristics of a participant within the platform, for which effects of interventions may be differential. Should be reflected in randomisation and statistical model. E.g. age cohort, severity of disease at enrolment.
- **Domain:** mutually exclusive and competing interventions comparing a common clinical mode of action or clinical context of use. E.g, types of statins or cephalosporins.

Platform definitions

- **State:** mutually exclusive and exhaustive categories, defined by characteristics of a participant within a platform, that are capable of changing over time (i.e. they can be dynamic). Used to define eligibility for domains that occur after the time of enrolment and included as a model covariate.
- **Stage:** indicates a domain that has reached a conclusion and then starts again with a different research question or estimand, such as a change in the target population or a different endpoint.

Trial integrity

Model complexity

Availability &
quality of data

Adaptive
randomisation

Model information
sharing

Contemporaneous
control group

Trial simulation

Decision Criteria

Data office

Decision criteria for adaptations



Frequentist	Bayesian
Focus on null hypothesis	Focus on alternative hypotheses
Probability of data	Probability of hypothesis
Analytical focus	Computational focus
Less flexible	More flexible
Poorly suited to sequential inference	Sequential inference a breeze
More familiar	Less familiar
Requires specialised software	Requires specialised software

Ease of decision making



Decision criteria for adaptations

Consider a trial with two arms: Treatment A and Treatment B

Objective: Is treatment B **superior** to A?

p_A : proportion cured on treatment A

p_B : proportion cured on treatment B

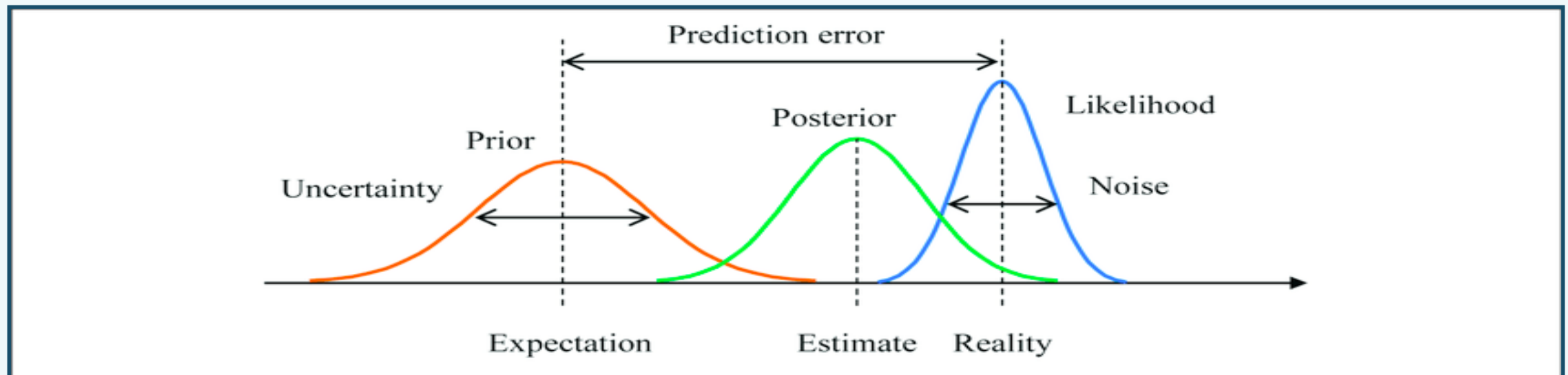
Where difference, $d = p_B - p_A$

Null hypothesis $H_0: d \leq 0$

Alternate hypothesis $H_1: d > 0$ (B superior if cure rate higher than A)

Decision criteria for adaptations

We perform **repeat analyses** as the **data accumulates** and **evaluate pre-specified decision criterion**. At each interim:



Decision criteria for adaptations

Step 1: Estimate probability of alternate hypotheses given current data:

