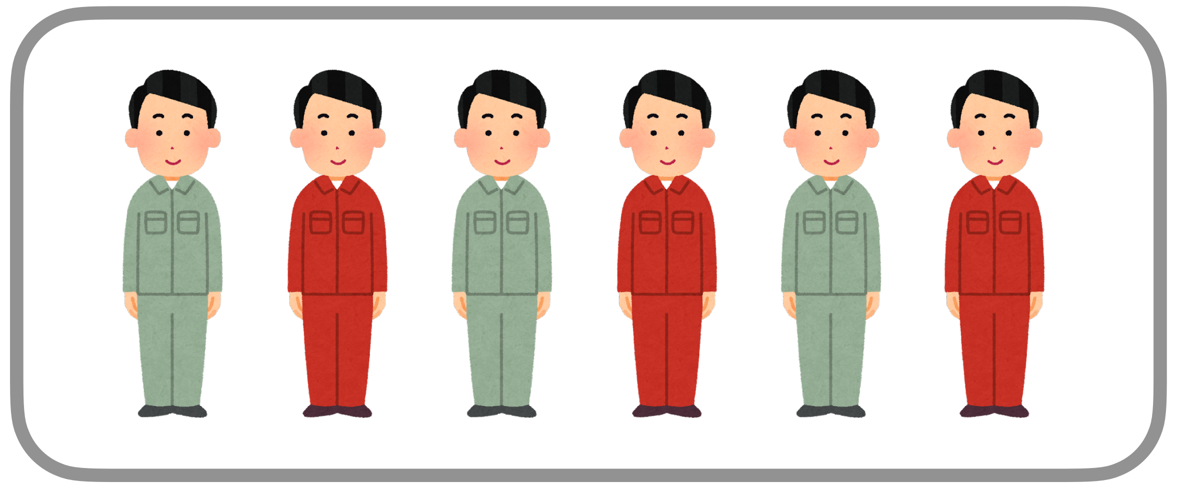
Selective Randomization Inference for Adaptive Clinical Studies

Tobias Freidling, Qingyuan Zhao, Zijun Gao

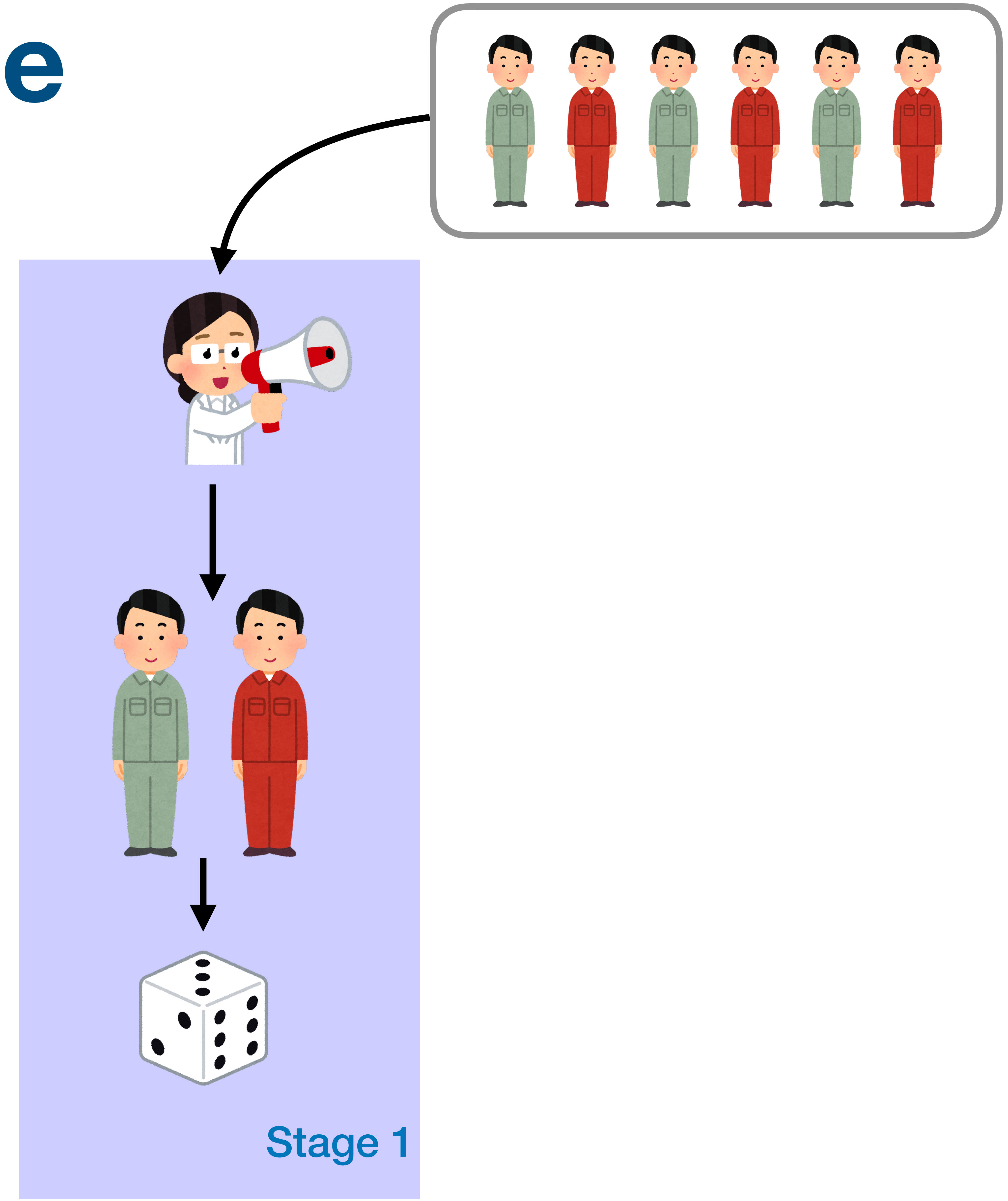
Response-Adaptive Randomisation in Clinical Trials Workshop 29/02/2024

