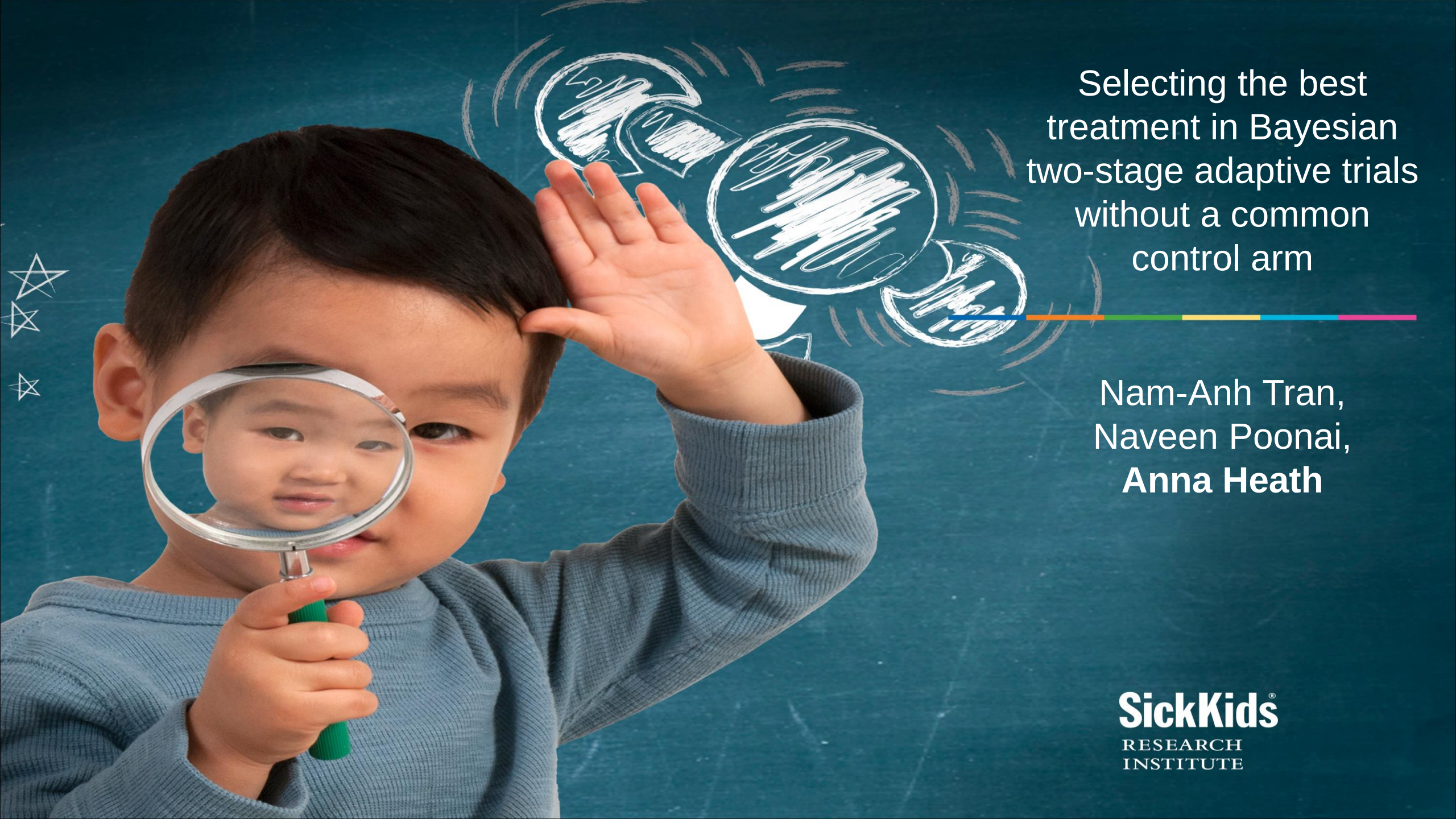
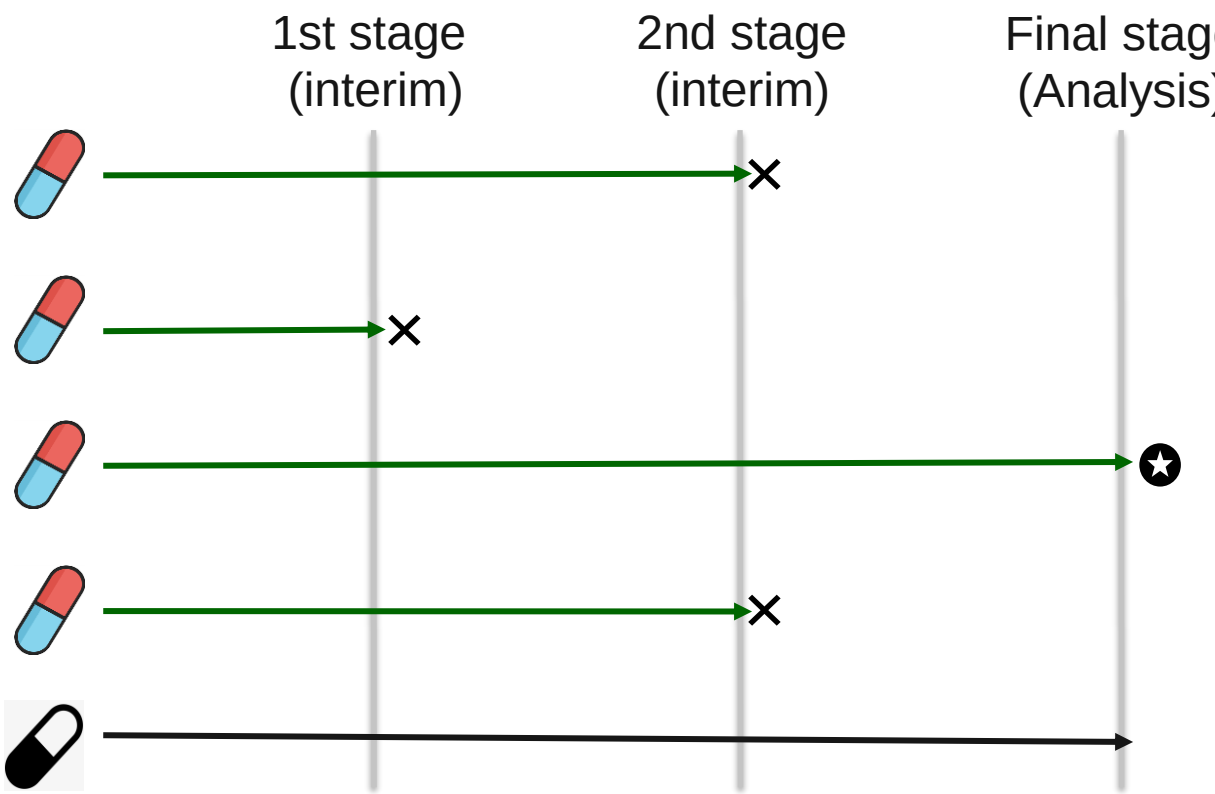