Superiority: probability B has a higher cure rate than A, $pr(H_1)=pr(d>0)$

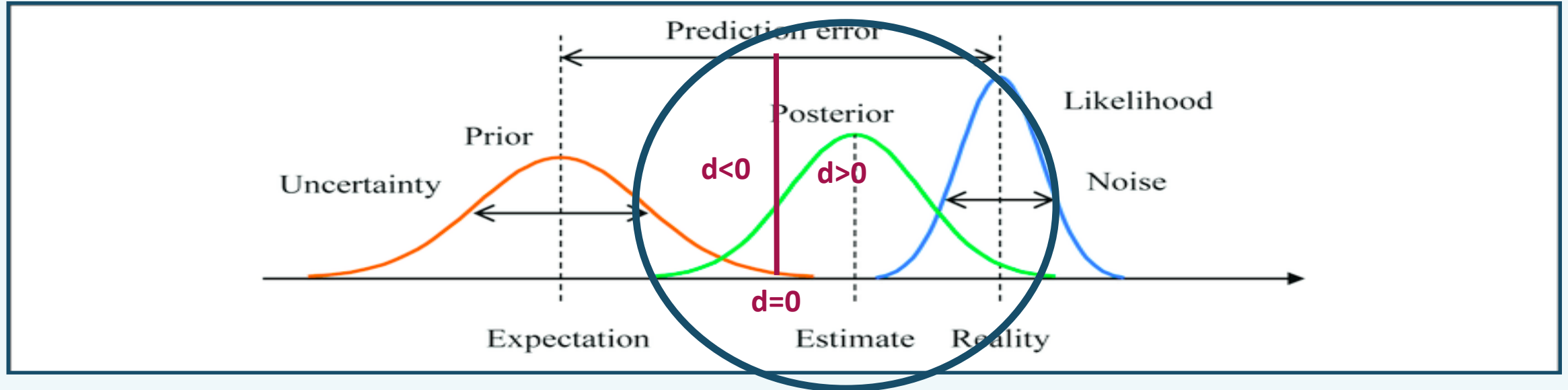
Step 2: Compare this probability to pre-defined **probability threshold** and follow protocol defined decision rules, e.g.

Superiority: if the probability that B has a higher cure rate than A is greater than 95% [$pr(d>0)>0.95$], then declare B superior and stop recruitment early

Futility: if the probability that B has a higher cure rate than A is less than 10% [$pr(d>0)<0.10$], then declare trial futile and stop recruitment early

Decision criteria for adaptations

Illustration of Step 2



Decision criteria for adaptations

Step 3: Implement decision rule in ongoing trial, *e.g.*

If **superiority** or **futility** threshold met → **STOP** recruitment early & report

Otherwise → **CONTINUE** recruitment to next interim or maximum recruitment

Note that decision criteria using Bayesian methods may be based on:

1. Data on participants who have reached endpoint: $\text{pr}(d>0 | \text{current data})$
2. All recruited participants, some yet to reach endpoint: $\text{pr}(d>0 | \text{current data \& missing data})$
3. All recruited participants plus future participants: $\text{pr}(d>0 | \text{current data \& missing data \& future data})$