Example

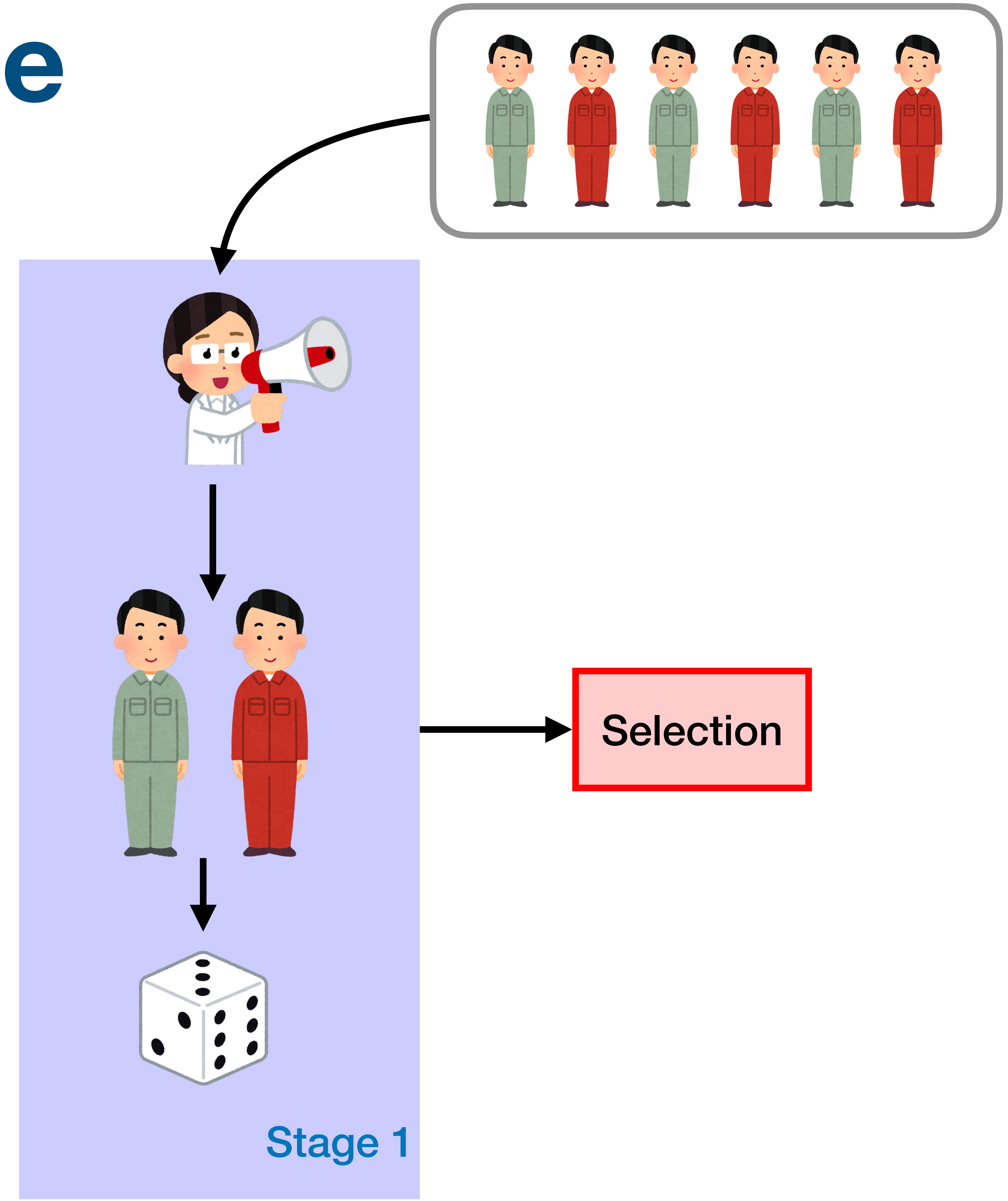
Example



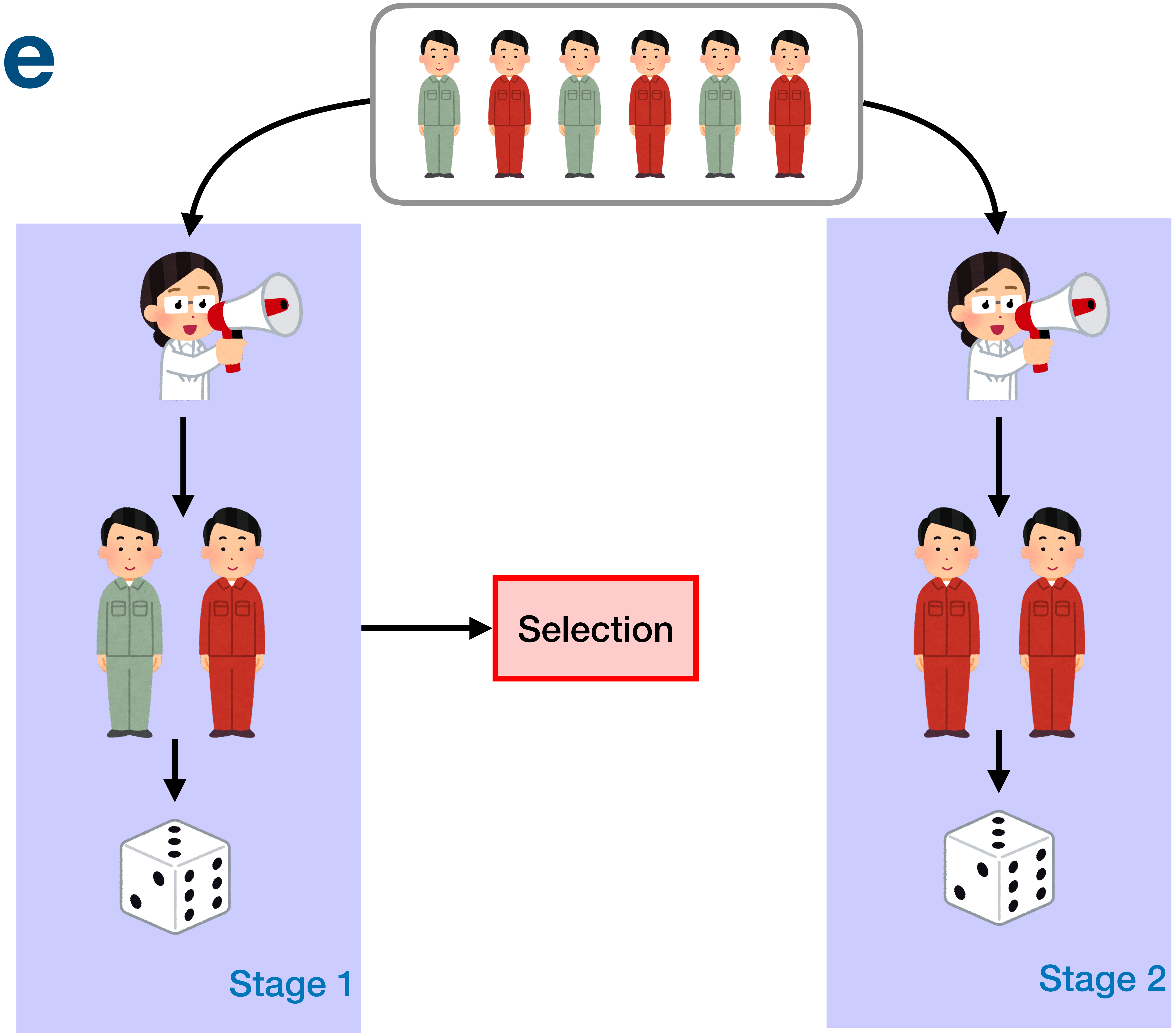
Example



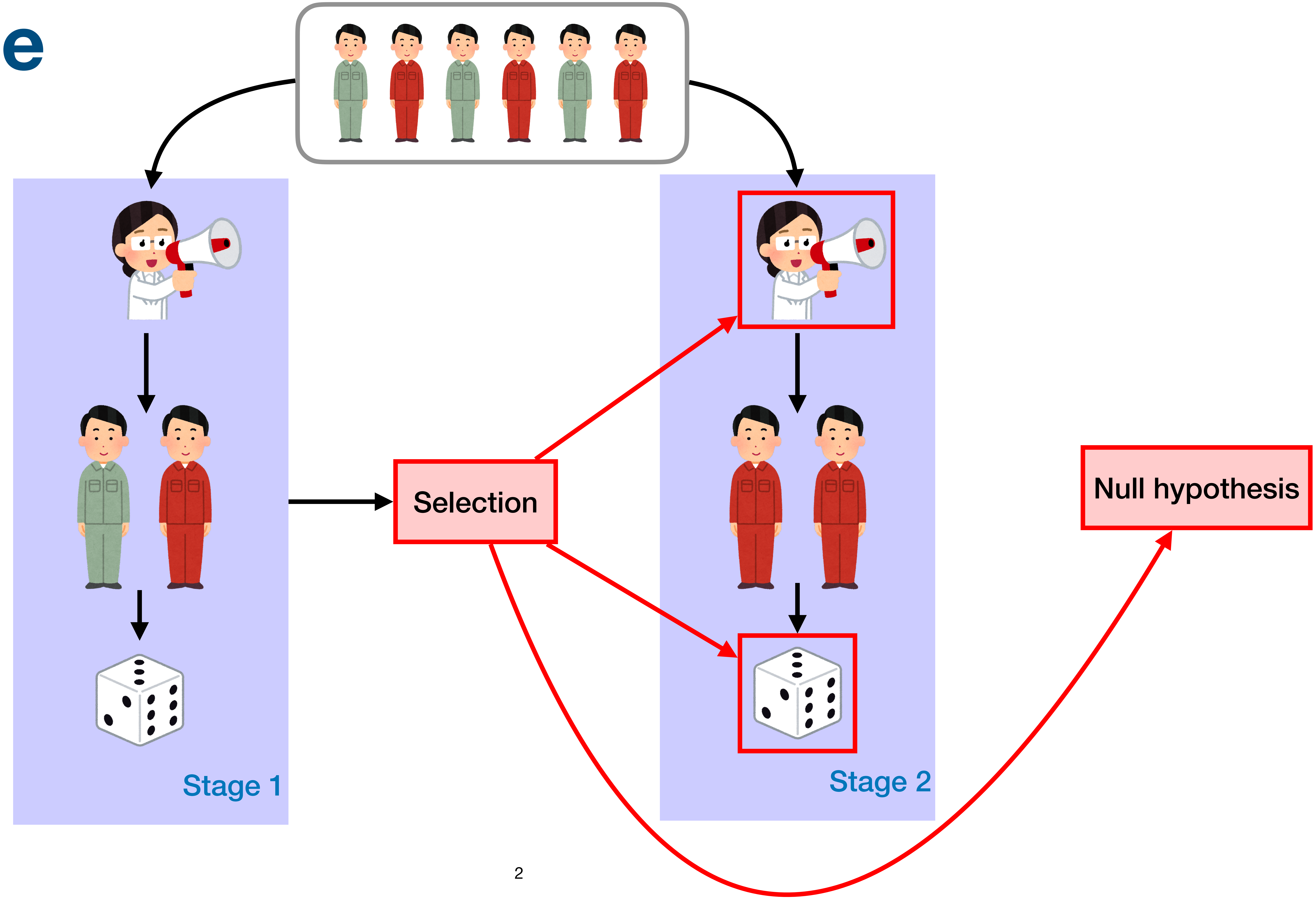
Example



Example



Example



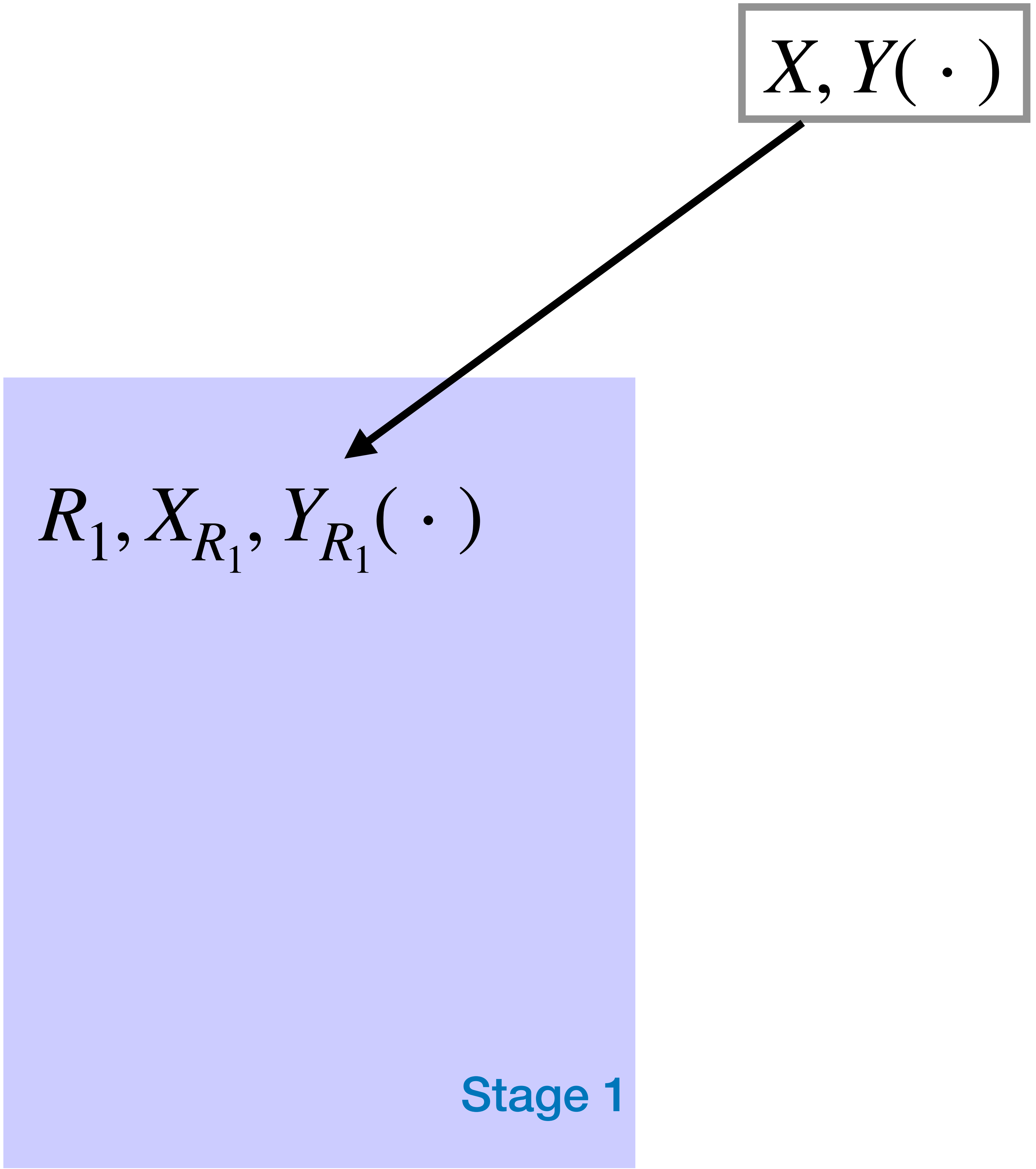
Graphical Model

