Selective Randomization Inference for Adaptive Experiments

Tobias Freidling

École polytechnique fédérale de Lausanne

University of Cambridge

ICORS 2025 — 22/05/2025

Collaborators

Qingyuan Zhao

Statistical Laboratory, DPMMS
University of Cambridge

Zijun Gao

Marshall School of Business
University of Southern California

Randomization Inference

Randomization Inference

- Randomized controlled trial

Fisher (1935), Pitman (1937), Zhang & Zhao (2023), Imbens & Rubin (2015)

Randomization Inference

- Randomized controlled trial
- Finite population perspective

Fisher (1935), Pitman (1937), Zhang & Zhao (2023),
Imbens & Rubin (2015)

Randomization Inference

- Randomized controlled trial
- Finite population perspective
- Inference without modelling- or i.i.d. data- assumptions

Fisher (1935), Pitman (1937), Zhang & Zhao (2023), Imbens & Rubin (2015)

Randomization Inference

- Randomized controlled trial
- Finite population perspective
- Inference without modelling- or i.i.d. data- assumptions
- Dataset: $(Y_i, Z_i, X_i)_{i=1,\ldots,n}$

Fisher (1935), Pitman (1937), Zhang & Zhao (2023), Imbens & Rubin (2015)

Randomization Inference

- Randomized controlled trial
- Finite population perspective
- Inference without modelling- or i.i.d. data- assumptions
- Dataset: $(Y_i, Z_i, X_i)_{i=1,\ldots,n}$
- Potential outcomes: $Y_i(0), Y_i(1)$ $\rightarrow$ Realized outcomes: $Y_i = Y_i(Z_i)$

Fisher (1935), Pitman (1937), Zhang & Zhao (2023), Imbens & Rubin (2015)

Randomization Inference

- Randomized controlled trial
- Finite population perspective
- Inference without modelling- or i.i.d. data- assumptions
- Dataset: $(Y_i, Z_i, X_i)_{i=1,\ldots,n}$
- Potential outcomes: $Y_i(0), Y_i(1)$ $\rightarrow$ Realized outcomes: $Y_i = Y_i(Z_i)$
- Distribution of Z is **known** and $Z \perp\!\!\!\perp Y(\,\cdot\,) \mid X$

Fisher (1935), Pitman (1937), Zhang & Zhao (2023),
Imbens & Rubin (2015)

Randomization Inference

Fisher (1935), Pitman (1937), Zhang & Zhao (2023)

Randomization Inference

i	Y	Y(0)	Y(1)	Z
1	5		5	1
2	7	7		0
3	-3	-3		0
4	0		0	1
…	…	…	…	…

Fisher (1935), Pitman (1937), Zhang & Zhao (2023)

Randomization Inference

- Null hypothesis: $Y_i(1) - Y_i(0) = 0$ for all i
- Test statistic: $T(Z, Y(\cdot))$, e.g. average outcome of treated minus control

i	Y	Y(0)	Y(1)	Z
1	5		5	1
2	7	7		0
3	-3	-3		0
4	0		0	1
…	…	…	…	…

Fisher (1935), Pitman (1937), Zhang & Zhao (2023)

Randomization Inference

- Null hypothesis: $Y_i(1) - Y_i(0) = 0$ for all i
- Test statistic: $T(Z, Y(\cdot))$, e.g. average outcome of treated minus control

i	Y	Y(0)	Y(1)	Z
1	5	5	5	1
2	7	7	7	0
3	-3	-3	-3	0
4	0	0	0	1
…	…	…	…	…

Fisher (1935), Pitman (1937), Zhang & Zhao (2023)

Randomization Inference

- Null hypothesis: $Y_i(1) - Y_i(0) = 0$ for all i
- Test statistic: $T(Z, Y(\cdot))$, e.g. average outcome of treated minus control
- Condition on $Y(\cdot)$ and compare observed value of statistic $T(Z, Y(\cdot))$ against values $T(Z^*, Y(\cdot))$ under alternative treatment assignments Z^*.

i	Y	Y(0)	Y(1)	Z
1	5	5	5	1
2	7	7	7	0
3	-3	-3	-3	0
4	0	0	0	1
…	…	…	…	…

Fisher (1935), Pitman (1937), Zhang & Zhao (2023)

Randomization Inference

- Null hypothesis: $Y_i(1) - Y_i(0) = 0$ for all i
- Test statistic: $T(Z, Y(\cdot))$, e.g. average outcome of treated minus control
- Condition on $Y(\cdot)$ and compare observed value of statistic $T(Z, Y(\cdot))$ against values $T(Z^*, Y(\cdot))$ under alternative treatment assignments Z^*.
- P-value:

$$\mathbb{P}(\, T(Z^*, Y(\cdot)) \leq T(Z, Y(\cdot)) \mid Y(\cdot), Z),$$

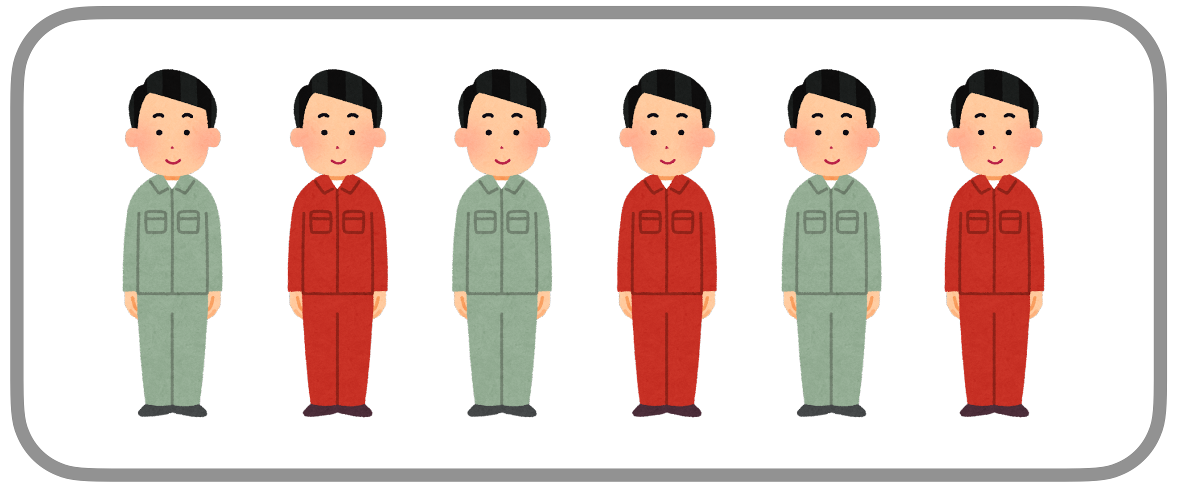
where $Z^* \overset{D}{=} Z$ and $Z^* \perp\!\!\!\perp Z \mid Y(\cdot)$

i	Y	Y(0)	Y(1)	Z
1	5	5	5	1
2	7	7	7	0
3	-3	-3	-3	0
4	0	0	0	1
…	…	…	…	…

Fisher (1935), Pitman (1937), Zhang & Zhao (2023)