Role of simulation

To determine optimum design parameters over a range of clinical scenarios that control the false-positive error, yield appropriate power and minimise participant exposure to harm.

The design parameters may include:

- Decision quantities (may be p-values, posterior probabilities, etc.)
- Decision threshold(s)
- Number and timing of interims
- Method of updating treatment allocation probabilities

Role of simulation

This is an iterative process between the investigators and the statistical team.

Inputs:

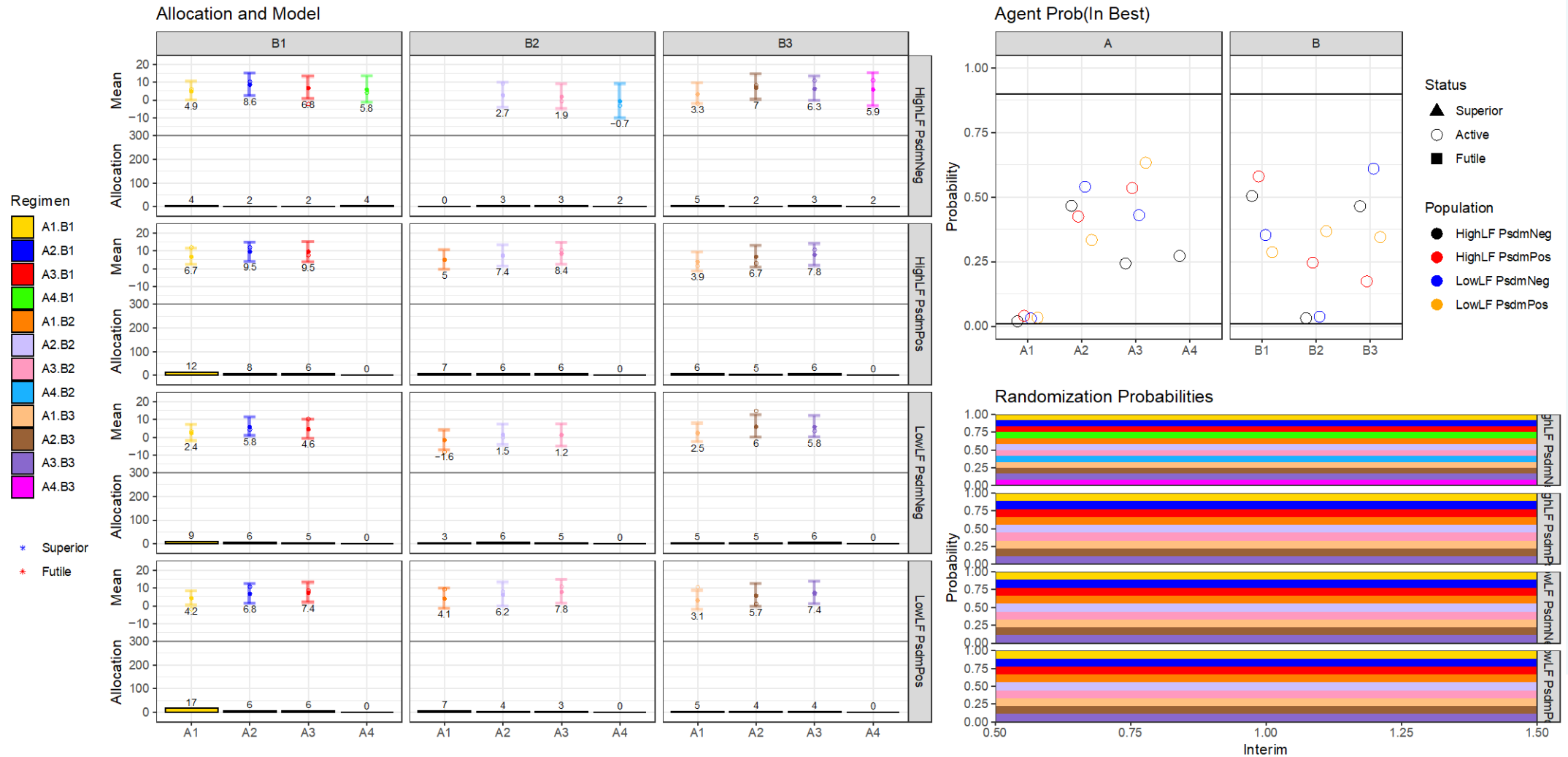
- Primary endpoint, include the time of assessment relative to randomisation
- Number of strata (and definition)
- Recruitment rate into each stratum
- Maximum sample size
- Number of domains and interventions within each domain
- Range of plausible and some more extreme clinical scenarios, by domain & stratum
- Statistical model
- Statistical distribution of primary endpoint and any prior information/publications
- Starting point for timing and number of interim analyses
- Starting point for decision criteria – superiority/non-inferiority/futility and decision quantities