Selecting the best treatment in Bayesian two-stage adaptive trials without a common control arm

Nam-Anh Tran,
Naveen Poonai,
Anna Heath

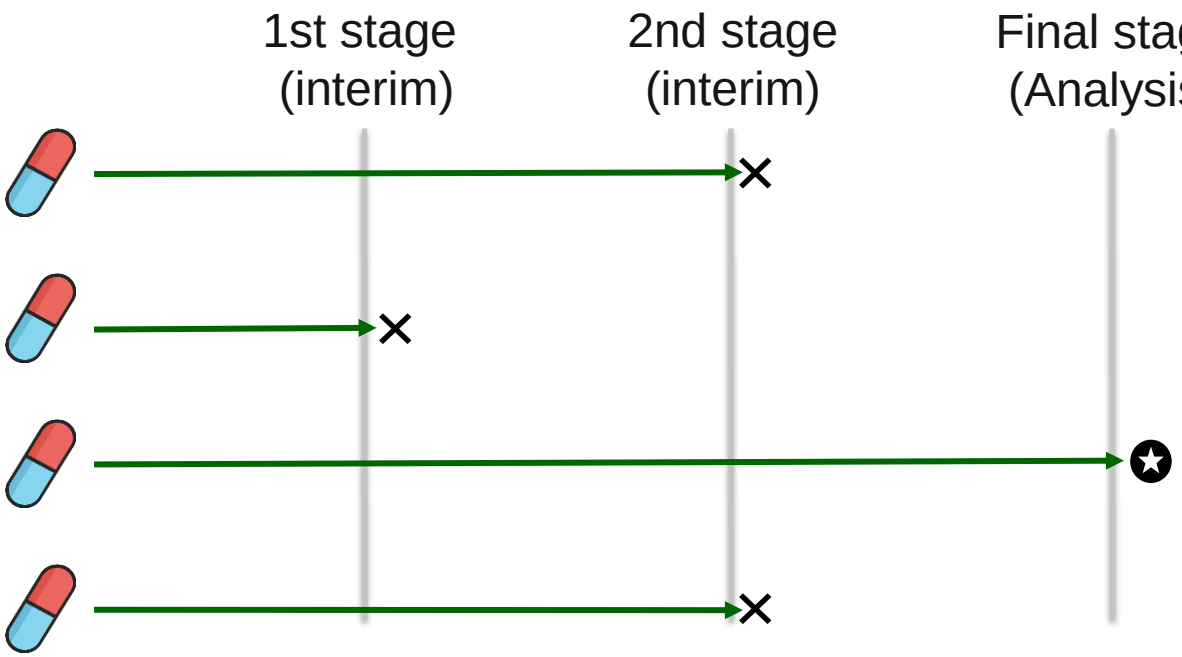
SickKids®
RESEARCH
INSTITUTE

An adaptive trial with multiple arms



- Multiple active treatments are evaluated through multiple stages.
- ***A common comparator is used through all stages, alongside treatments, to make pairwise comparisons.***
- Futile treatments are stopped at interim stages.
- Potential treatments are assessed at the analysis stage to select the superior treatment.
- The sample size at the final analysis is pre-specified.