Example

Example

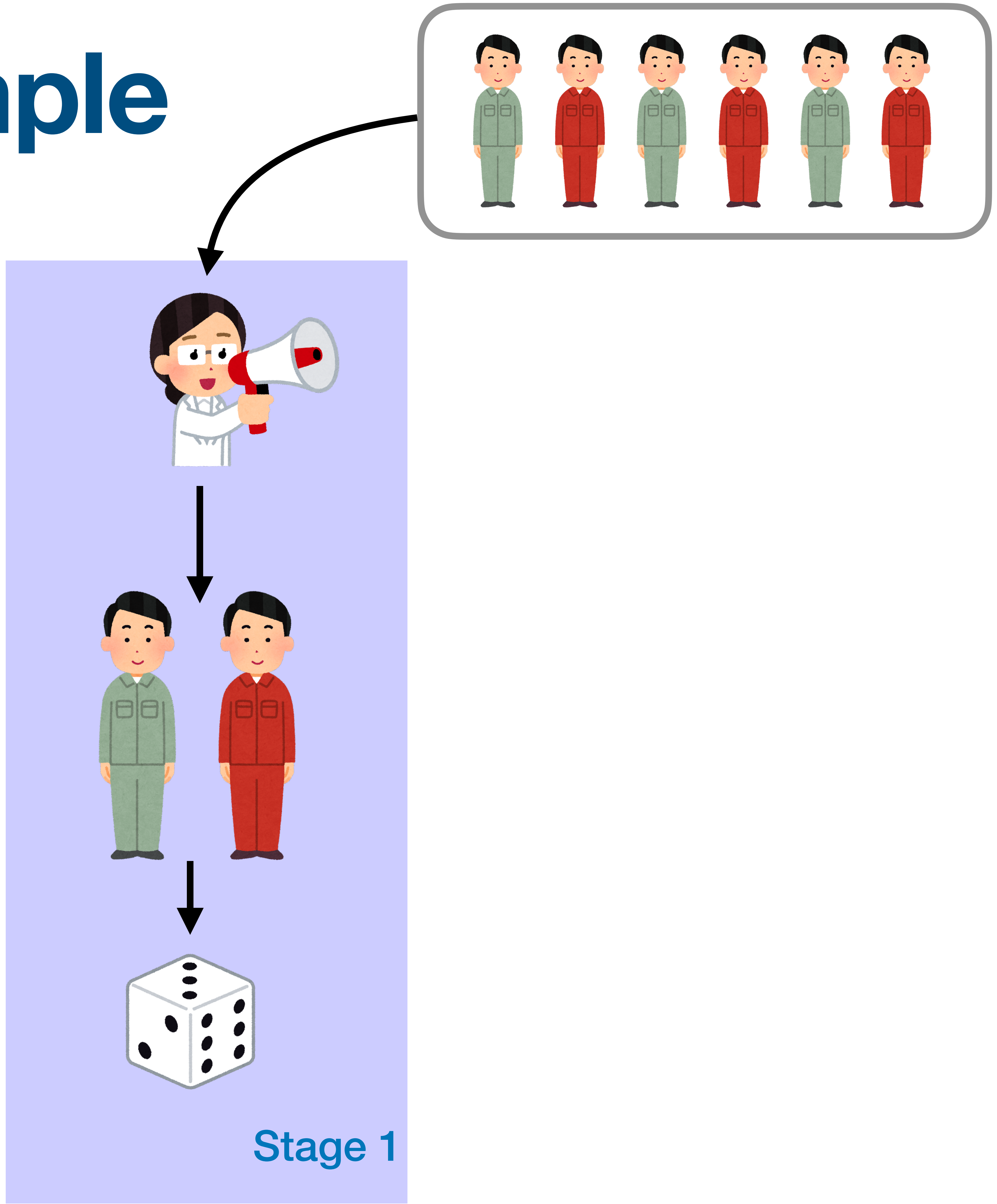



High genetic risk


Low genetic risk

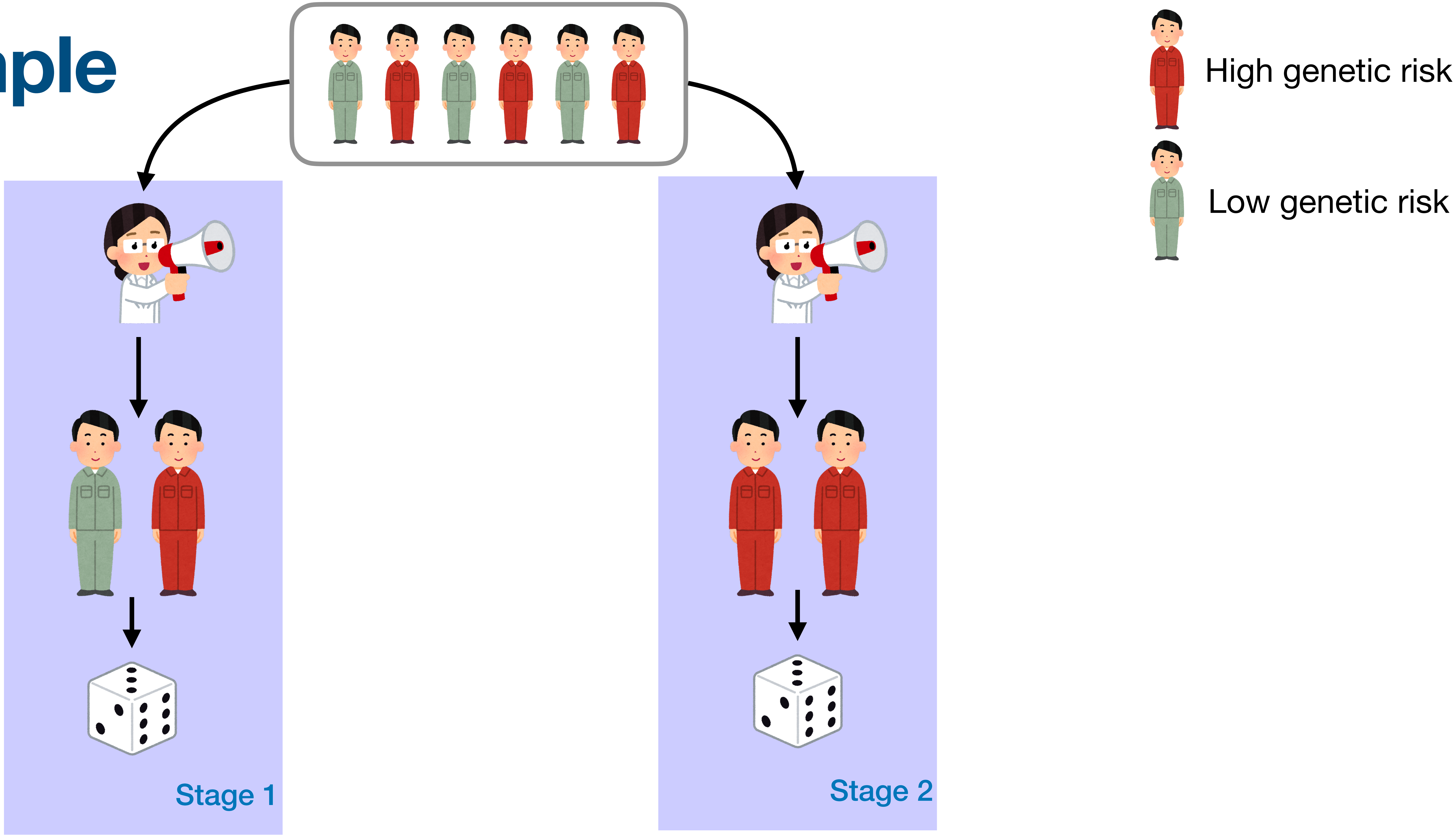
FOURIER trial: Marston et al. (2020)

Example



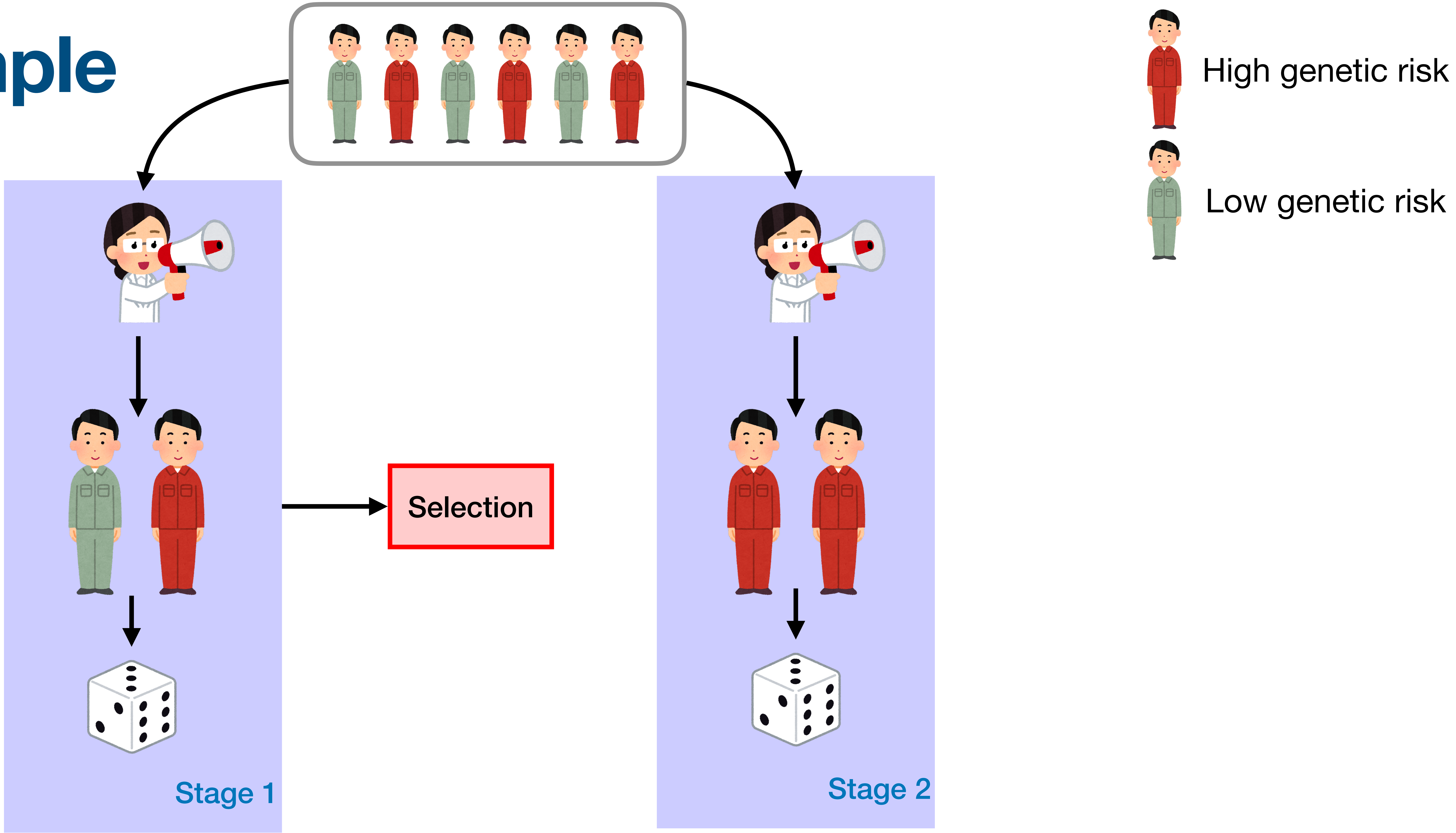
FOURIER trial: Marston et al. (2020)