Role of simulation: example clinical scenarios

Simulation Scenario	Simulated Change in Endpoint		Description of scenario
	Domain A	Domain B B1 / B2 / B3	
1	A1	4 / 4 / 4	Null scenario. Each Regimen associated with a 4.0 increase
	A2	4 / 4 / 4	
	A3	4 / 4 / 4	
	A4	4 / 4 / 4	
2	A1	4 / 4 / 4	Moderate increase for A2 in all Strata (8.0 vs 4.0)
	A2	8 / 8 / 8	
	A3	4 / 4 / 4	
	A4	4 / 4 / 4	
3	A1	4 / 8 / 8	Moderate increase for B2 and B3 Strata (8.0 vs 4.0)
	A2	4 / 8 / 8	
	A3	4 / 8 / 8	
	A4	4 / 8 / 8	

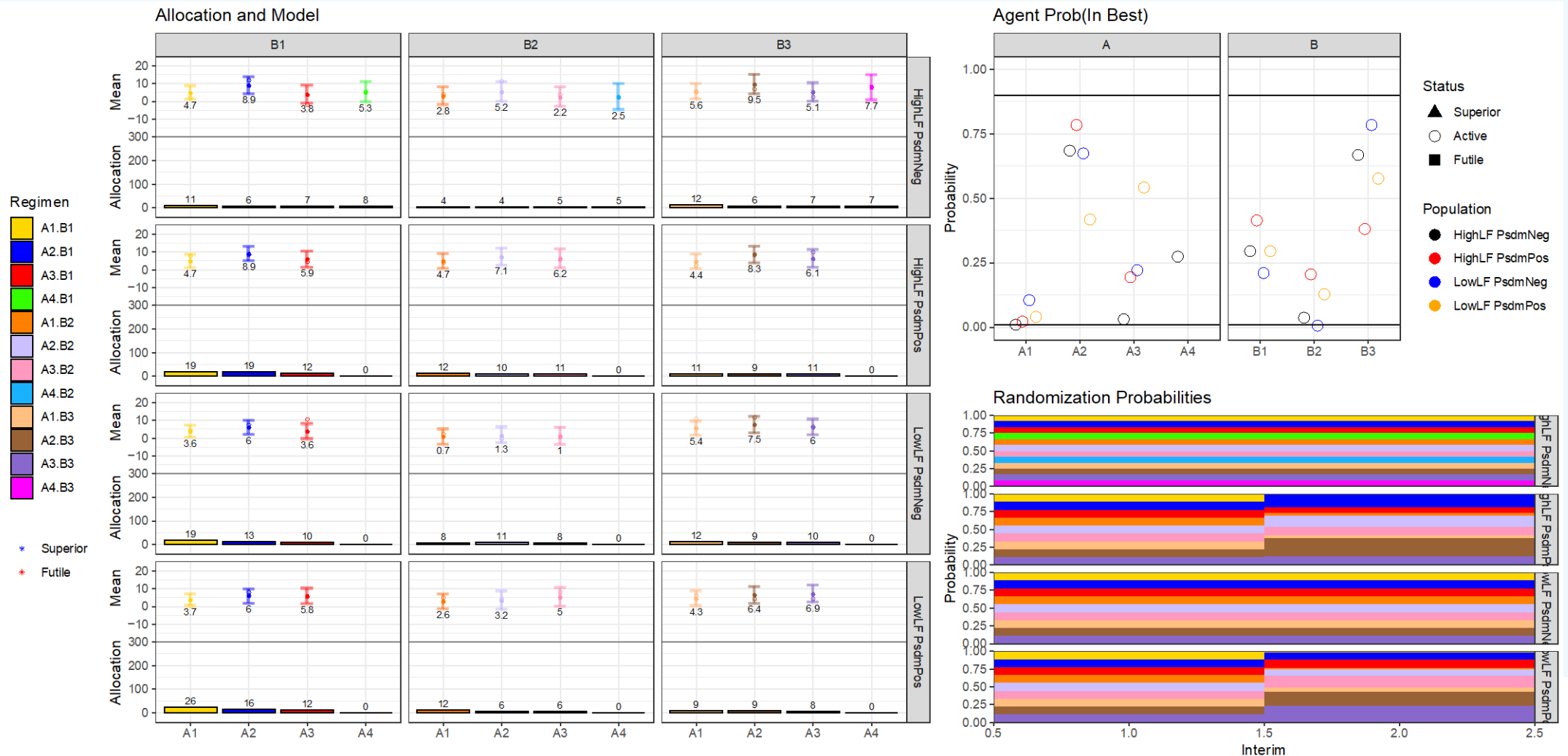
Example simulated trial: Interim 1

Look #1: N = 200



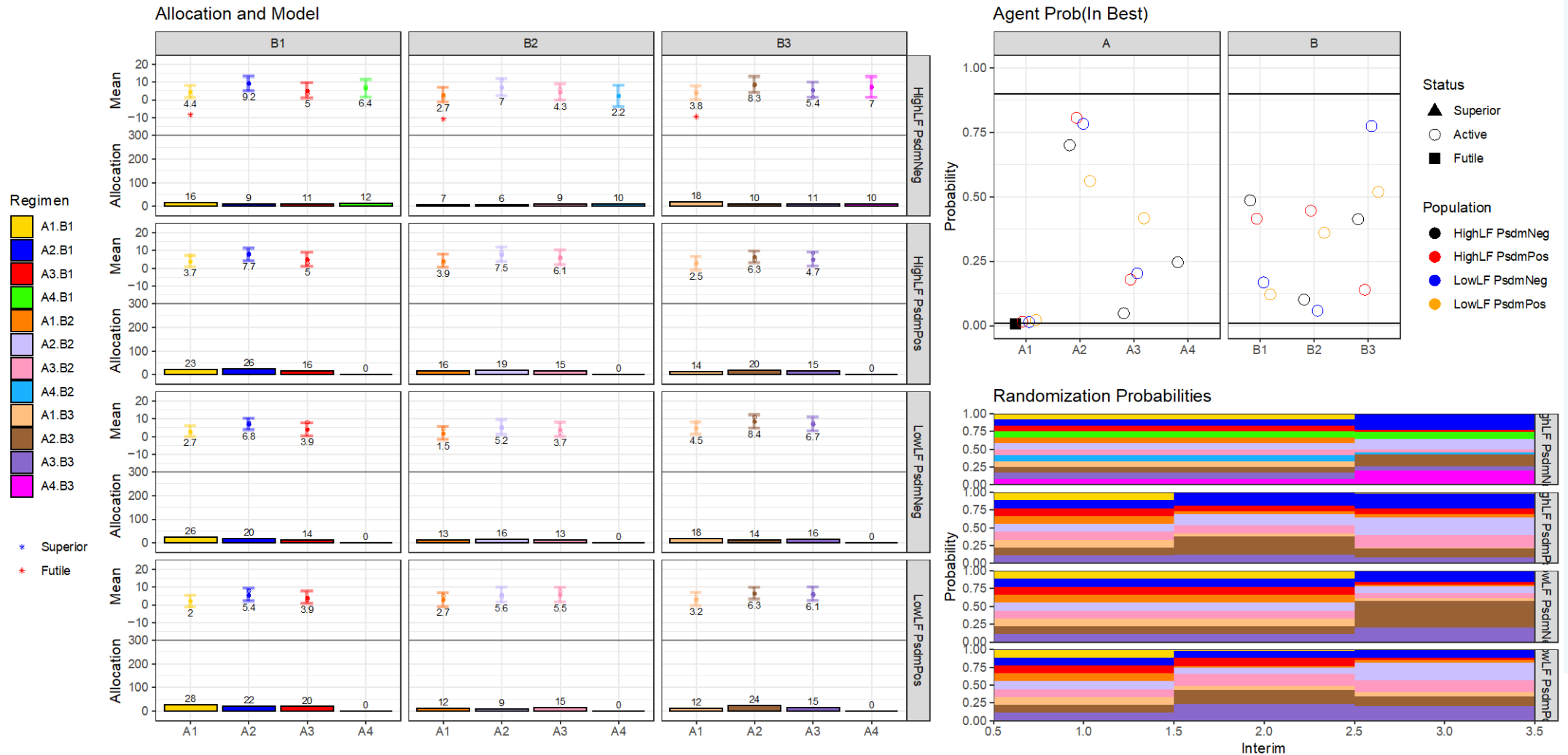
Example simulated trial: Interim 2

Look #2: N = 400



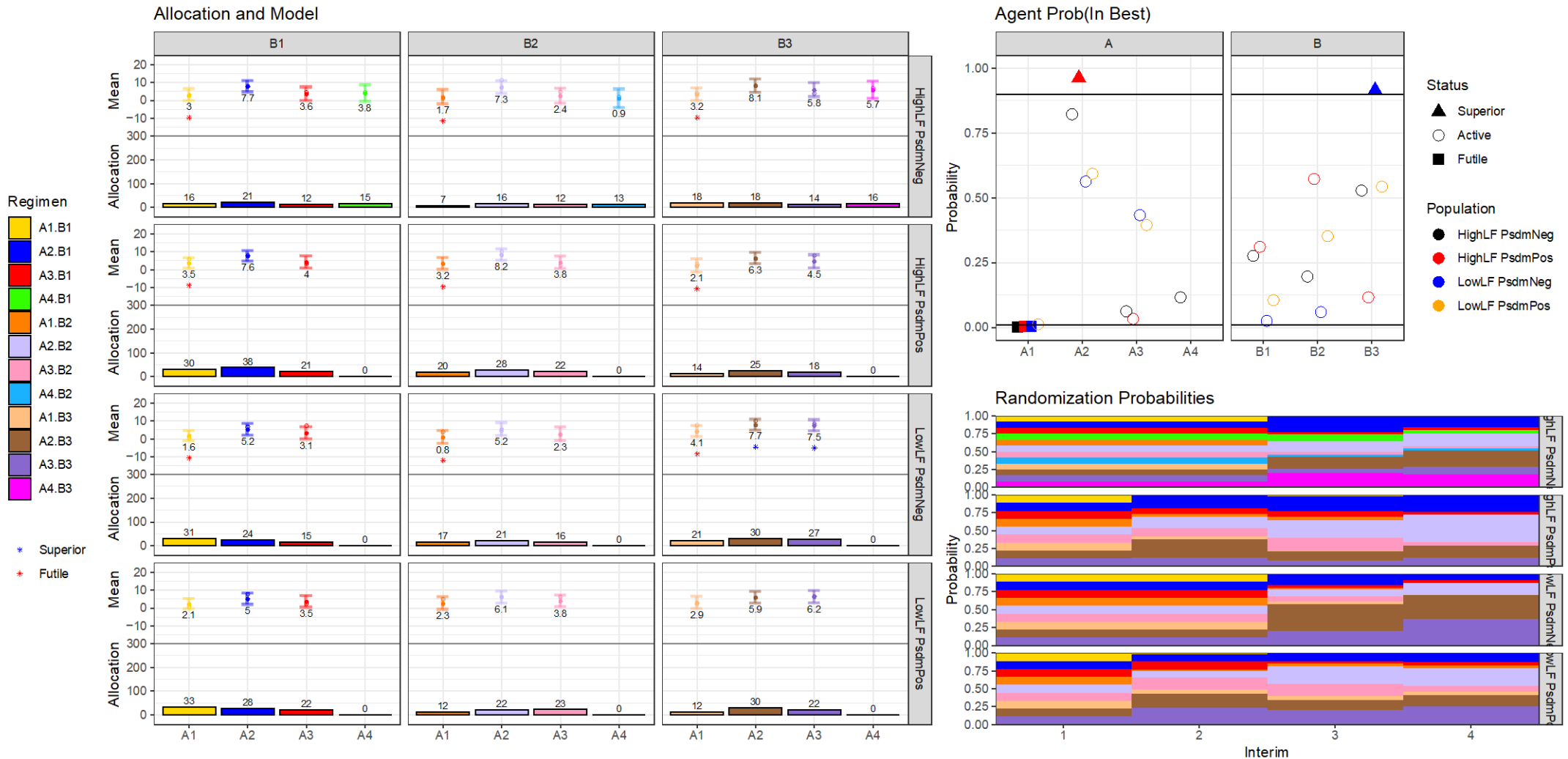
Example simulated trial: Interim 3

Look #3: N = 600



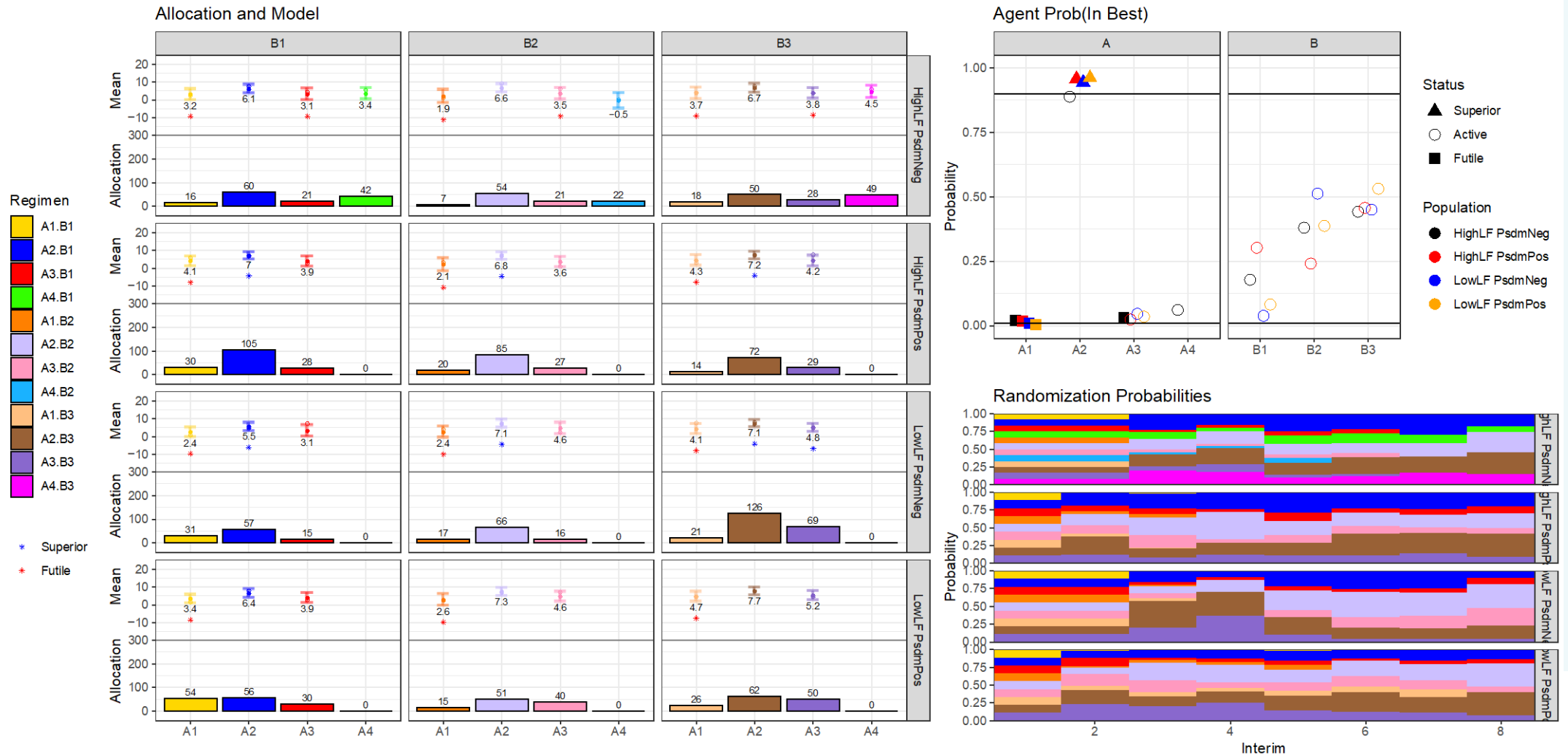
Example simulated trial: Interim 4

Look #4: N = 800



Example simulated trial: Interim 8

Look #8: N = 1600



Role of simulation

Simulation is used to generate 1000's of trials for each scenario (under a range of design parameters) and each scenario is summarised as a set of trial operating characteristics.

- The null scenario provides the false-positive error
- Other scenarios provide the power to detect clinically relevant treatment effects
- The average sample size under each scenario can be calculated
- Other useful operating characteristics are: proportion of trials when ineffective treatments are dropped (by interim) and number of participants exposed to ineffective treatments.