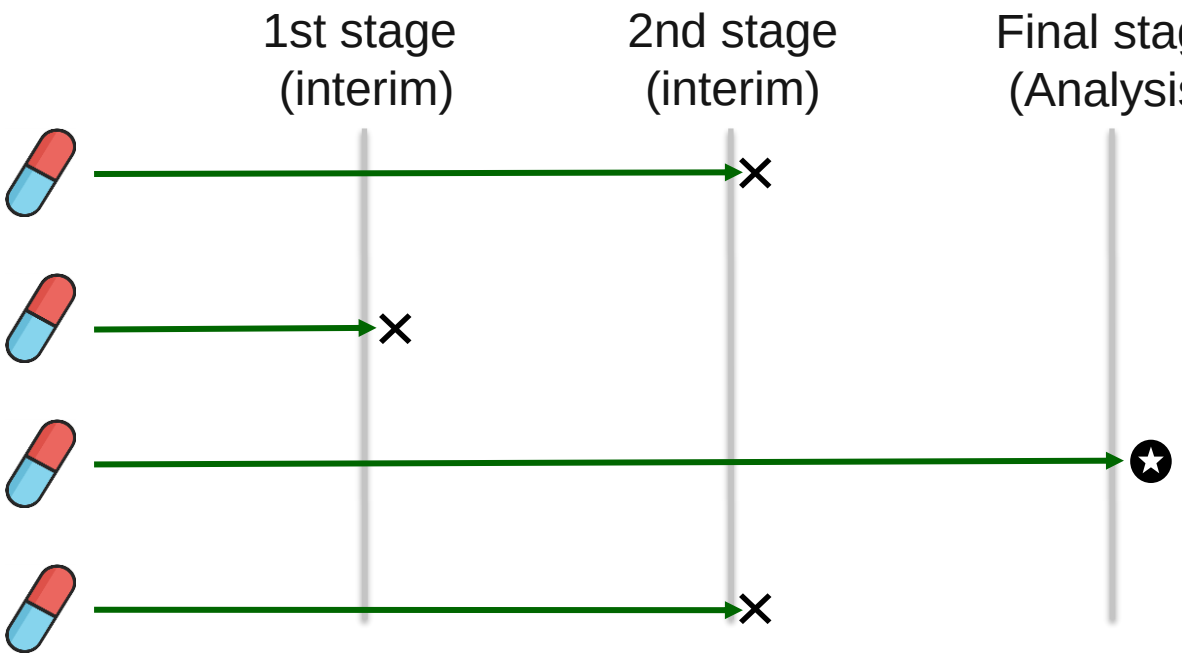
An adaptive trial with multiple arms



- Multiple active treatments are evaluated through multiple stages.
- Futile treatments are stopped at interim stages.
- Potential treatments are assessed at the analysis stage to select the superior treatment.
- The sample size at the final analysis is pre-specified

How to evaluate active treatments without a comparator?