Graphical Model

$X, Y(\cdot)$

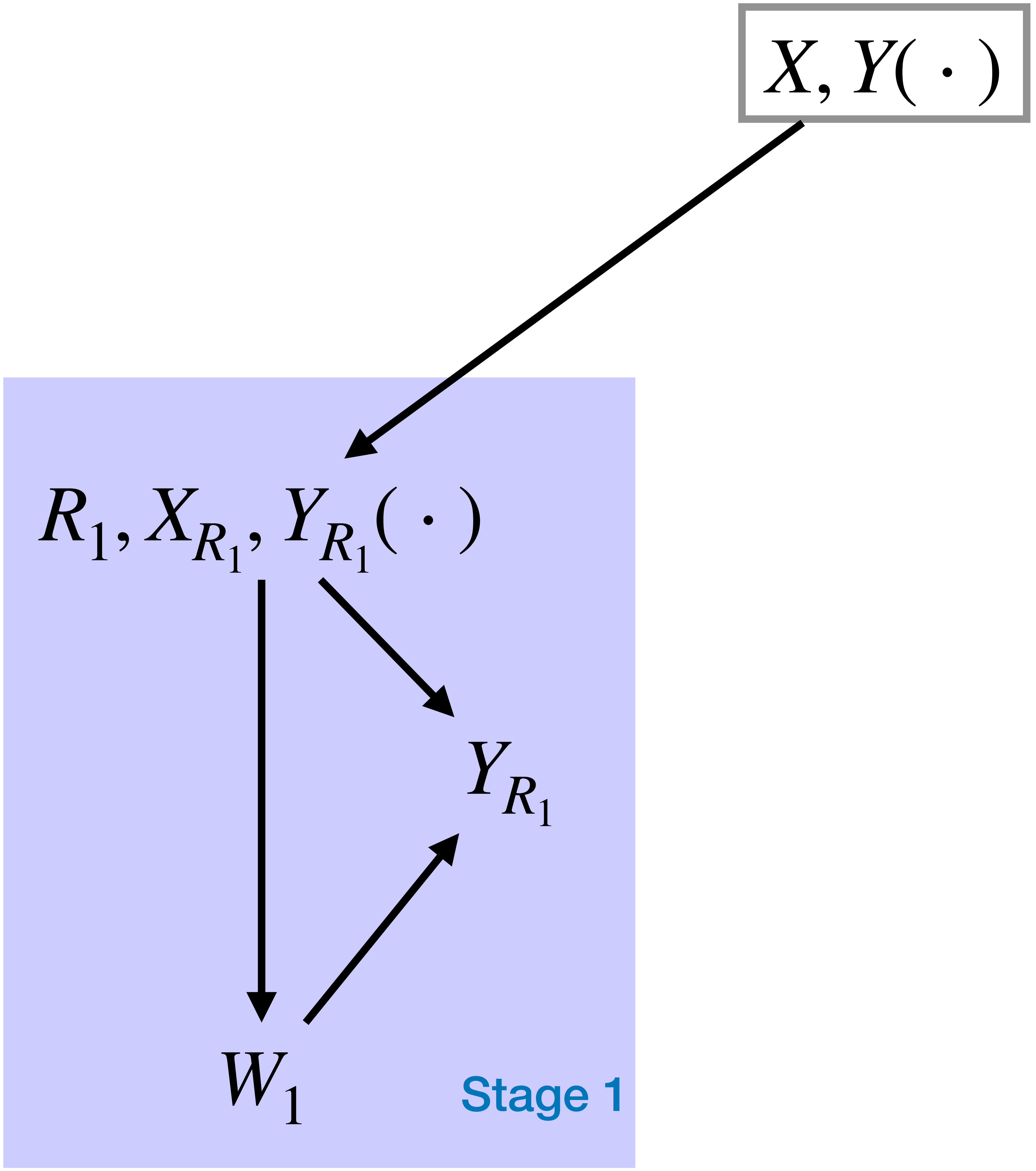
- Covariates: X
- Potential outcomes: $Y(\cdot)$

Graphical Model



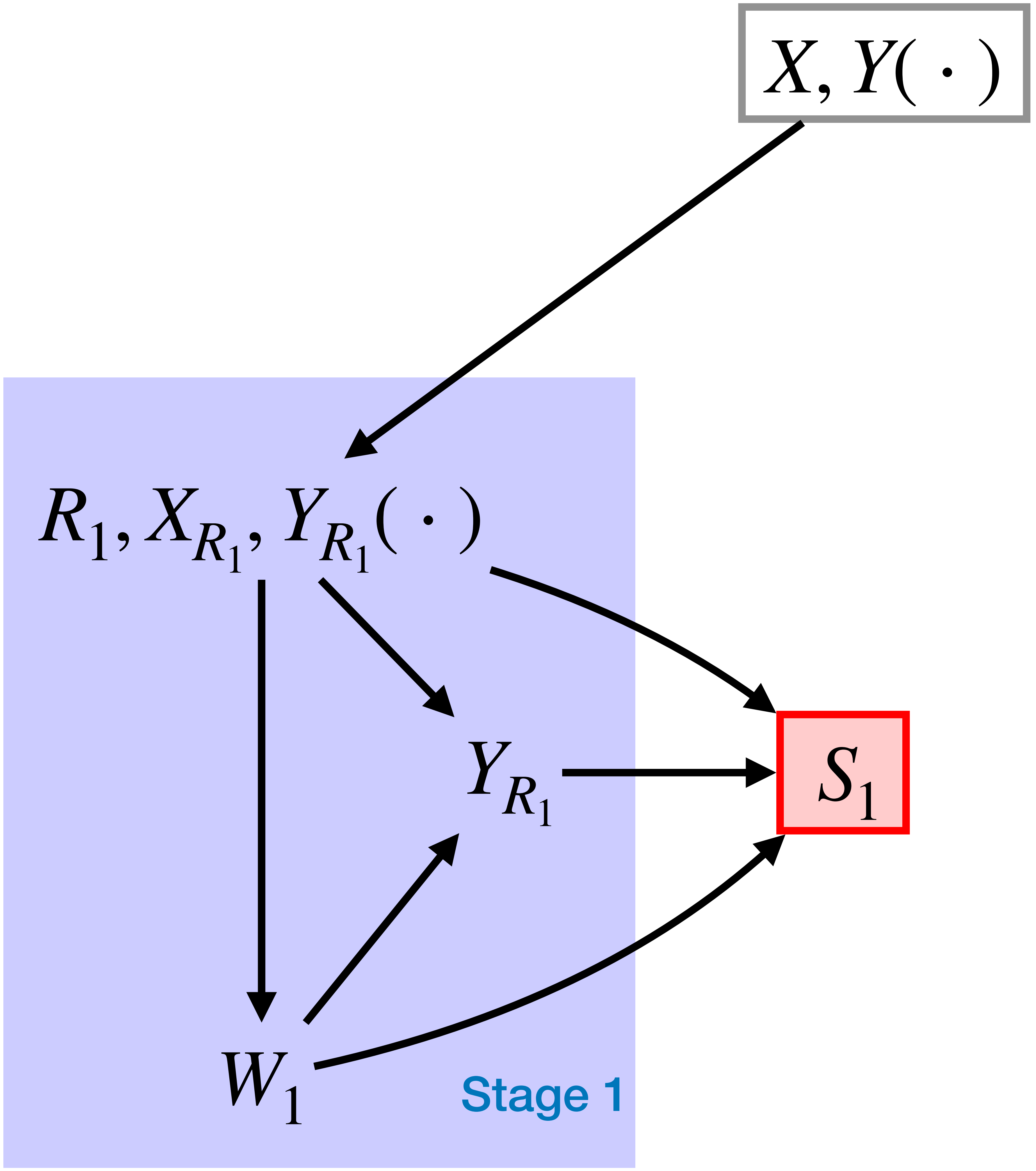
- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: R_1, R_2