Example



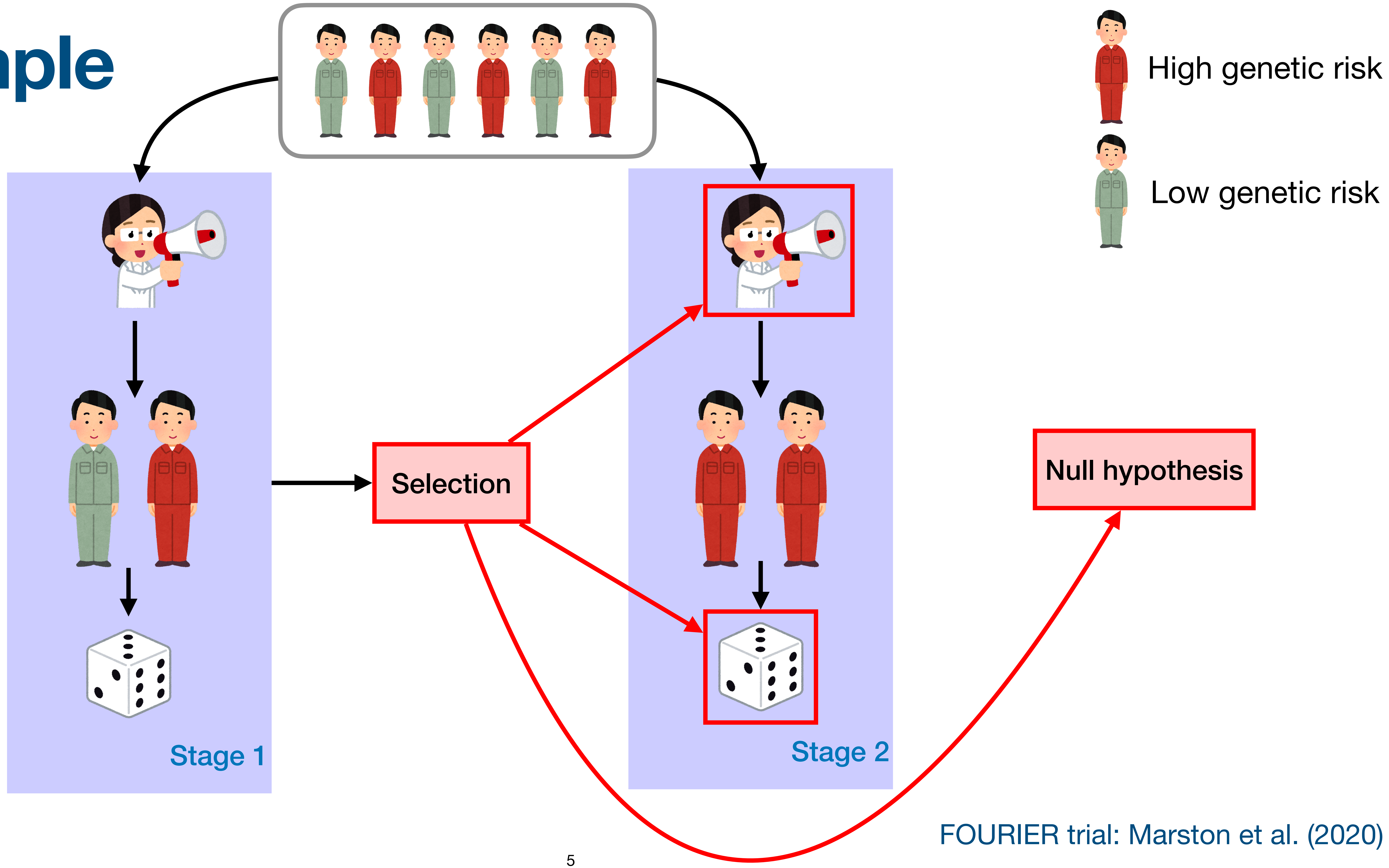
FOURIER trial: Marston et al. (2020)

Example



FOURIER trial: Marston et al. (2020)

Example



FOURIER trial: Marston et al. (2020)

Graphical Model

Graphical Model

Pearl (2009)

Graphical Model

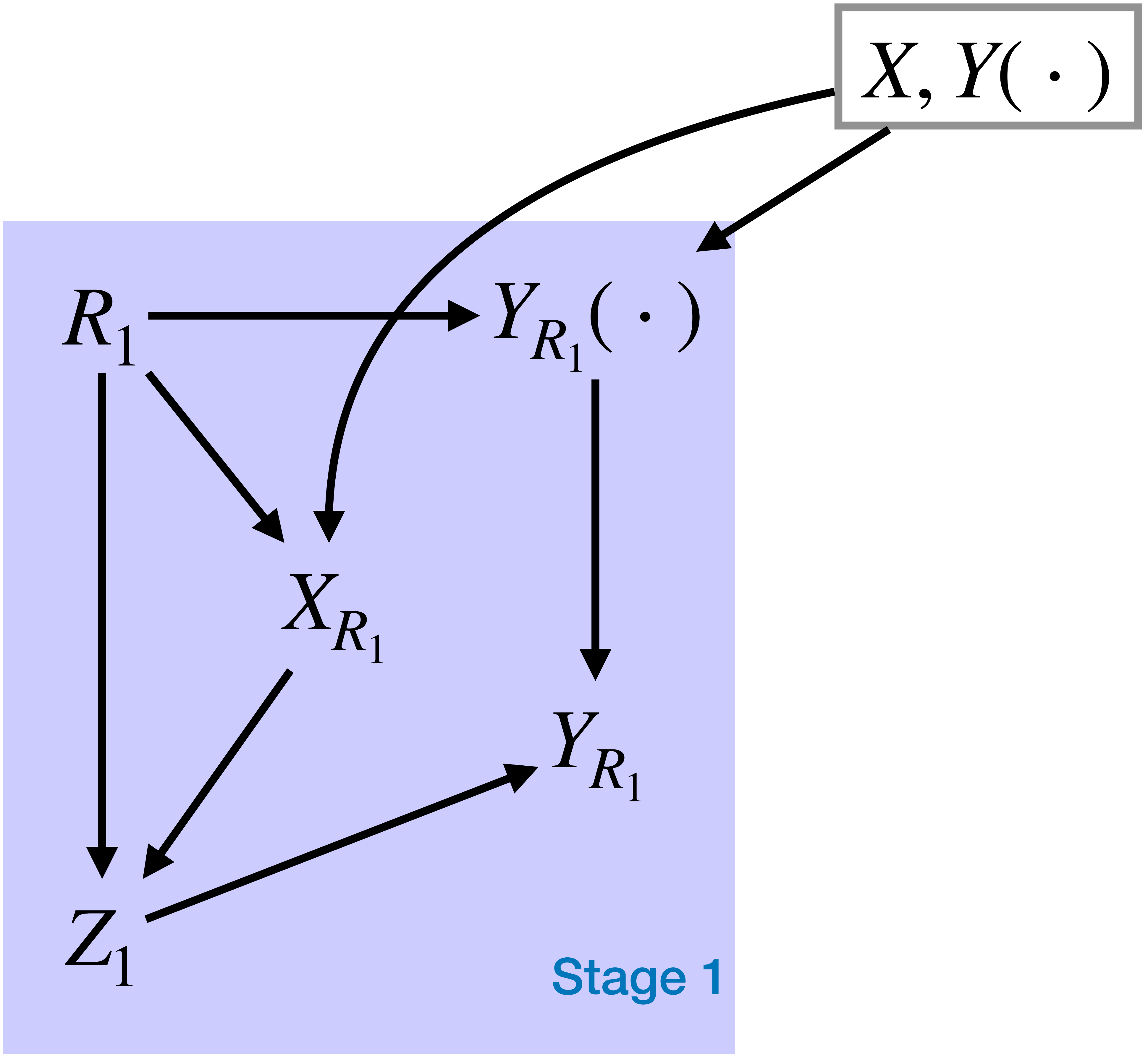
- Covariates: X
- Potential outcomes: $Y(\cdot)$

$$\boxed{X, Y(\cdot)}$$

Pearl (2009)