An adaptive trial with multiple arms

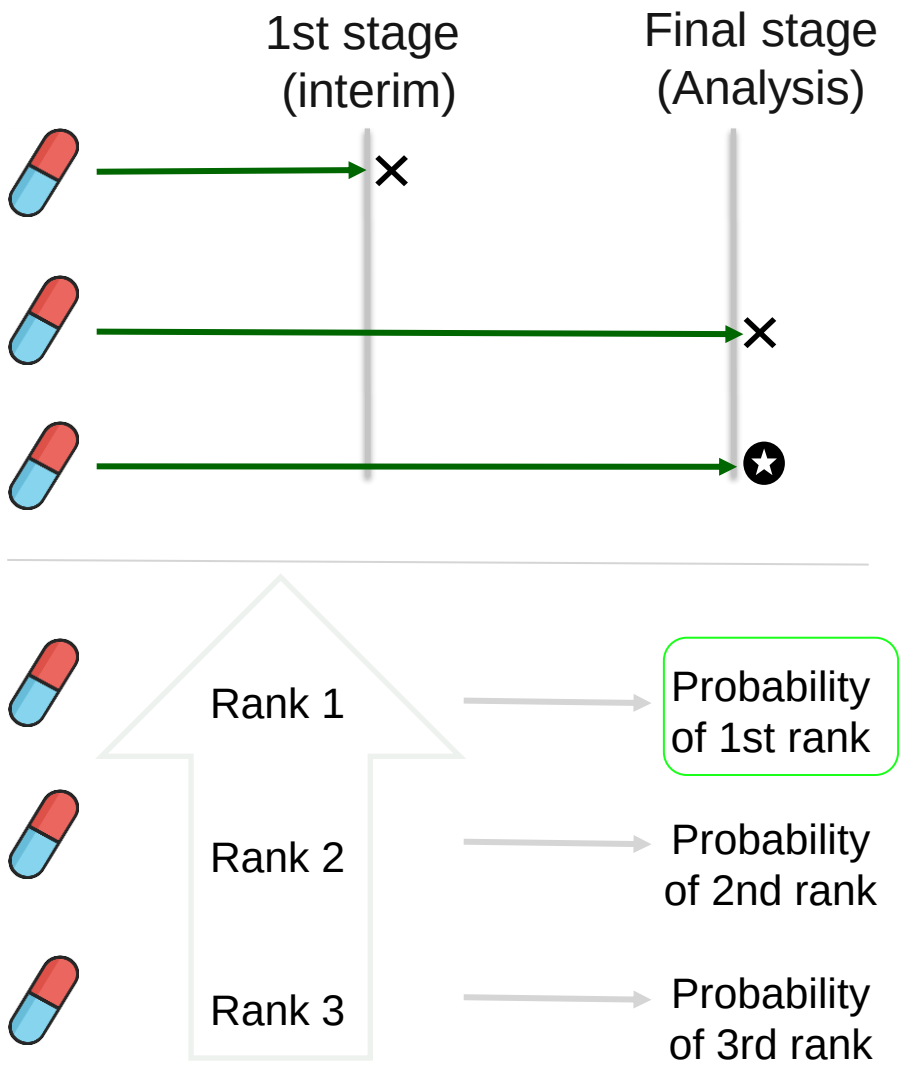


- Multiple active treatments are evaluated through multiple stages.
- Futile treatments are stopped at interim stages.
- Potential treatments are assessed at the analysis stage to select the superior treatment.
- The sample size at the final analysis is pre-specified.

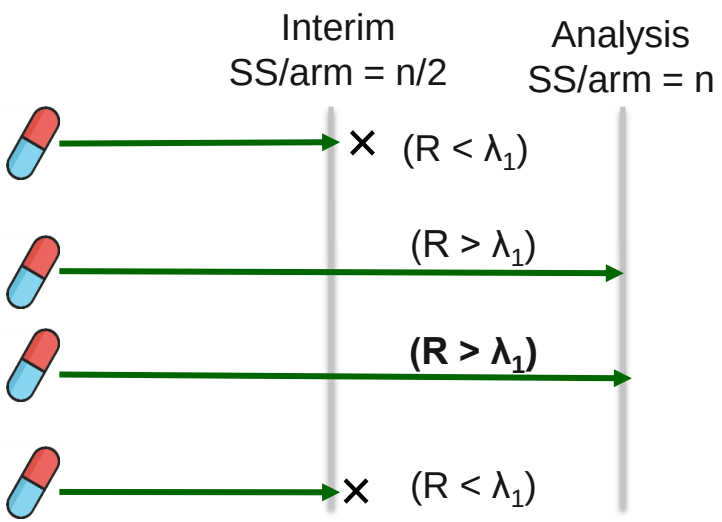
- **Treatments are ranked based on their performance.**
- **We employ a ranking scheme outputting the possibility that each considered treatment is the best.**

Our adaptive trial with multiple arms and two stages

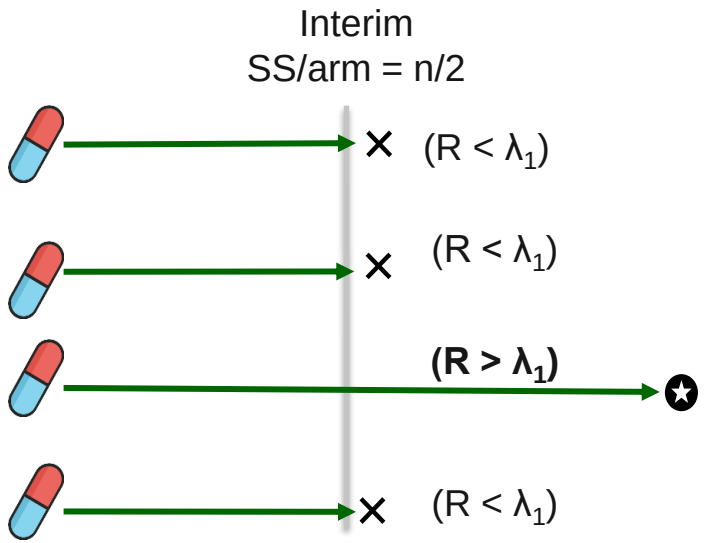
- Multiple arms and two stages; no common control.
- Treatments are ranked using **the surface under the cumulative ranking curve (SUCRA)** or **the probability of first rank (PFR).**
- These two approaches require the Bayesian framework.