Graphical Model



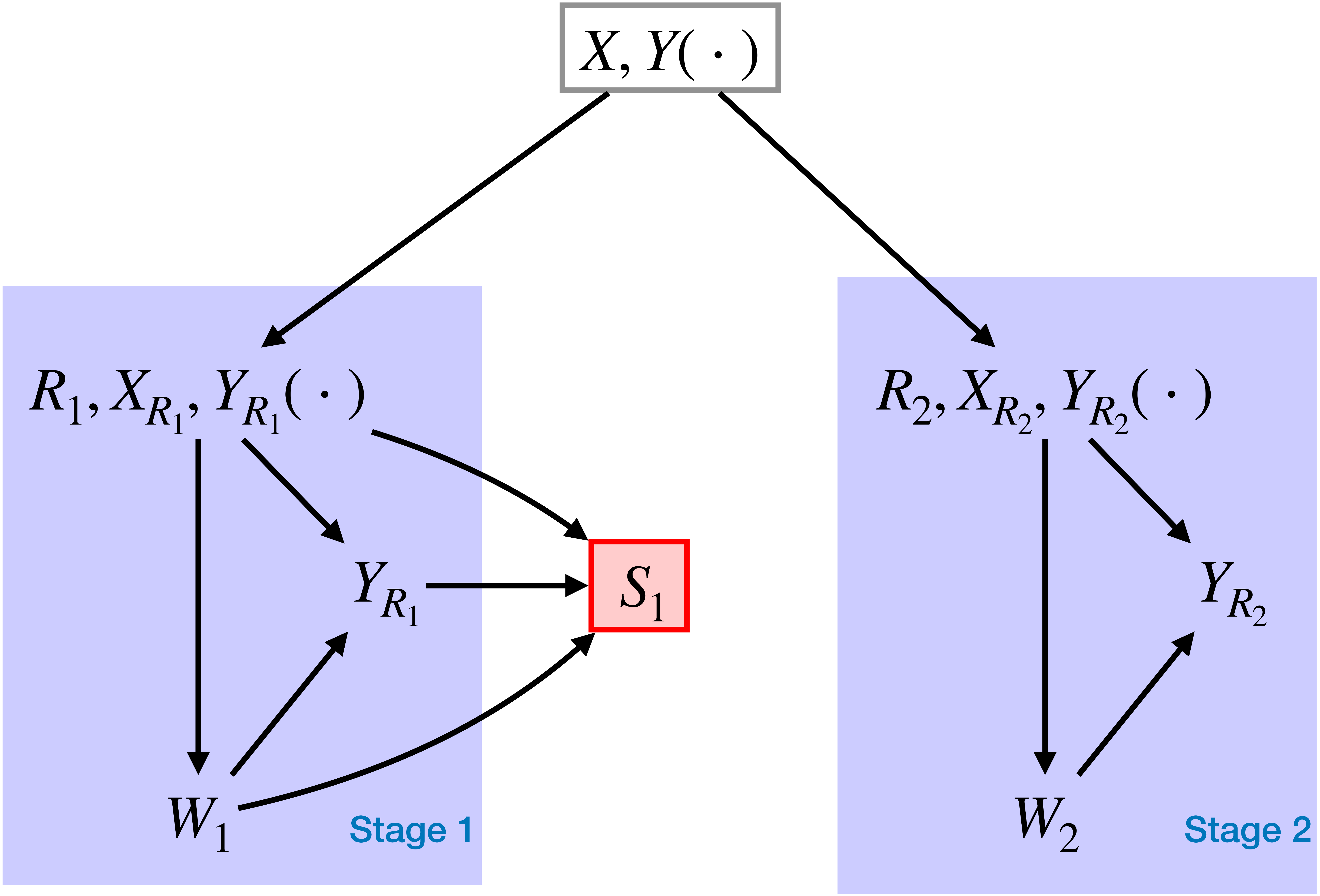
- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: R_1, R_2
- Treatments: W_1, W_2
- Observed outcomes: $Y = Y(W)$

Graphical Model



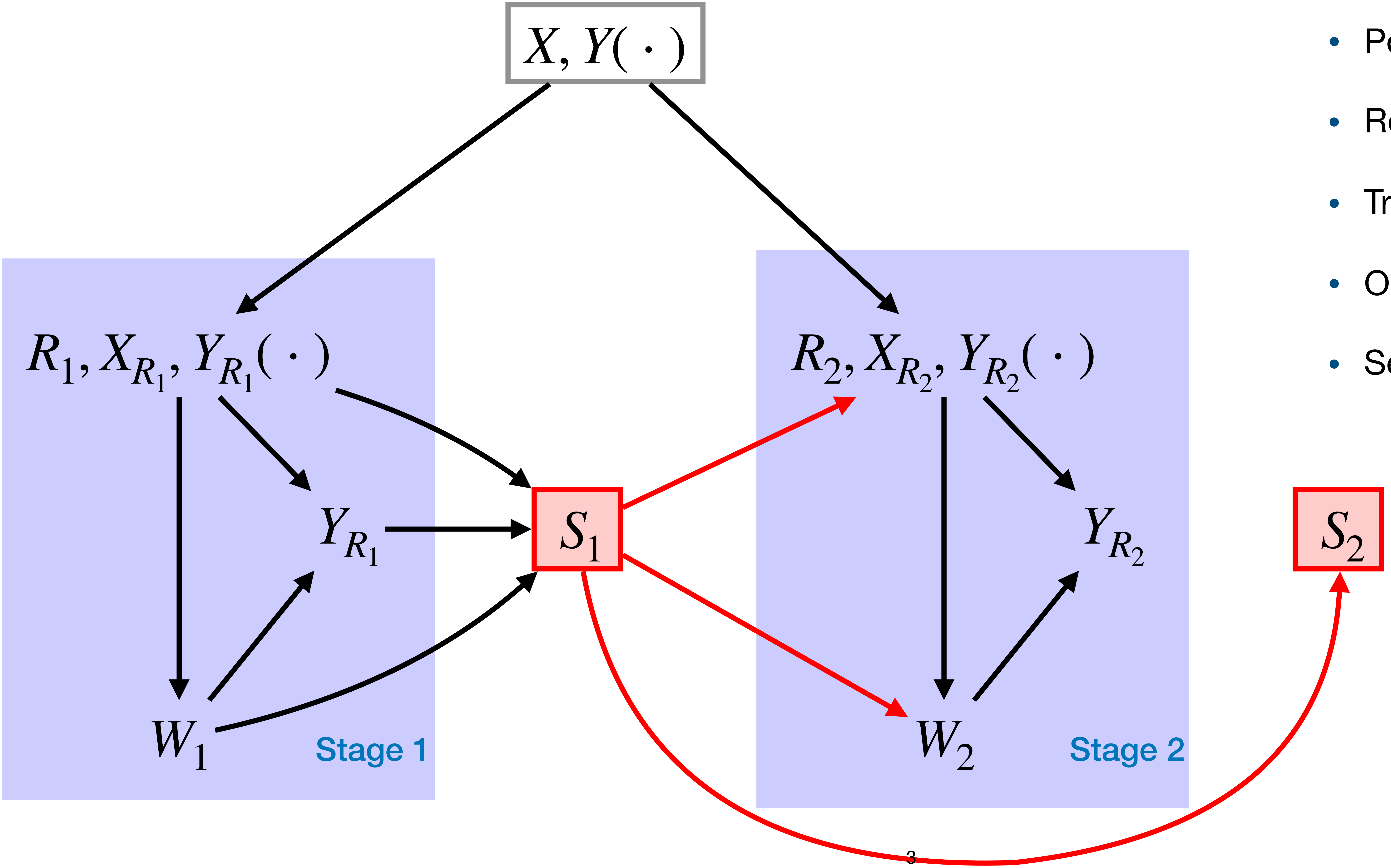
- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: R_1, R_2
- Treatments: W_1, W_2
- Observed outcomes: $Y = Y(W)$
- Selective choice: S_1, S_2