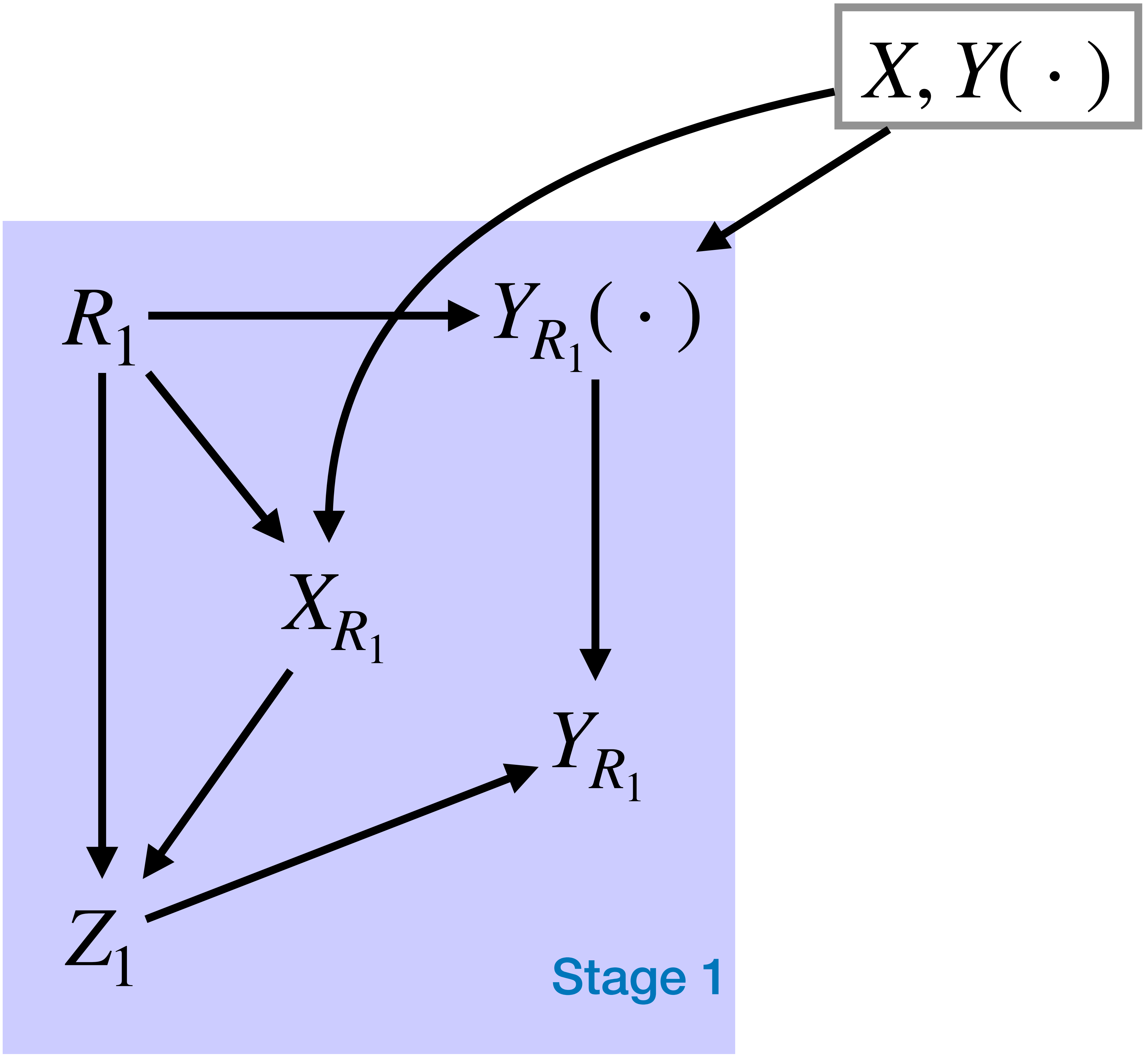
Graphical Model

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



Pearl (2009)

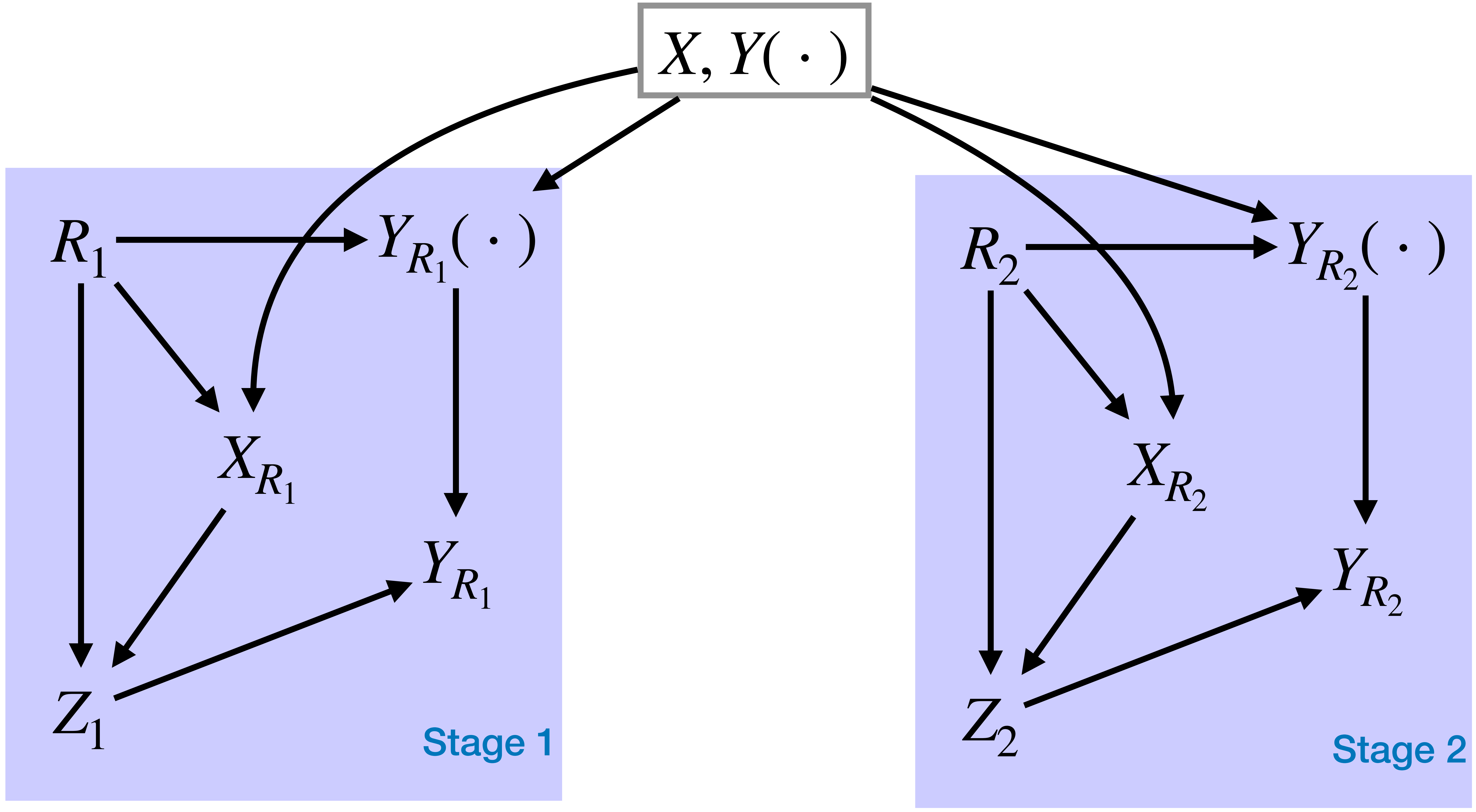
Graphical Model



- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: $R_k \subseteq [n]$
- Treatments: Z_k
- Observed outcomes: $Y = Y(Z)$