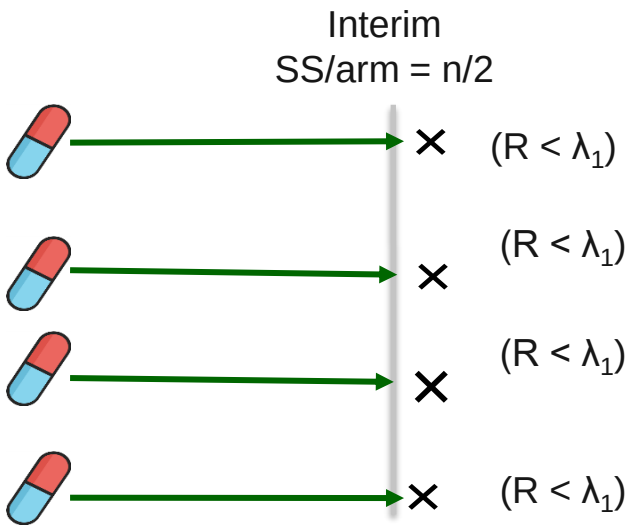
The structure of our proposed trial



Re-evaluated



Early terminated,
Superiority is declared

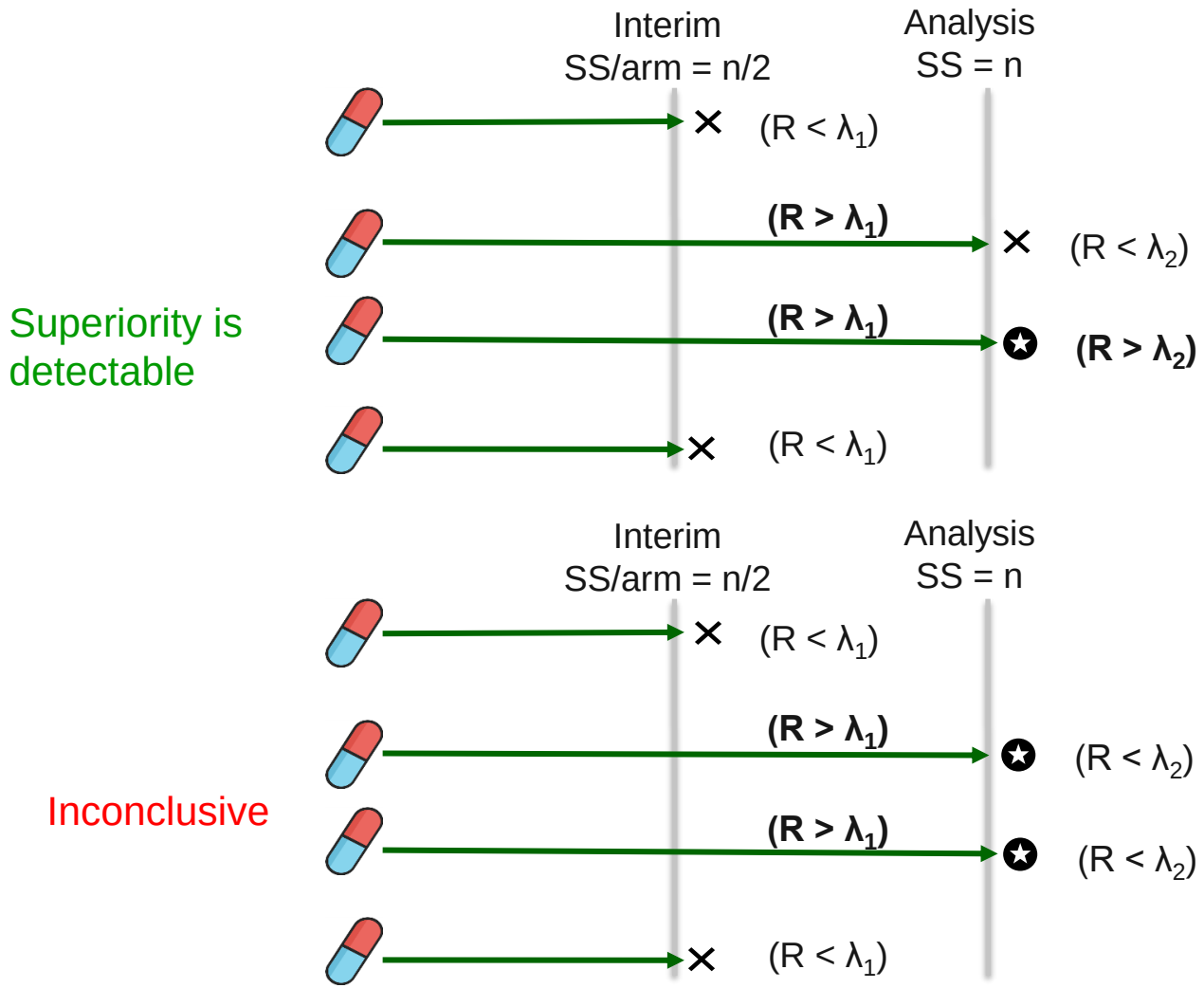


Inconclusive

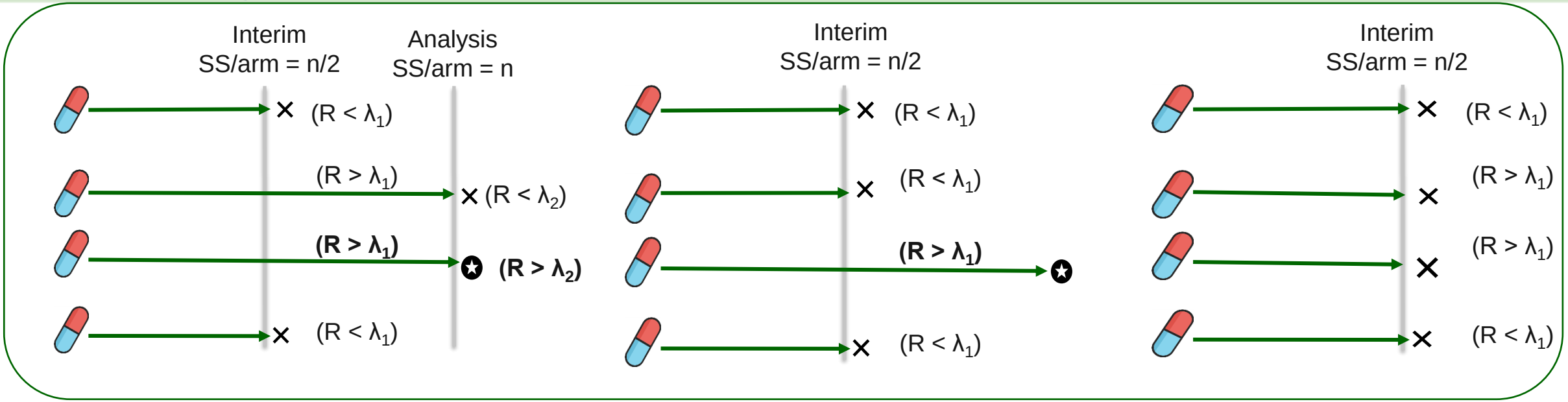
- Determine the maximum sample size n.
- **Interim stage**: calculate the ranking criterion (SUCRA or PFR) of K treatment after ~n/2 patients:
  - Drop treatments which satisfy the pre-defined dropping condition D1.
  - All other treatments are re-evaluated in the final stage.
  - If (K-1) treatments are dropped, the trial is terminated, and the remaining treatment is superior.
  - If K treatments are dropped, the trial is inconclusive.
- Dropping condition D1: $R < \lambda_1$. (R is the ranking criterion)