Graphical Model



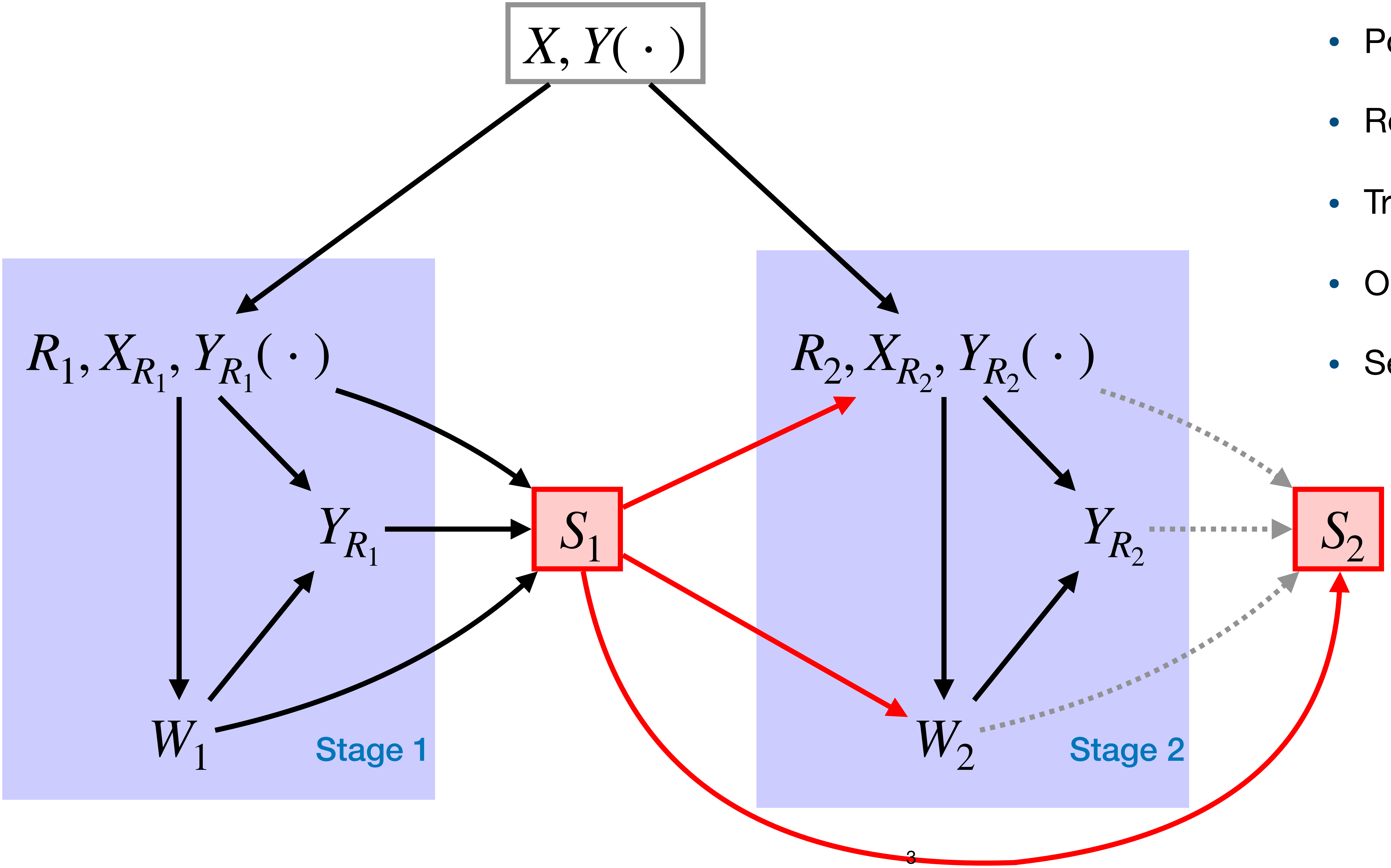
- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: R_1, R_2
- Treatments: W_1, W_2
- Observed outcomes: $Y = Y(W)$
- Selective choice: S_1, S_2

Graphical Model



- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: R_1, R_2
- Treatments: W_1, W_2
- Observed outcomes: $Y = Y(W)$
- Selective choice: S_1, S_2

Graphical Model



- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: R_1, R_2
- Treatments: W_1, W_2
- Observed outcomes: $Y = Y(W)$
- Selective choice: S_1, S_2

Randomization Inference

Randomization Inference

- Strength: no modelling assumptions, no i.i.d. data
- Distribution of $W = (W_1, W_2)$ is known

Randomization Inference

- Strength: no modelling assumptions, no i.i.d. data
- Distribution of $W = (W_1, W_2)$ is known
- Null hypothesis: $Y_i(1) - Y_i(0) = 0$ for all/subset of units

Randomization Inference

- Strength: no modelling assumptions, no i.i.d. data
- Distribution of $W = (W_1, W_2)$ is known
- Null hypothesis: $Y_i(1) - Y_i(0) = 0$ for all/subset of units
- Condition on $Z = (R, X_R, Y_R(\,\cdot\,))$ and compare observed value of statistic $T(W, Z)$ against values $T(W^*, Z)$ under alternative treatment assignments W^*.
- $\mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z)$ $\qquad W^* \stackrel{D}{=} W$ and $W^* \perp\!\!\!\perp W \mid Z$

Fisher (1935), Zhang & Zhao (2023)

Randomization Inference

- Strength: no modelling assumptions, no i.i.d. data
- Distribution of $W = (W_1, W_2)$ is known
- Null hypothesis: $Y_i(1) - Y_i(0) = 0$ for all/subset of units
- Condition on $Z = (R, X_R, Y_R(\cdot))$ and compare observed value of statistic $T(W, Z)$ against values $T(W^*, Z)$ under alternative treatment assignments W^*.
- $\mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z)$ $\qquad W^* \overset{D}{=} W$ and $W^* \perp\!\!\!\perp W \mid Z$
- Problem: double dipping

Fisher (1935), Zhang & Zhao (2023)

Selective Randomization Inference

Selective Randomization Inference

- Traditional statistics: model & null hypothesis → data → inference
- Here: data → model & null hypothesis → inference

Selective Randomization Inference

- Traditional statistics: model & null hypothesis → data → inference
- Here: data → model & null hypothesis → inference
- Type-I error inflation: comparing to W_1^* that choose different stage-II design than W_1

Selective Randomization Inference

- Traditional statistics: model & null hypothesis → data → inference
- Here: data → model & null hypothesis → inference
- Type-I error inflation: comparing to W_1^* that choose different stage-II design than W_1
- Solutions:
    - Data splitting (Cox, 1975): $\mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z, W_1^* = W_1)$

Selective Randomization Inference

- Traditional statistics: model & null hypothesis → data → inference
- Here: data → model & null hypothesis → inference
- Type-I error inflation: comparing to W_1^* that choose different stage-II design than W_1
- Solutions:
  - Data splitting (Cox, 1975): $\mathbb{P}^*(T(W^*, Z) \leq T(W, Z) \mid W, Z, W_1^* = W_1)$
  - Selective inference (Lee et al., 2016; Fithian et al., 2017): regression models etc.

Selective Randomization Inference

- Traditional statistics: model & null hypothesis → data → inference
- Here: data → model & null hypothesis → inference
- Type-I error inflation: comparing to W_1^* that choose different stage-II design than W_1
- Solutions:
    - Data splitting (Cox, 1975): $\mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z, W_1^* = W_1)$
    - Selective inference (Lee et al., 2016; Fithian et al., 2017): regression models etc.
    - Selective randomization inference:

$$p_S(W) = \mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z,\ S(W_1^*) = S(W_1)\ )$$

Inference

$$p_S(W) = \mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z,\ S(W^*) = S(W)\ )$$

Inference

$$p_S(W) = \mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z,\ S(W^*) = S(W)\ )$$

- General framework for L treatments and K stages

Inference

$$p_S(W) = \mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z,\ S(W^*) = S(W)\ )$$

- General framework for L treatments and K stages
- Confidence intervals:
  - test $Y_i(1) - Y_i(0) = \beta$ for different β
  - $(1 - \alpha)$ confidence interval: $\{\beta \colon p_S^\beta(W) \geq \alpha\}$

Inference

$$p_S(W) = \mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z,\ S(W^*) = S(W)\ )$$

- General framework for L treatments and K stages
- Confidence intervals:
  - test $Y_i(1) - Y_i(0) = \beta$ for different β
  - $(1 - \alpha)$ confidence interval: $\{\beta : p_S^\beta(W) \geq \alpha\}$
- Estimation: β such that $p_S^\beta(W) = 0.5$

Inference

$$p_S(W) = \mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z,\ S(W^*) = S(W)\ )$$

- General framework for L treatments and K stages
- Confidence intervals:
  - test $Y_i(1) - Y_i(0) = \beta$ for different β
  - $(1 - \alpha)$ confidence interval: $\{\beta : p_S^\beta(W) \geq \alpha\}$
- Estimation: β such that $p_S^\beta(W) = 0.5$
- Data carving: non-adaptive hold out units

Computation

$$p_S(W) = \mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z,\ S(W^*) = S(W)\ )$$