Pearl (2009)

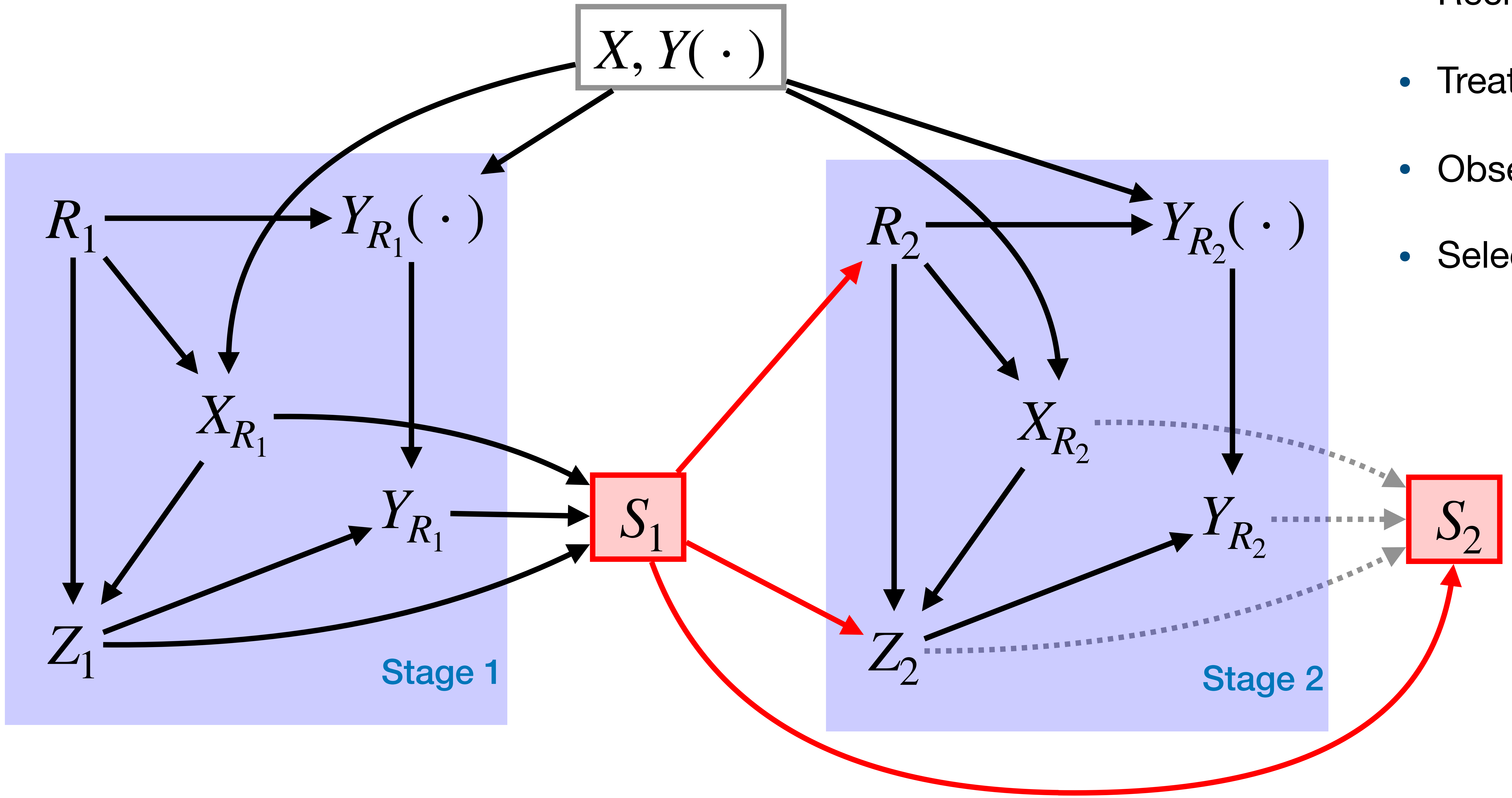
Graphical Model



- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: $R_k \subseteq [n]$
- Treatments: Z_k
- Observed outcomes: $Y = Y(Z)$

Pearl (2009)

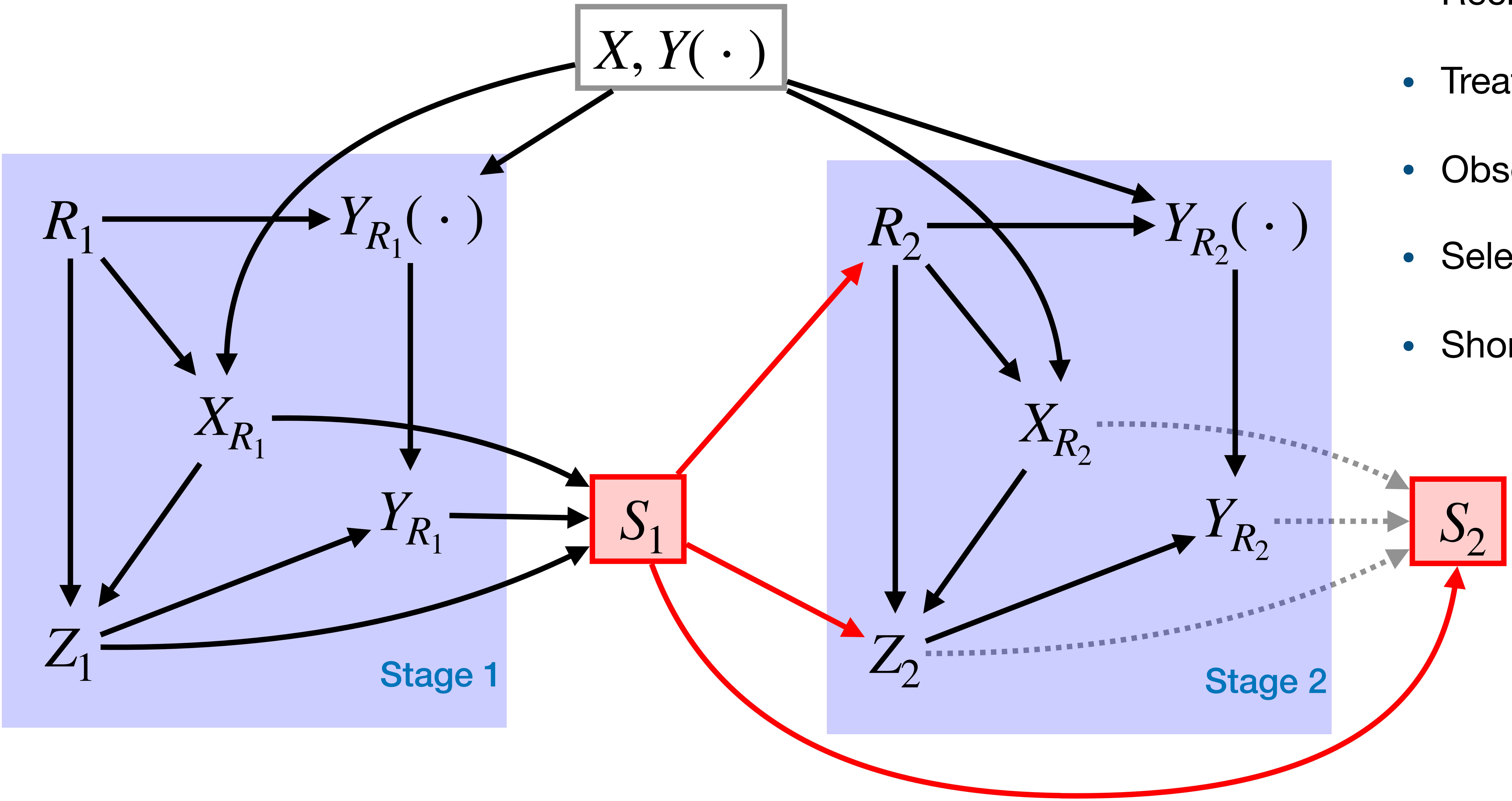
Graphical Model



- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: $R_k \subseteq [n]$
- Treatments: Z_k
- Observed outcomes: $Y = Y(Z)$
- Selective choice: S_k