The structure of our proposed trial

- **Final analysis stage**: Only remaining treatments are evaluated with sample size of n per arm:
  - If one treatment satisfies the superiority condition D2, it will be declared superior.
  - Otherwise, the trial is inconclusive.
- Superiority condition D2: $R > \lambda_2$ (R is computed among the remaining treatments)



Specifying the thresholds for D1 and D2



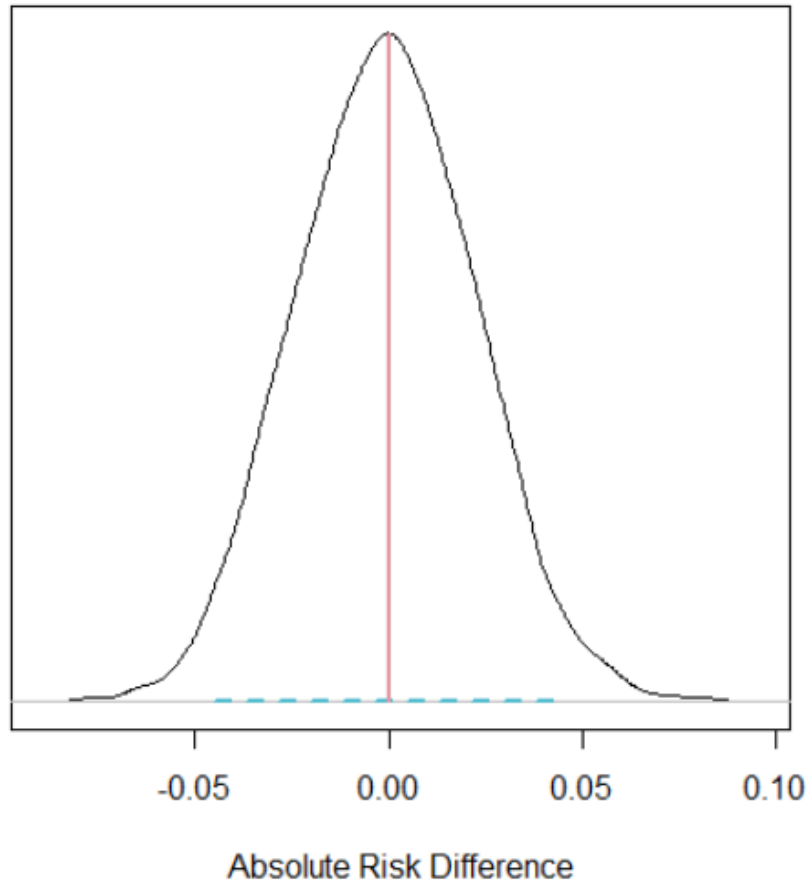
- The thresholds λ_1 and λ_2 are chosen to ensure a Type I error below 0.05 and similar across the two ranking approaches.
- The thresholds λ_1 and λ_2 vary depending on K and the ranking criterion
- Power is simulated to compare two ranking approaches.