Role of simulation: example operating characteristics

Null scenario 1

False-positive error

Moderate increase to A2

Power

Simulation Scenario	Intervention	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Combined
1	A1	0.000	0.005	0.002	0.002	0.009
	A2	0.001	0.011	0.008	0.006	0.026
	A3	0.001	0.013	0.007	0.015	0.034
	A4	0.021	0.000	0.000	0.000	0.021
	B1	0.002	0.005	0.004	0.002	0.013
	B2	0.005	0.007	0.013	0.014	0.037
	B3	0.010	0.013	0.013	0.014	0.021
2	A1	0.00	0.000	0.000	0.000	0.000
	A2	0.601	0.868	0.905	0.918	0.998
	A3	0.000	0.000	0.000	0.000	0.000
	A4	0.000	0.000	0.000	0.000	0.000
	B1	0.004	0.009	0.006	0.009	0.025
	B2	0.013	0.025	0.022	0.022	0.070
	B3	0.019	0.014	0.023	0.028	0.069



Platform design criticisms

Wason et al. *BMC Medicine* (2019) 17:152
<https://doi.org/10.1186/s12916-019-1186-1>

OPINION

When to keep are not always

James M. S. Wason^{1,2*}

Abstract

Background: Adaptive designs are used to make changes to a trial as participants are enrolled to maximize their benefit or to minimize their size or the enrolment of patients. Their use in many clinical trials provide little efficiency. In our experience, factors mentioned in methodological papers are not actually what the problems are.

Main text: In this paper we discuss situations when the usual randomization when increased practice is needed.

Conclusion: Adaptive designs should be aware that they do not always provide an advantage. There should always be careful consideration of the potential benefits and disadvantages of an adaptive design.

Clinical Infectious Diseases

INVITED ARTICLE

INNOVATIONS IN DESIGN, EDUCATION

Resist the Temptation Randomization

Michael Proschan¹ and Scott Evans²

¹Mathematical Statistician, Biostatistics Research Branch, Director, Biostatistics Center, Milken Institute School of Public Health

Response-adaptive randomization (RAR) of participants randomized to inferior treatment groups causes many problems, including (1) sample-size distributions that can cause problems in analyzing results, and (2) the problems of RAR are most acute in the infectious disease settings where temporal trends are common.

Randomization, the most powerful tool in clinical trials, provides. Use of RAR is discouraged.

Keywords. response-adaptive randomization; temporal trend; platform trials; frequentist approach; Bayesian approach.

The NEW ENGLAND JOURNAL of MEDICINE

CORRESPONDENCE



Platform Trials — Beware the Noncomparable Control Group

TO THE EDITOR: The coronavirus disease 2019 (Covid-19) pandemic has highlighted the crucial role of randomized trials in guiding clinical practice and the need for designs that provide rapid evaluation of multiple interventions. Multi-group randomized clinical trials in which multiple experimental treatment groups are compared with a single control group allow for an efficient use of resources in that a separate control group does not need to be generated for each com-

group could bias the results of a trial is provided in Figure 1. Consider the decline in in-hospital mortality from Covid-19 that occurred over a 2-month period in the spring of 2020² and a hypothetical trial that compared a control treatment with an ineffective new agent that was not included in the randomization until the second month. If comparisons were made between the patients who received the control treatment during the 2-month period (April–May 2020) and

Platform design criticisms

Clinical Infectious Diseases

CORRESPONDENCE

The Temptation of Overgeneralizing Response-adaptive Randomization

TO THE EDITOR—We read with interest the recent article by Proschan and Evans [1] on the use of response-adaptive randomization (RAR) and its potential problems; however, these problems are neither

ne
pr
et
th
m
an
ce
th
ta

isting procedures that do not have these problems.

A discussion of many of the specific points of Proschan and Evans is presented by Robertson et al [2], which we briefly summarize for the points highlighted in the abstract of Proschan and

Sirkis et al.
BMC Medical Research Methodology (2022) 22:216
<https://doi.org/10.1186/s12874-022-01691-w>

BMC Medical Research
Methodology

RESEARCH

Open Access



Should RECOVERY have used response adaptive randomisation? Evidence from a simulation study

Tamir Sirkis^{1*}, Benjamin Jones² and Jack Bowden¹

Conclusions: Using response-adaptive randomisation within RECOVERY could have increased the number of patients receiving the optimal COVID-19 treatment during the trial, while reducing the number of patients needed to attain the same study power as the original study. This would likely have reduced patient deaths during the trial and lead to dexamethasone being declared effective sooner. Deciding how to balance the needs of patients within a trial and future patients who have yet to fall ill is an important ethical question for the trials community to address.

This design feature has recently been implemented within the REMAP-CAP platform trial.

55 years ago, has been fully vetted, and advocates of RAR recognize its extreme deficiencies [5]. We would encourage the use of the early controversy around the ECMO trial to learn when and how RAR might be used appropriately, rather

5. Burton PR, Guilan EC, Hussey MR. Interpreting the clinical trials of extracorporeal membrane oxygenation in the treatment of persistent pulmonary hypertension of the newborn. *Semin Neonatal* 1997; 2:69–79.
6. Rosenberger WF, Lachin JM. *Randomization in clinical trials*. Wiley Series in Probability and Statistics. Hoboken, New Jersey: John Wiley & Sons, 2016.