Computation

$$p_S(W) = \mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z,\ S(W^*) = S(W)\ )$$

- Monte Carlo approximation: Generate m feasible samples $(w_j^*)_{j=1}^m$, i.e. $S(w) = S(w_j^*)$, and compute

$$\frac{\sum_{i=1}^m \mathbf{1}\{T(w_j^*, Z) \leq T(w, Z)\}\ \mathbb{P}^*(W^* = w^* \mid Z)}{\sum_{i=1}^m \mathbb{P}^*(W^* = w^* \mid Z)}$$

Computation

$$p_S(W) = \mathbb{P}^*(\ T(W^*, Z) \leq T(W, Z) \mid W, Z,\ S(W^*) = S(W)\ )$$

- Monte Carlo approximation: Generate m feasible samples $(w_j^*)_{j=1}^m$, i.e. $S(w) = S(w_j^*)$, and compute

$$\frac{\sum_{i=1}^m \mathbf{1}\{T(w_j^*, Z) \leq T(w, Z)\}\ \mathbb{P}^*(W^* = w^* \mid Z)}{\sum_{i=1}^m \mathbb{P}^*(W^* = w^* \mid Z)}$$

- Rejection sampling, Markov Chain Monte Carlo (MCMC)

Simulation Study

Simulation Study

- 2 stages, 2 treatments $W_i \in \{0,1\}$, 2 groups $X_i \in \{0,1\}$
- Potential outcomes: $Y_i(0) = Y_i(1) \sim N(0,1)$ i.i.d.
- First stage: 50 patients

Simulation Study

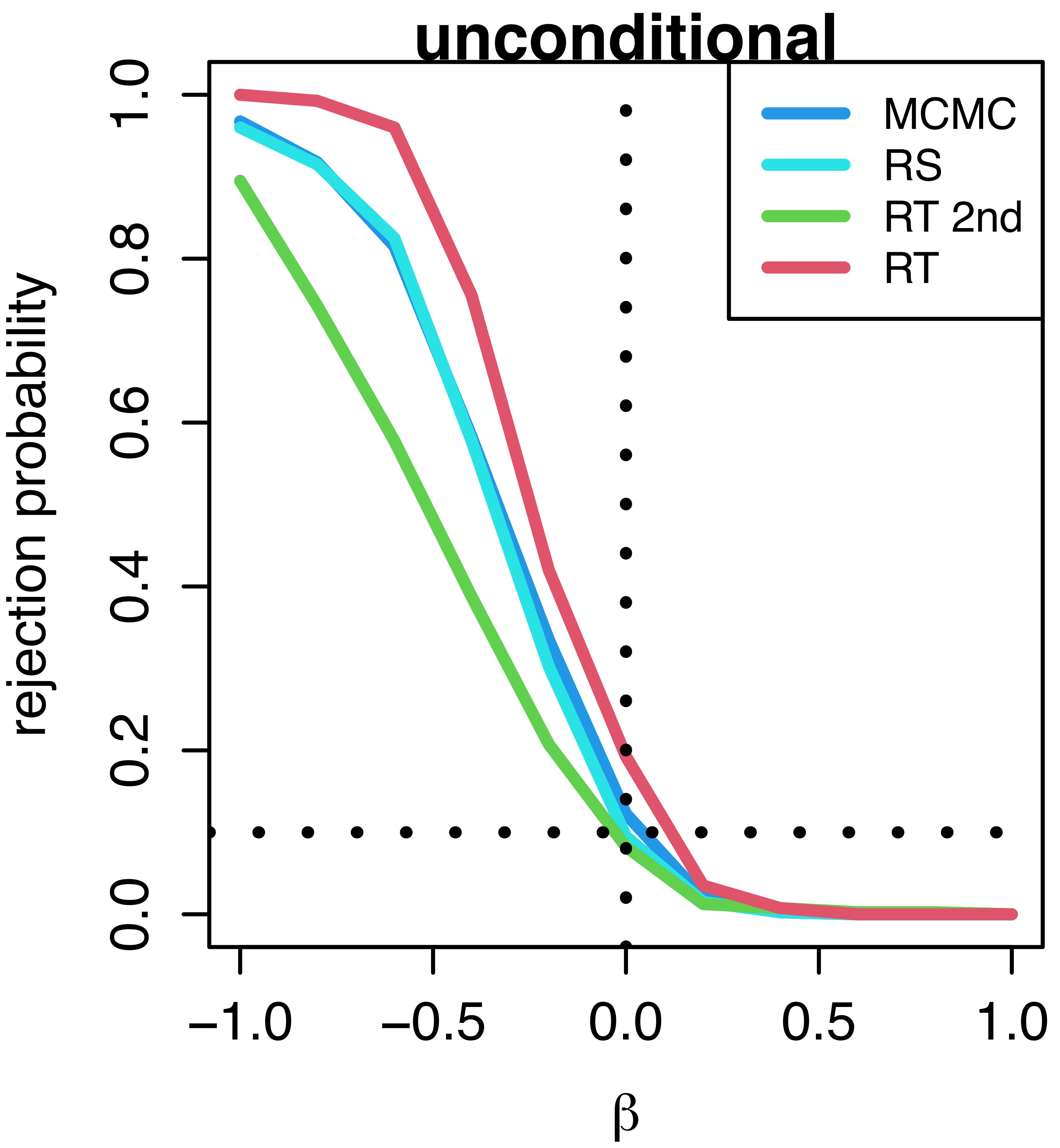
- 2 stages, 2 treatments $W_i \in \{0,1\}$, 2 groups $X_i \in \{0,1\}$
- Potential outcomes: $Y_i(0) = Y_i(1) \sim N(0,1)$ i.i.d.
- First stage: 50 patients
- $\Delta =$ standardized difference in SATEs between groups

Simulation Study

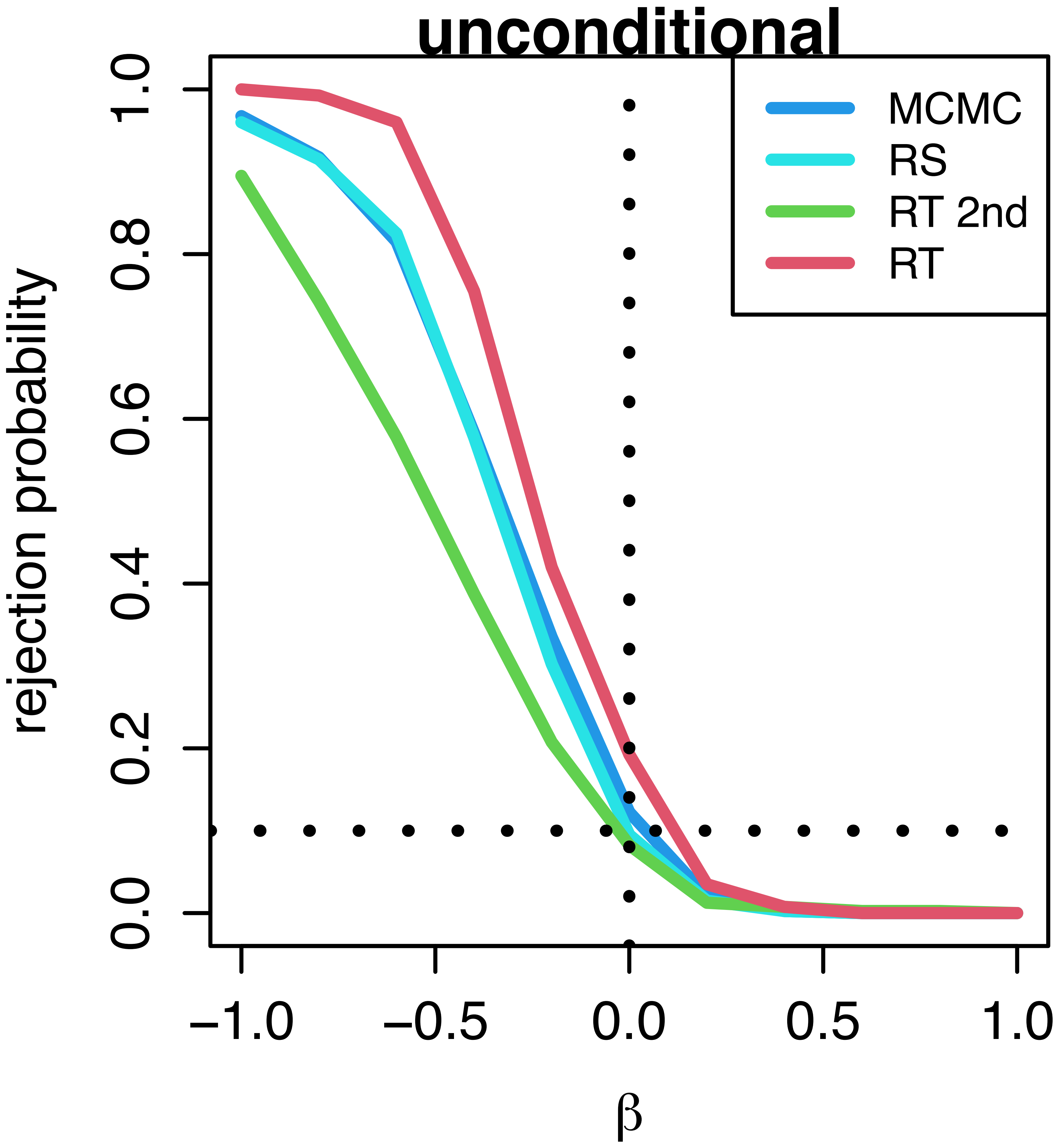
- 2 stages, 2 treatments $W_i \in \{0,1\}$, 2 groups $X_i \in \{0,1\}$
- Potential outcomes: $Y_i(0) = Y_i(1) \sim N(0,1)$ i.i.d.
- First stage: 50 patients
- $\Delta =$ standardized difference in SATEs between groups
- Selection variable:

$$S = \begin{cases} 0, & \Phi(0.2) \leq \Delta \leq \Phi(0.8), & \text{recruit 13/12 in stage II,} \\ 1, & \Delta < \Phi(0.2), & \text{recruit 25 from group 1 in stage II,} \\ 2, & \Delta > \Phi(0.8), & \text{recruit 25 from group 2 in stage II.} \end{cases}$$