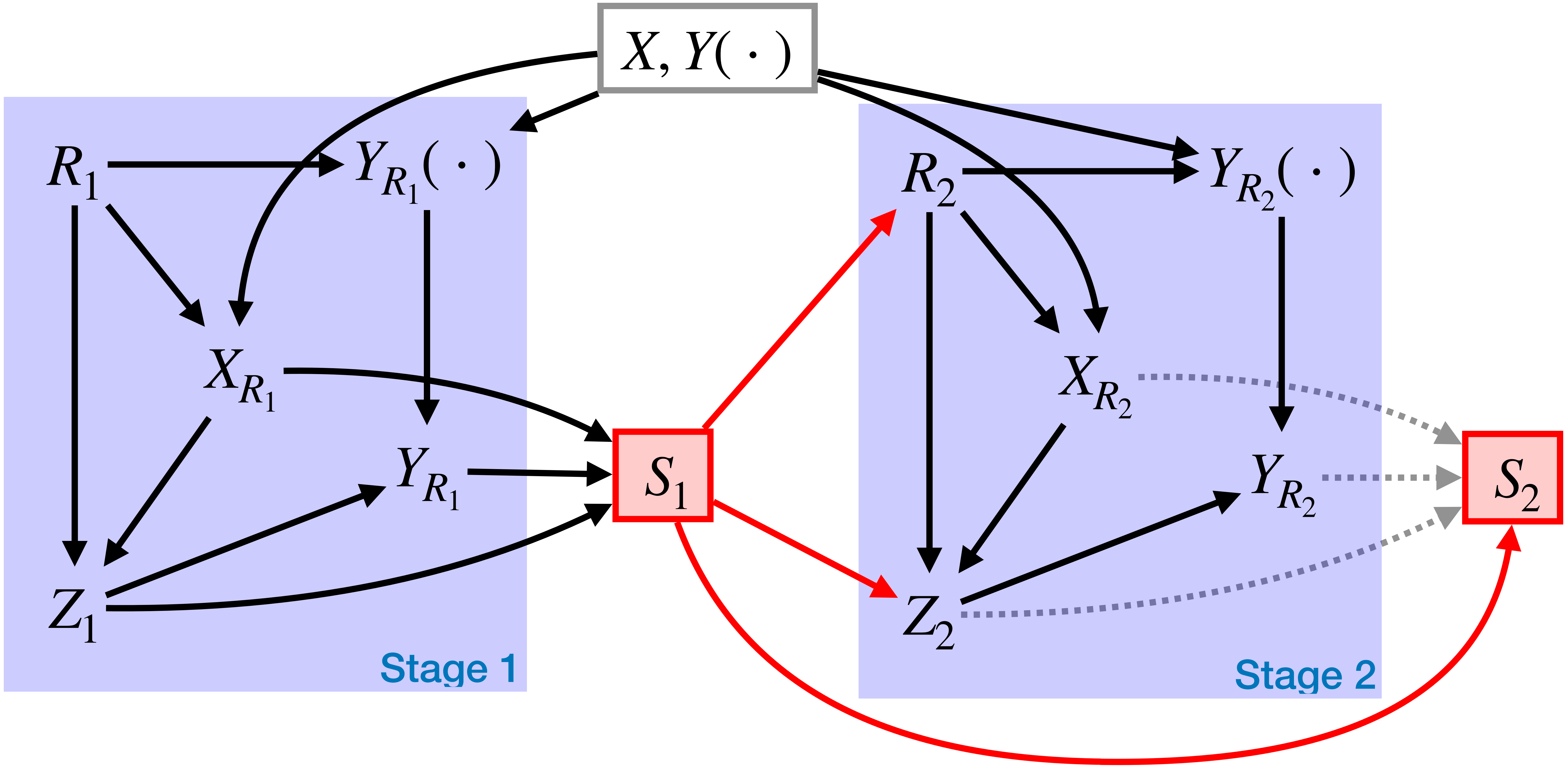
Pearl (2009)