Platform design criticisms

1. Temporal changes & non-concurrent controls can produce biased effect estimates: **model time effects in the primary analysis and perform sensitivity analyses**
2. Increased potential for selection & operational bias: **trial governance, including defined roles for blinded and unblinded study personnel**
3. More resources needed to initiate trial: **but may be resource saving overall**
4. Slow collection/availability of high quality data: **efficient use of technology/automation**
5. Need to custom-build infrastructure and response-adaptive-randomisation module

Platform design criticisms

Automation can increase **quality**:

- Complex assessment of eligibility & collection of consent (time points, domains, sub-studies)
- Randomisation to a combination of interventions (dependent on ineligibilities)
- Electronic prompts & data capture
- Trial monitoring

Automation can increase **efficiency**:

- Collection & review of data
- Repeat analyses on accumulating data
- Some adaptations (e.g, RAR or arm dropping)
- Generation of standardised reports

AuTOMatic: Adaptive Trial of MessAging to improve Immunisation Coverage



SmartVax interrogates BP/MD to find which children are due for a vaccine



Randomised to one of 13 different groups



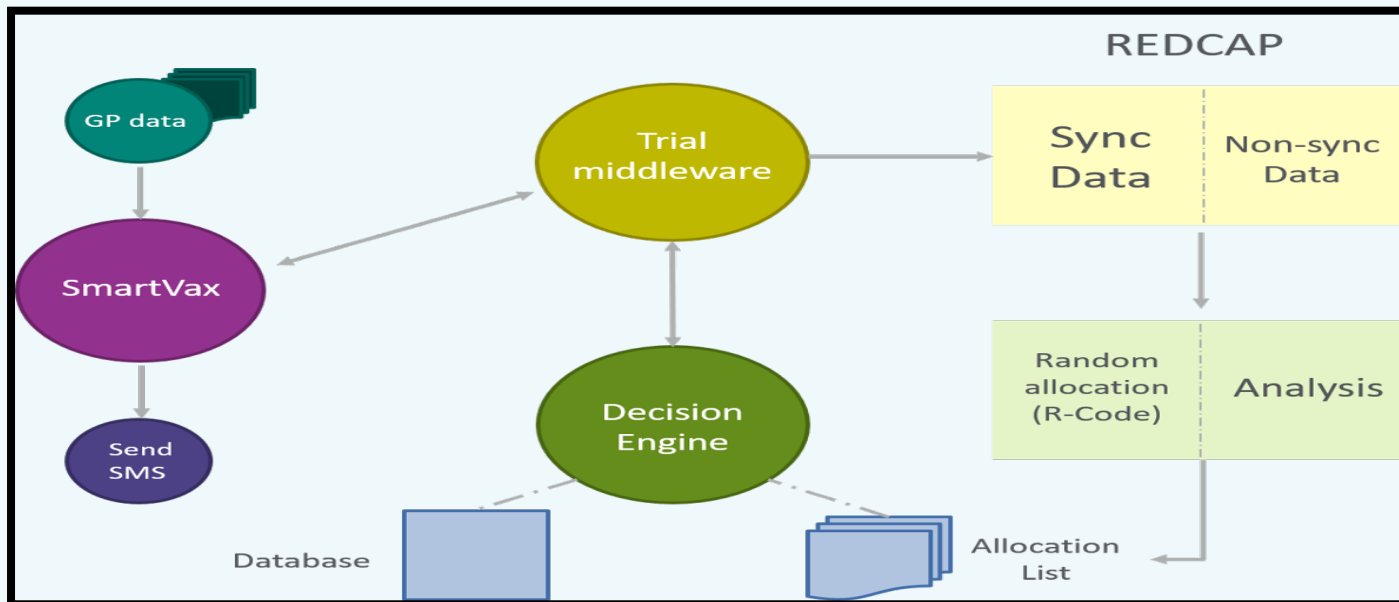
Parent receives an SMS (or no SMS if in control group)



SmartVax interrogates BP/MD to find which children have received their vaccine.



Analyse which group is best and randomise future children to these groups



Demonstration of a semi-automated structure with the flexibility to incorporate new functionality or modules and expand to federated databases



Australian
Clinical
Trials
Alliance

<https://www.telethonkids.org.au/projects/automatic-adaptive-trial-of-messaging/>

www.clinicaltrialsalliance.org.au

Platform design criticisms

6. Response adaptive randomisation: allocate proportional to posterior probability that an intervention is the best, **accounting for sample size & variability**

Further RAR options to explore in simulation:

- If only 2 treatment arms: optimal treatment allocation 1:1
- Consider fixing allocation to control arm (to maintain power)
- Consider delaying start of RAR until a minimum number allocated or after the early interims
- Consider whether floor (minimum) or ceiling (maximum) allocation probabilities needed

To embrace this complexity, there is a global need for training and capacity building