Power Analysis

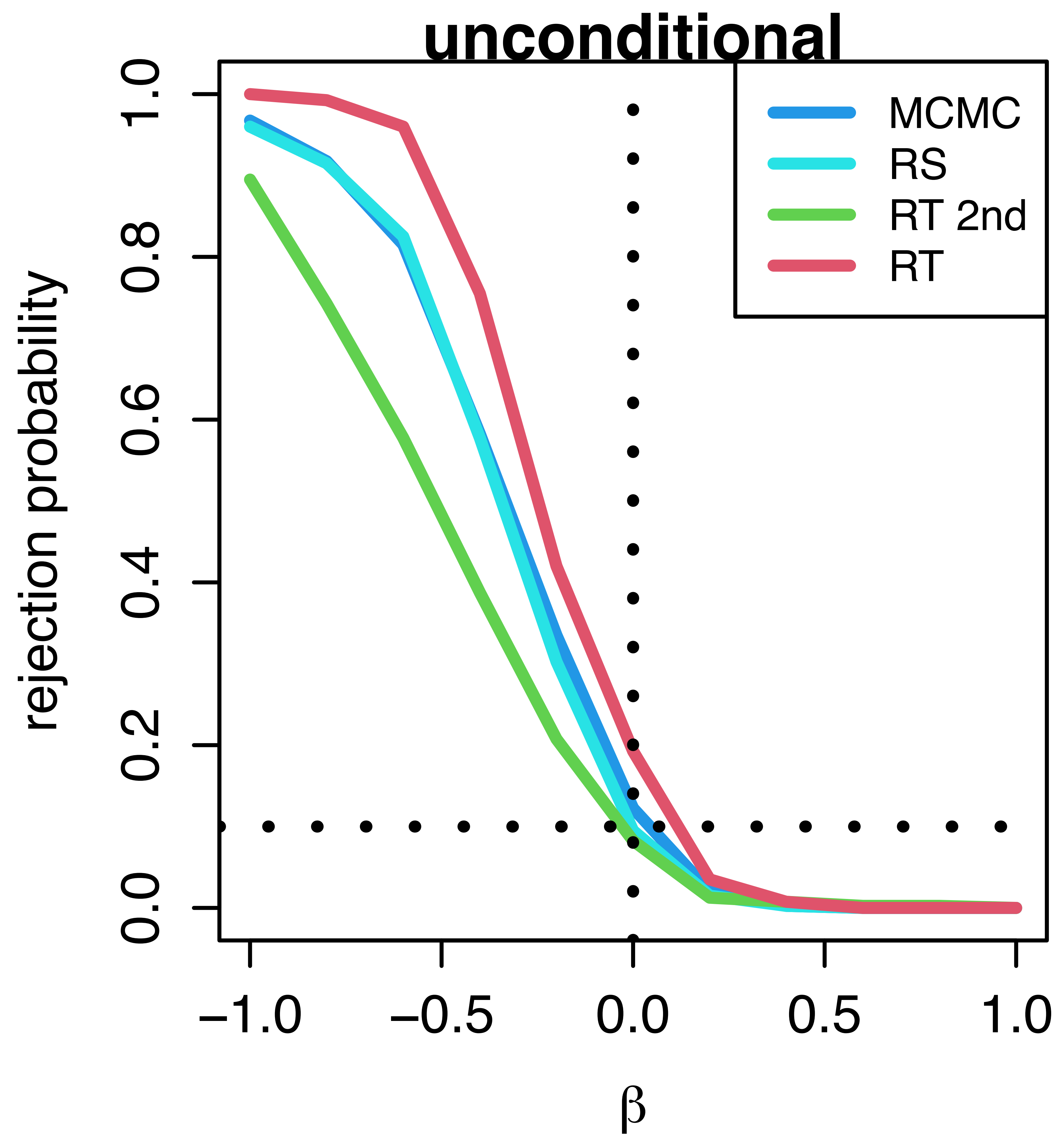


Power Analysis



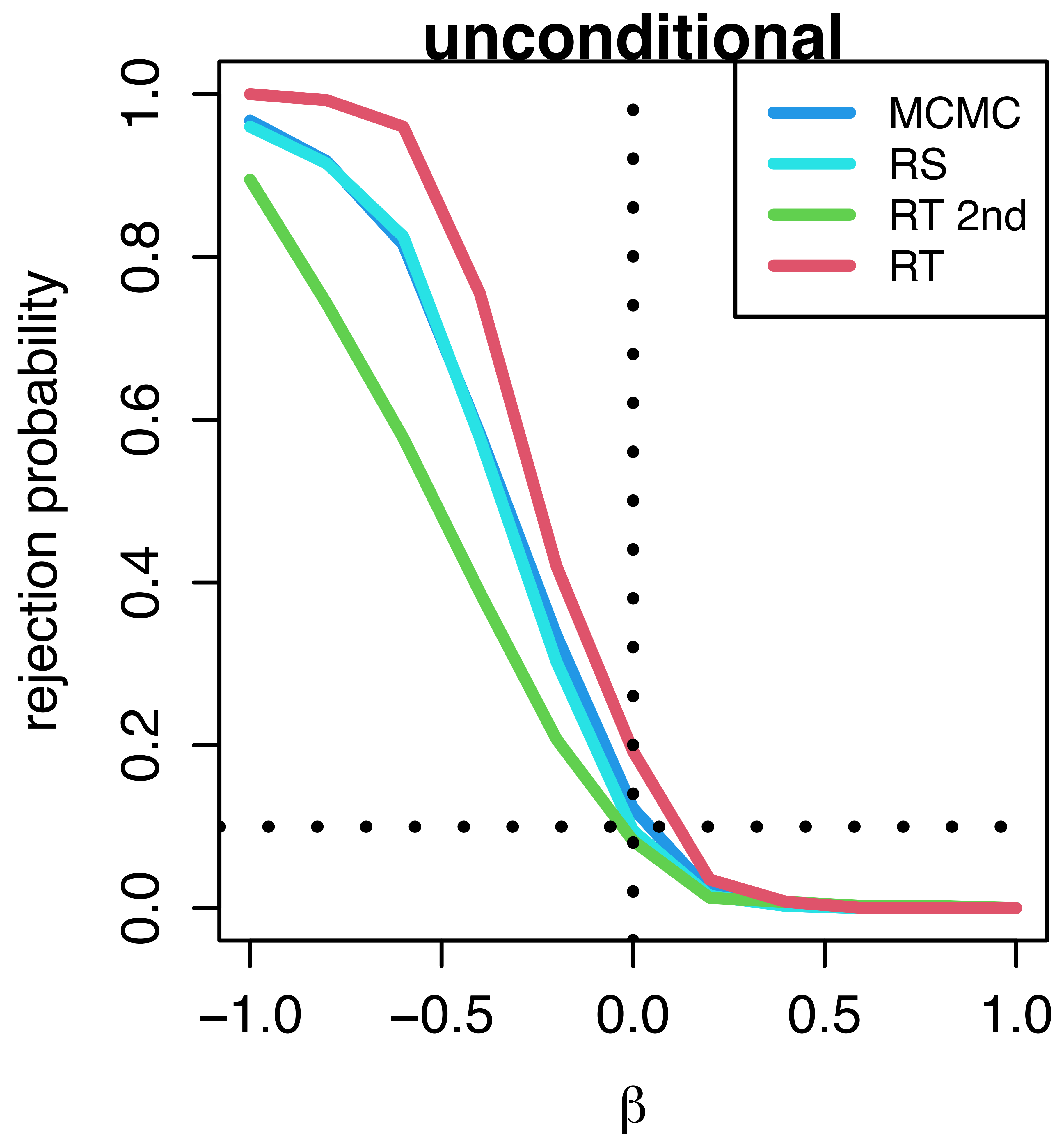
- Standard randomization inference does not control type-I error.
- Randomization inference on 2nd stage is valid but has low power.