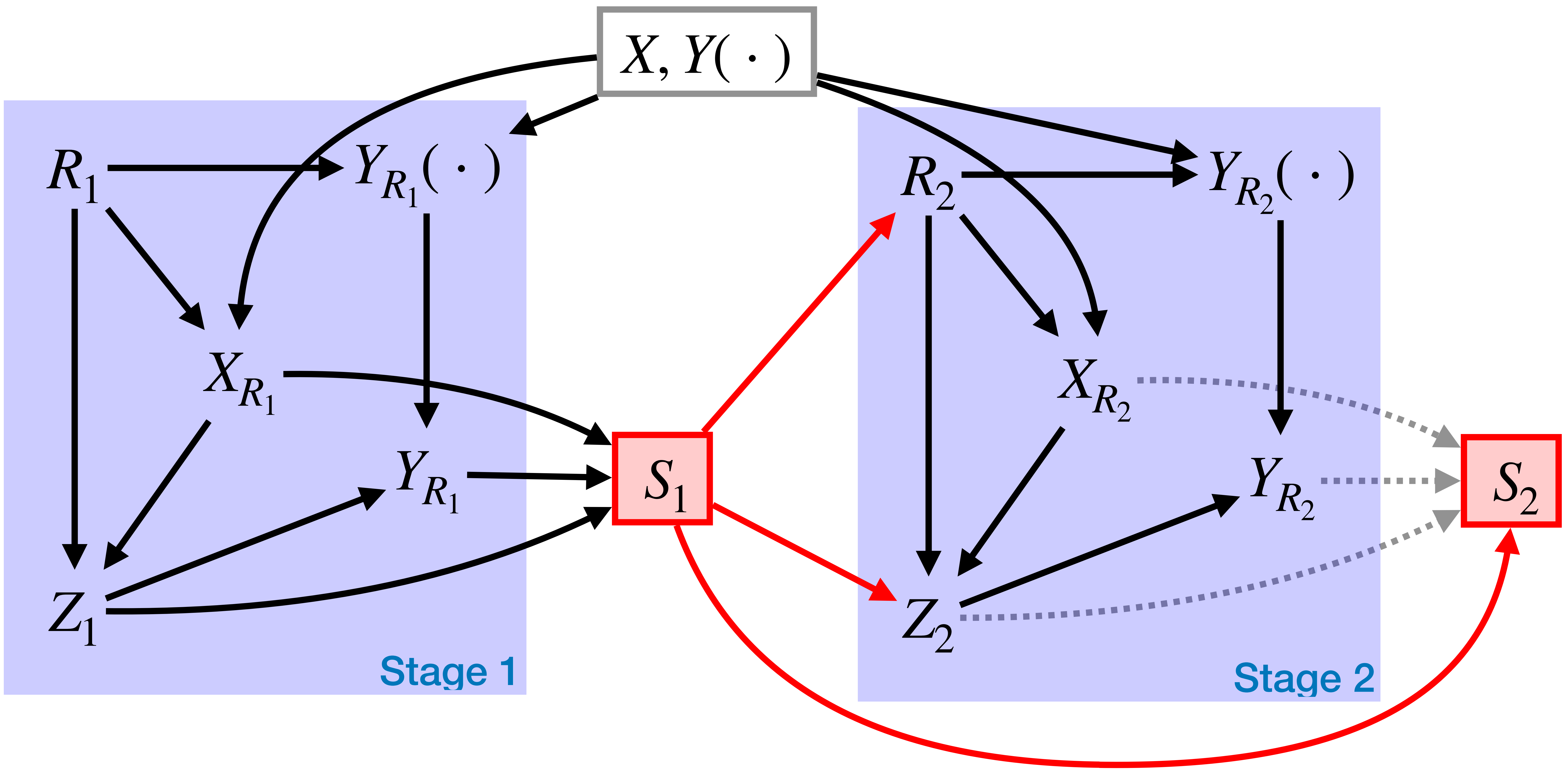
Graphical Model



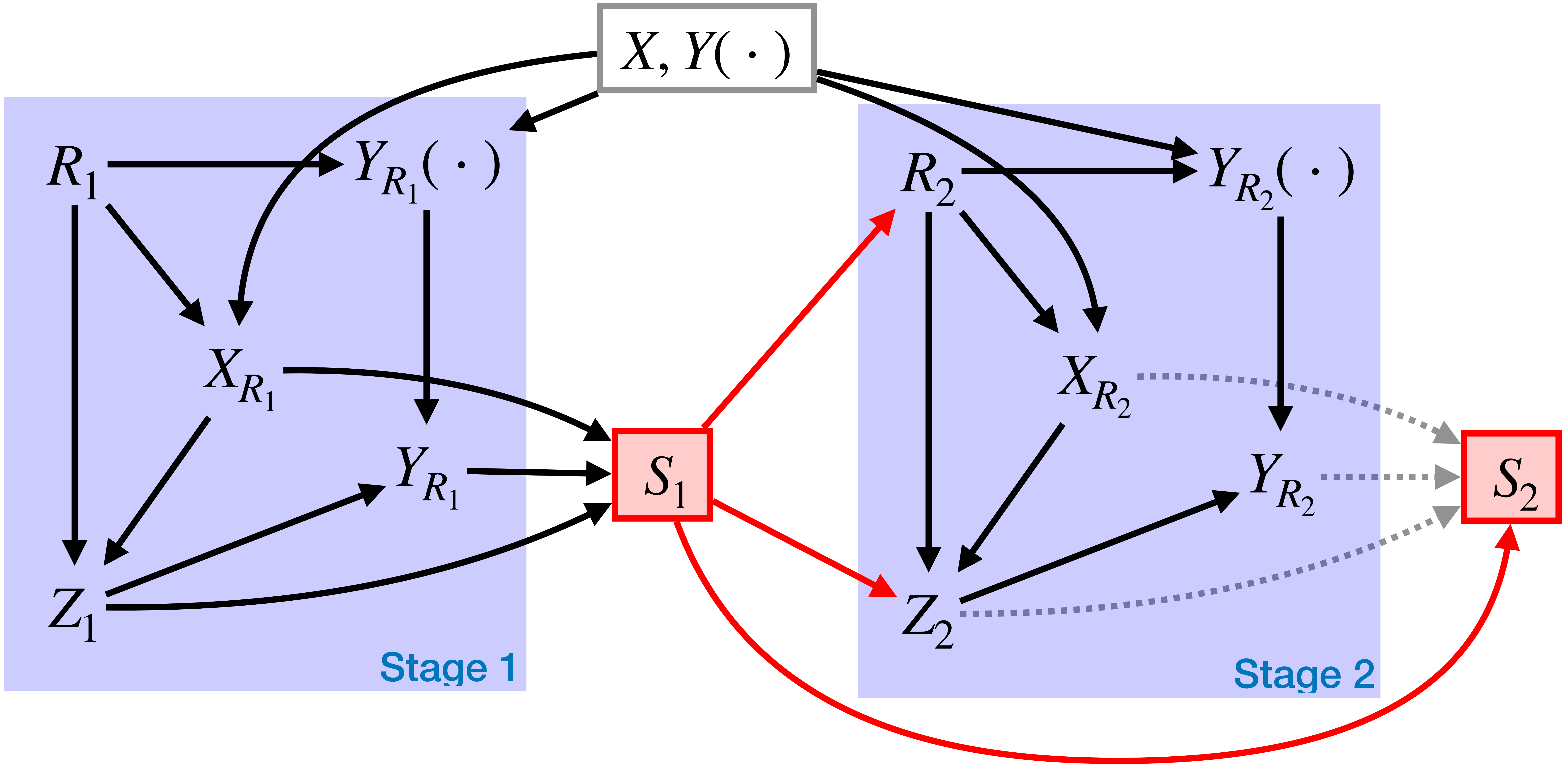
- Covariates: X
- Potential outcomes: $Y(\cdot)$
- Recruitment: $R_k \subseteq [n]$
- Treatments: Z_k
- Observed outcomes: $Y = Y(Z)$
- Selective choice: S_k
- Short-hand: $W = (R, X_R, Y_R(\cdot))$

Pearl (2009)


$X, Y(\cdot)$
R_1
$Y_{R_1}(\cdot)$
X_{R_1}
Y_{R_1}
Z_1
Stage 1
S_1
R_2
$Y_{R_2}(\cdot)$
X_{R_2}
Y_{R_2}
Z_2
Stage 2
S_2

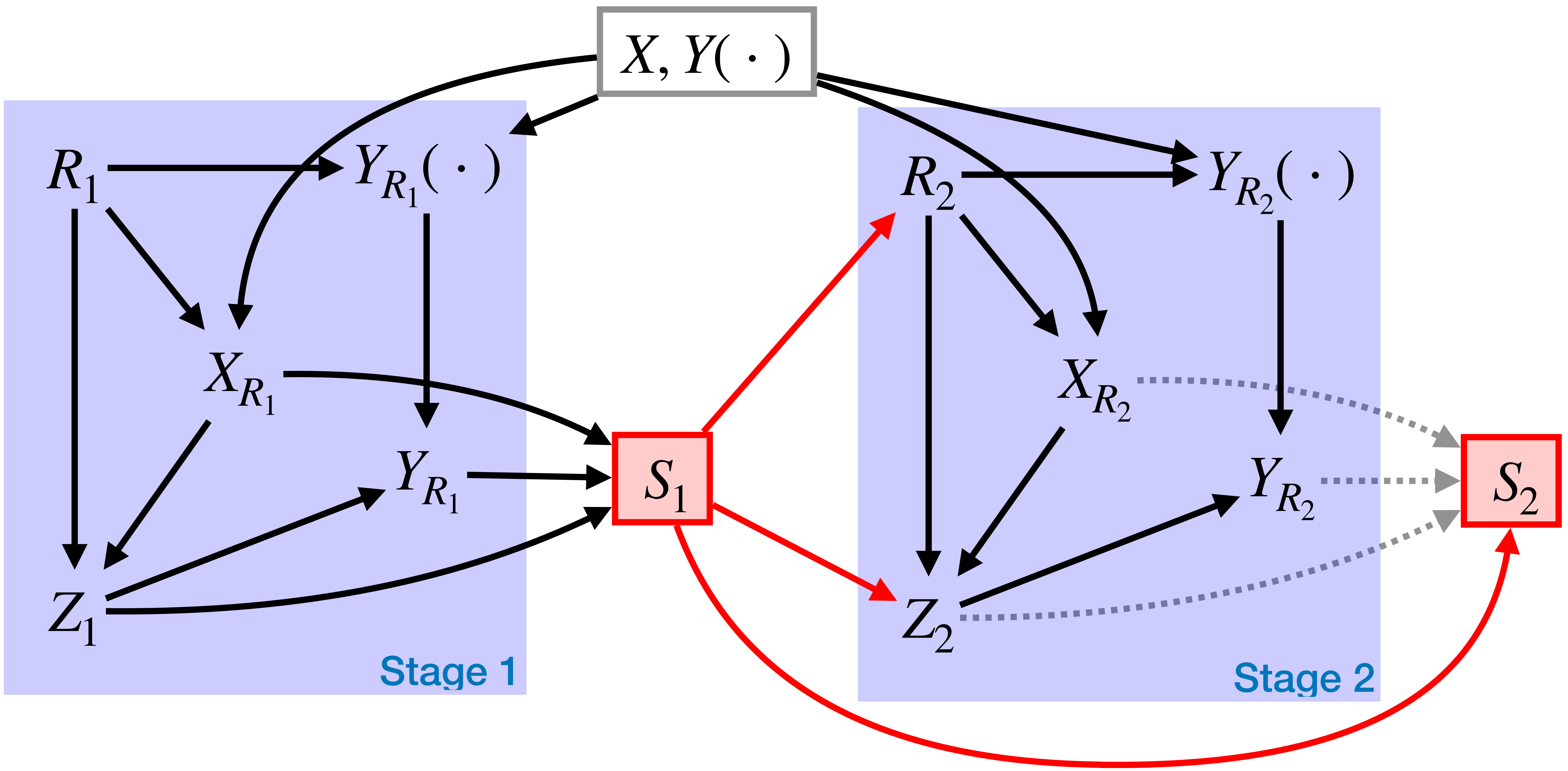


- **Assumption (A1):** Distribution of $Z = (Z_1, \ldots, Z_K)$ is known



- **Assumption (A1):** Distribution of $Z = (Z_1, \ldots, Z_K)$ is known
- **Assumption (A2):** Conditional independence of present and post potential outcomes

$$Z_k \perp\!\!\!\perp Y_{R_{[k]}}(\cdot) \mid R_{[k]}, X_{R_{[k]}}, Y_{R_{[k-1]}}, Z_{[k-1]} \quad \forall k \in [K]$$



- **Assumption (A1):** Distribution of $Z = (Z_1, \ldots, Z_K)$ is known
- **Assumption (A2):** Conditional independence of present and post potential outcomes

  $Z_k \perp\!\!\!\perp Y_{R_{[k]}}(\cdot) \mid R_{[k]}, X_{R_{[k]}}, Y_{R_{[k-1]}}, Z_{[k-1]} \quad \forall k \in [K]$

- **Assumption (A3):** S_k captures all the information that is passed on

  $R_k, X_{R_k}, Y_{R_k}(\cdot) \perp\!\!\!\perp Z_{[k-1]} \mid W_{[k-1]}, S_{k-1} \quad \forall k \in [K]$