Determining the maximum sample size

Average Length Criterion *in multi-arm studies*

- Simulate multiple datasets from the prior predictive distribution for a fixed sample size.
- For each dataset, calculate the length of the 95% credible interval for the average effect.
- Determine the average length of these intervals across the datasets for each treatment and extract the maximum length.
- Select the sample size where the maximum average length is less than a pre-specified cut-point.

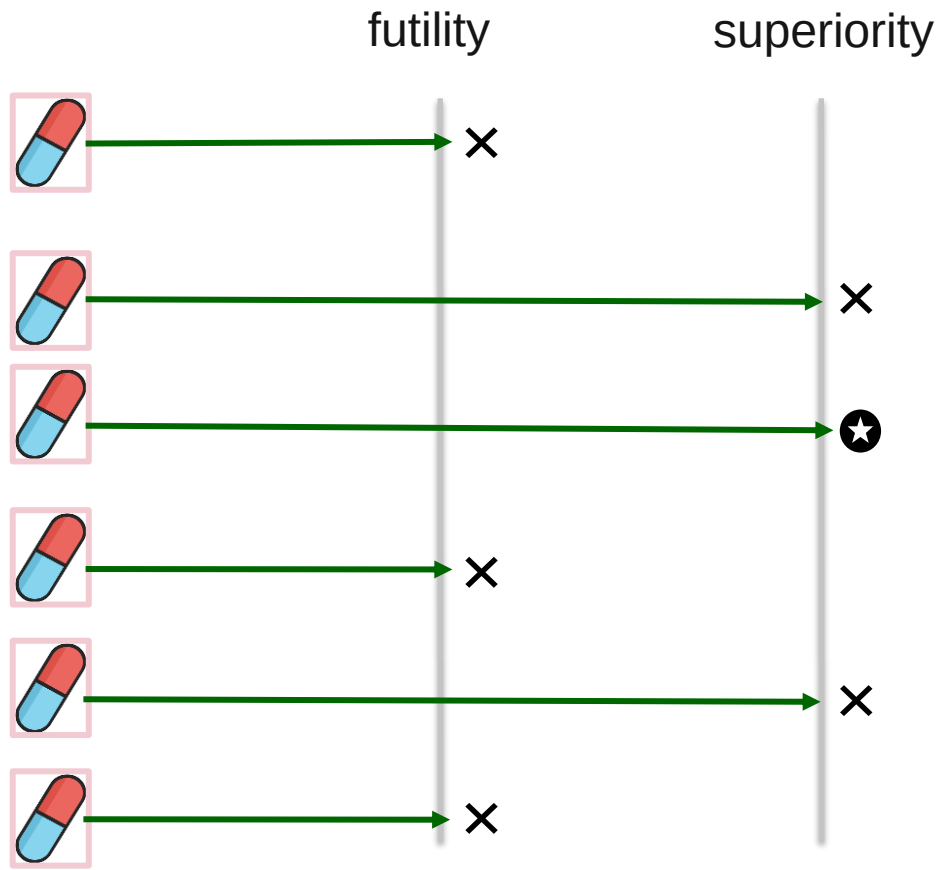
Sample size	Average length			
	Trt 1	Trt 2	Trt 3	Max
85	1.31	1.21	1.46	1.46
90	1.29	1.32	1.35	1.35
95	1.25	1.11	1.23	1.25
100	1.21	1.19	1.22	1.22



Accounting for parameter uncertainty in the design

- Simulate a common mean outcome from a design prior.
- Simulate data conditional on this common mean outcome.
- Obtain posterior distribution and calculate the SUCRA and PFR.
- *Type I error:* The proportion of trials that conclude superiority of any intervention.

- Simulate the arm specific mean outcome from separate priors – determine the simulation specific optimal treatment.
- Simulate data for each arm using specific mean outcome.
- Obtain posterior distributions and calculate the SUCRA or PFR.
- *Power*: The proportion of trials that conclude superiority of the simulation specific optimal intervention.