Power Analysis



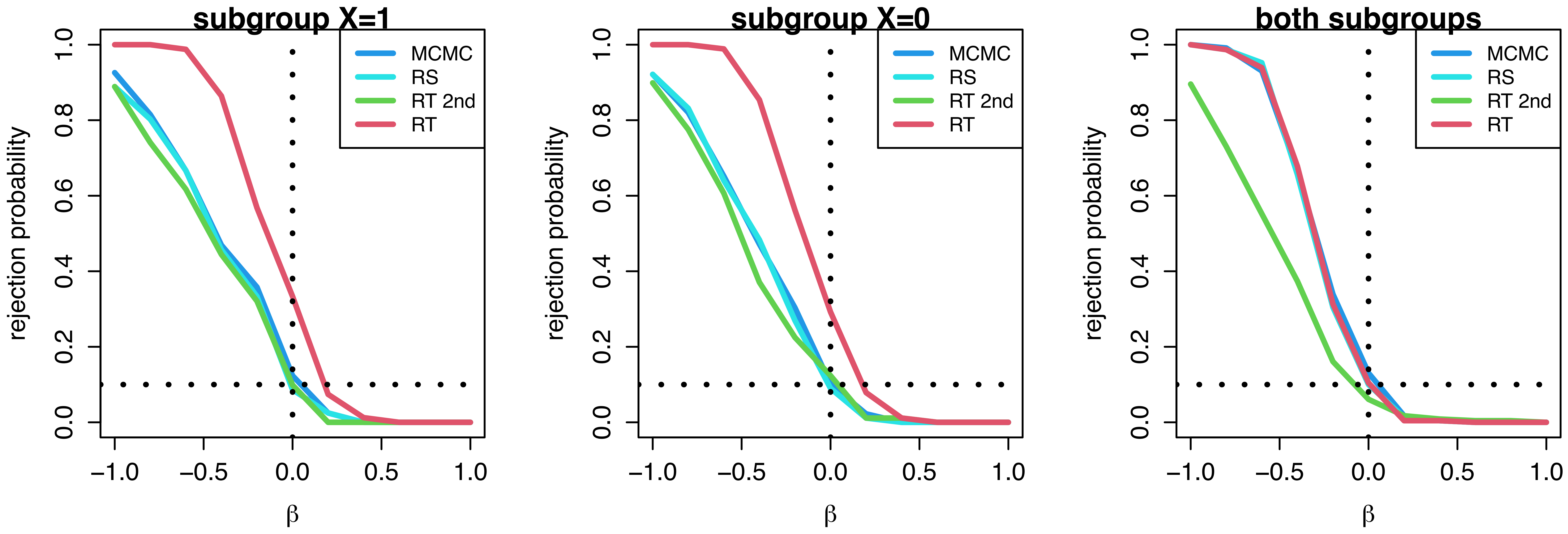
- Standard randomization inference does not control type-I error.
- Randomization inference on 2nd stage is valid but has low power.
- Selective randomization inference is valid and more powerful.

Power Analysis

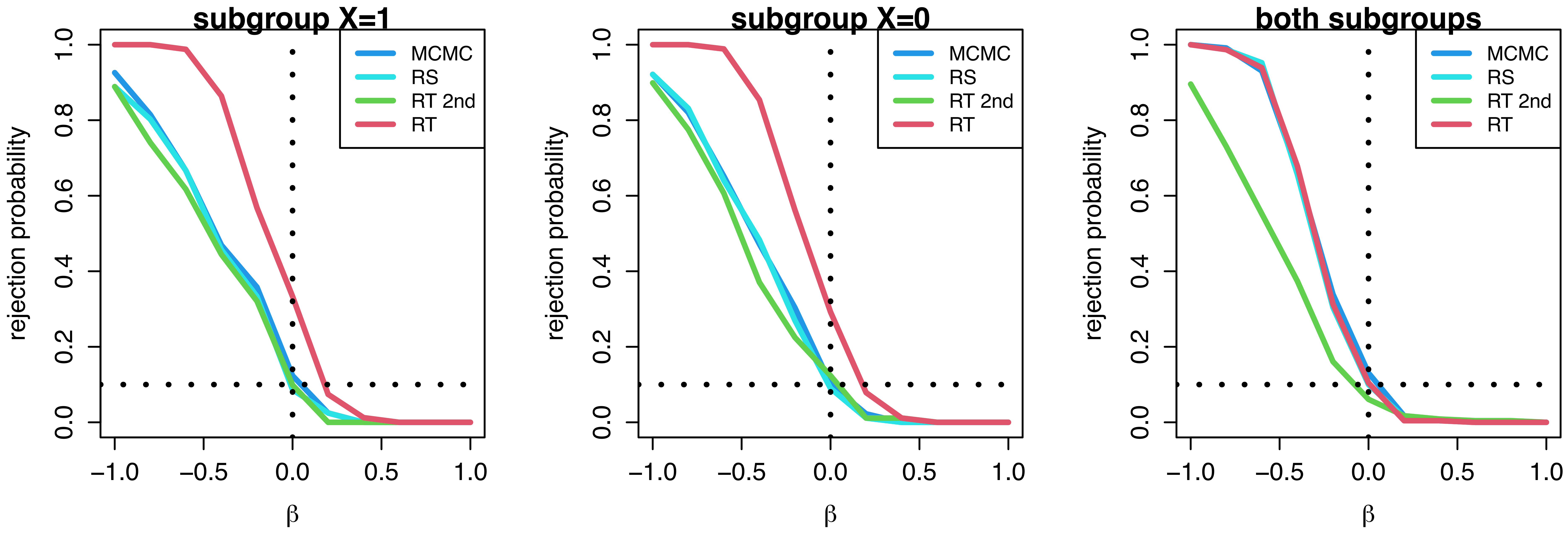


- Standard randomization inference does not control type-I error.
- Randomization inference on 2nd stage is valid but has low power.
- Selective randomization inference is valid and more powerful.
- Rejection sampling and MCMC lead to very similar approximations.

Power Analysis

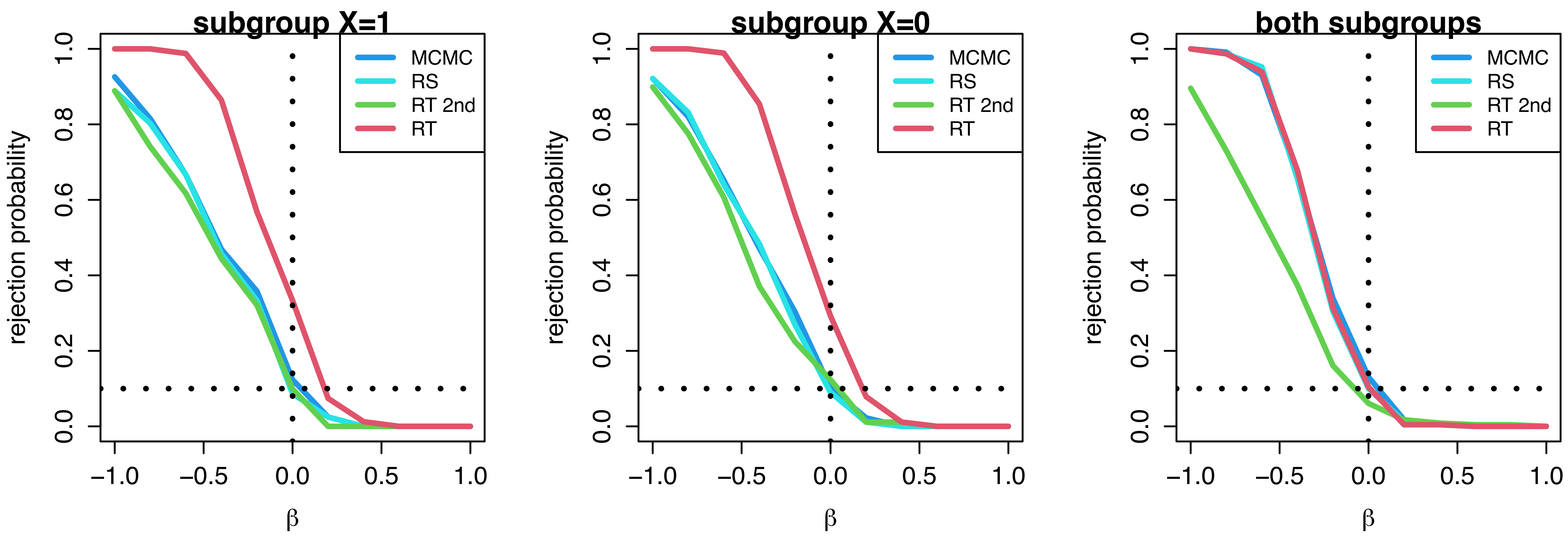


Power Analysis



- Type-I error control in every subgroup

Power Analysis



- Type-I error control in every subgroup
- Gain in power when there is a lot of “randomness left”

Thanks for your attention!

taf40@cam.ac.uk

References

Cox, D.R. (1975) ‘A note on data-splitting for the evaluation of significance levels’, Biometrika, 62(2), pp. 441–444.

Fisher, R. A. (1935). ‘The design of experiments’, Edinburgh: Oliver & Boyd.

Fithian, W., Sun, D. and Taylor, J. (2017) ‘Optimal Inference After Model Selection’, arXiv:1410.2597

Lee, J.D., Sun, D.L., Sun, Y., Taylor J. (2016) ‘Exact post-selection inference, with application to the lasso’, The Annals of Statistics, 44(3).

Zhang, Y. and Zhao, Q. (2023) ‘What is a Randomization Test?’, Journal of the American Statistical Association, 0(0), pp. 1–15.