Comparing the two ranking criteria

Type 1 Error

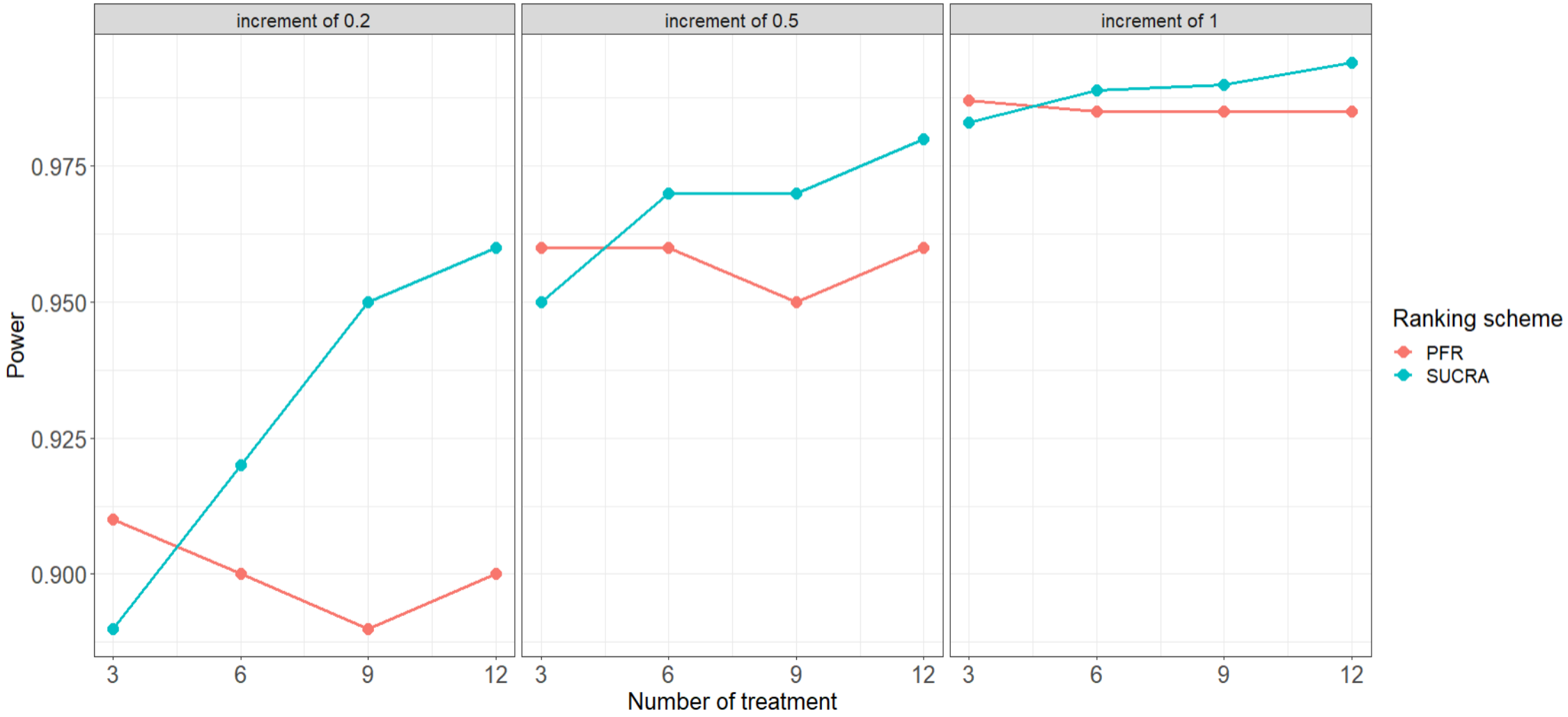
- Vary the number of interventions; K = 3, 6, 9, 12
- Determine the decision thresholds λ_1 and λ_2

Power

- Vary the difference in the prior mean effectiveness between the interventions; α = 0.2, 0.5, 1.0.
- Each prior has the same variance in this simulation study.

K	Ranking scheme	λ_1	λ_2
3	PFR	0.07	0.99
3	**SUCRA**	**0.3**	**0.95**
6	PFR	0.085	0.975
6	**SUCRA**	**0.051**	**0.95**
9	PFR	0.045	0.95
9	**SUCRA**	**0.09**	**0.975**
12	PFR	0.078	0.95
12	**SUCRA**	**0.005**	**0.975**

Comparing the SUCRA and PFR



Anxiolysis for laceration repair in children: A multicenter adaptive randomized trial (ALICE)

- Children aged 1-13 years with simple laceration repair
- Three interventions: Dexmedetomidine or Midazolam or N_2O
- Observational Scale of Behavioral Distress – Revised (OSBD-R) score

- Design priors for Midazolam and N_2O were from a published study.
- For Dexmedetomidine, we had individual-level data from a dose finding study.

	Least Square Means			
Period	**M**	**MN**	**N**	**SC**
OSBD scores*				
Baseline	0.1	0.2	0.3	0.33
Inject lidocaine	1.5	0.7	0.7	2.4
Cleaning	1.2	0.4	0.6	2.0
Suturing	1.9	0.7	0.4	2.0
Recovery	0.1	0.6	0.3	0.3
Suturer satisfaction†	7.5	8.0	8.2	6.6
Recovery time (min)‡	30	28	21	20

Results for the ALICE trial

Sample Size	Dexmedetomidine	Midazolam	N_2O	Max
80	1.06	1.05	1.05	1.06
85	1.05	1.03	1.03	1.05
90	1.00	0.99	0.99	1.00
95	**0.96**	**0.95**	**0.95**	**0.96**
100	0.95	0.94	0.94	0.95

Ranking scheme	λ_1	λ_2	Type I error	Power
PFR	**0.07**	**0.99**	**0.043**	**0.965**
SUCRA	0.3	0.95	0.042	0.957

Conclusions

- The proposed trial design detects the superior treatment among a set of active comparators.
- The thresholds for decision making should be calibrated.
- The SUCRA is better able to detect the true optimal treatment as the number of active comparators increases.
- For three interventions, as in our study, the PFR is superior.
- Other ranking methods could also be considered.