Selective Randomization Inference

Selective Randomization Inference

- Standard Randomization Inference $\rightarrow$ Using data twice (double dipping)
- Comparing to $Z*$ that choose different stage-II design or null hypothesis than Z

Selective Randomization Inference

- Standard Randomization Inference $\rightarrow$ Using data twice (double dipping)
- Comparing to Z^* that choose different stage-II design or null hypothesis than Z
- Result: Type-I error inflation

Selective Randomization Inference

- Standard Randomization Inference $\rightarrow$ Using data twice (double dipping)
- Comparing to Z^* that choose different stage-II design or null hypothesis than Z
- Result: Type-I error inflation
- Solutions:
  - Data splitting (Cox, 1975): $\mathbb{P}(\, T(Z^*, W) \leq T(Z, W) \mid W, Z, Z_1^* = Z_1)$, where $K = 2$

Selective Randomization Inference

- Standard Randomization Inference $\rightarrow$ Using data twice (double dipping)
- Comparing to Z^* that choose different stage-II design or null hypothesis than Z
- Result: Type-I error inflation
- Solutions:
  - Data splitting (Cox, 1975): $\mathbb{P}(\,T(Z^*, W) \leq T(Z, W) \mid W, Z, Z_1^* = Z_1)$, where $K = 2$
  - Selective inference (Lee et al., 2016; Fithian et al., 2017): regression models etc.

Selective Randomization Inference

- Standard Randomization Inference $\rightarrow$ Using data twice (double dipping)
- Comparing to Z^* that choose different stage-II design or null hypothesis than Z
- Result: Type-I error inflation
- Solutions:
  - Data splitting (Cox, 1975): $\mathbb{P}(\, T(Z^*, W) \leq T(Z, W) \mid W, Z, Z_1^* = Z_1)$, where $K = 2$
  - Selective inference (Lee et al., 2016; Fithian et al., 2017): regression models etc.
  - Selective randomization inference:

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

Type-I Error

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

Type-I Error

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

- **Proposition:** P_{sel} controls the selective type-I error, i.e.

$$\mathbb{P}(P_{sel} \leq \alpha \mid W, S(Z) = s) \leq \alpha \qquad \forall\, s$$

Type-I Error

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

- **Proposition:** P_{sel} controls the selective type-I error, i.e.

$$\mathbb{P}(P_{sel} \leq \alpha \mid W, S(Z) = s) \leq \alpha \qquad \forall\, s$$

- Formula for p-value: $P_{sel} = \dfrac{\sum_{z^*} \mathbf{1}\{T(z^*, W) \leq T(Z, W)\}\ \mathbf{1}\{S(z^*) = S(Z)\}\ q\left(z^* \mid W\right)}{\sum_{z^*} \mathbf{1}\{S(z^*) = S(Z)\}\ q\left(z^* \mid W\right)}$

- $q(z \mid w) = \prod_{k=1}^{K} \mathbb{P}(Z_k = z_k \mid R_k = r_k, X_{R_k} = x_{R_k}, S_{k-1} = s_{k-1})$ under slightly stronger assumptions

Computation

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

Computation

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

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

$$\hat{P}_M := \frac{1 + \sum_{j=1}^M \mathbf{1}\{T(z_j^*, W) \leq T(Z, W)\}}{1 + M}.$$

Computation

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

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

$$\hat{P}_M := \frac{1 + \sum_{j=1}^M \mathbf{1}\{T(z_j^*, W) \leq T(Z, W)\}}{1 + M}.$$

- Rejection sampling, Markov Chain Monte Carlo (MCMC)

Inference

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

Inference

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

- Confidence intervals:

Inference

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

- Confidence intervals:
  - test $Y_i(1) - Y_i(0) = \tau$ for different τ
  - $(1 - \alpha)$ confidence interval: $C_{1-\alpha} = \{\tau : P_{sel}(\tau) \geq \alpha\}$
- Estimation: τ such that $P_{sel}(\tau) = 0.5$

Simulation Study

Simulation Study

- 2 stages, 2 treatments $Z_i \in \{0,1\}$, 2 groups $X_i \in \{\text{low}, \text{high}\}$
- Potential outcomes: $Y_i(0) = Y_i(1) \sim N(0,1)$ i.i.d.
- First stage: 100 patients, Second stage: 40 patients

Simulation Study

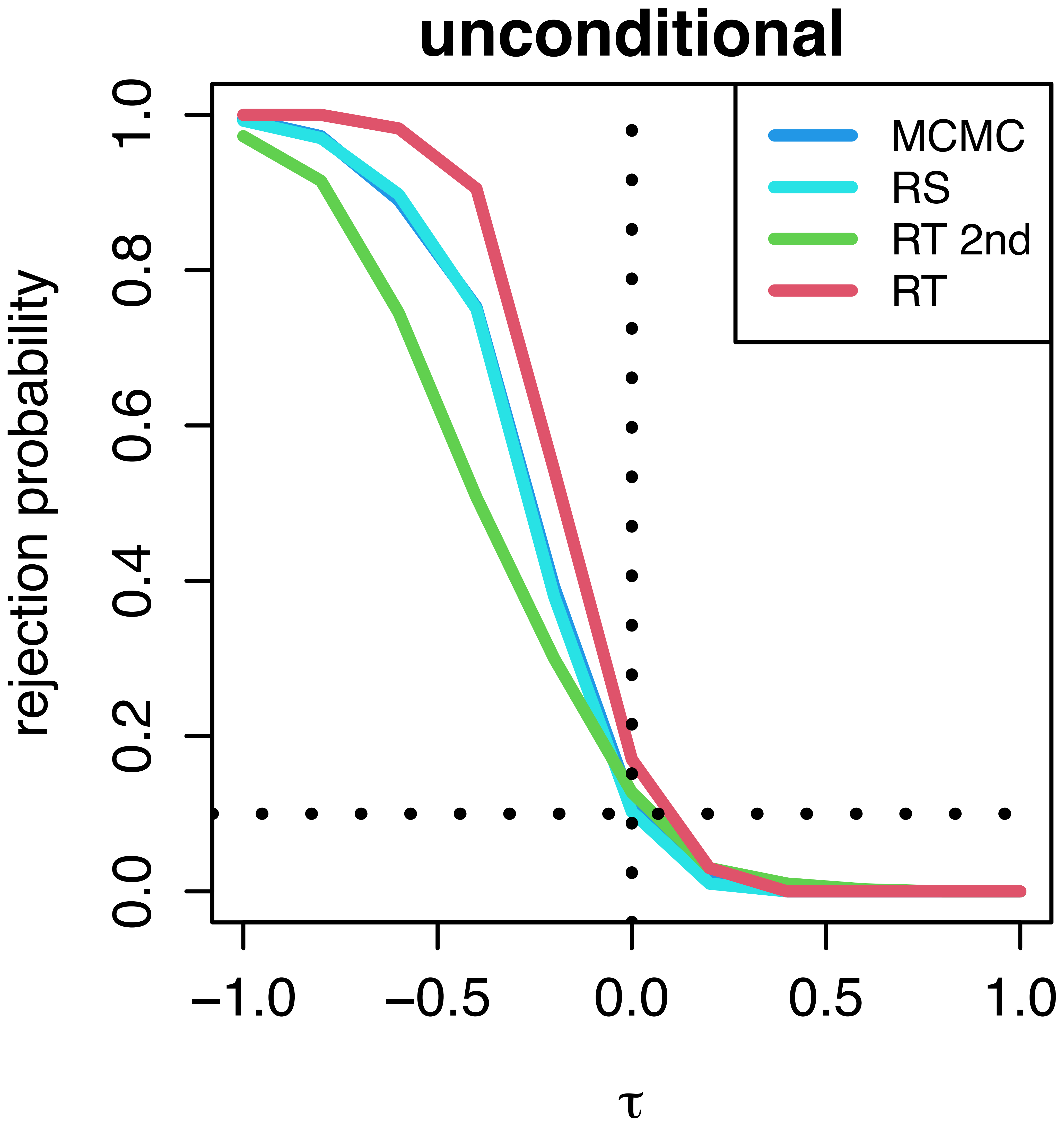
- 2 stages, 2 treatments $Z_i \in \{0,1\}$, 2 groups $X_i \in \{\text{low, high}\}$
- Potential outcomes: $Y_i(0) = Y_i(1) \sim N(0,1)$ i.i.d.
- First stage: 100 patients, Second stage: 40 patients
- $\Delta =$ standardized difference in SATEs between groups

Simulation Study

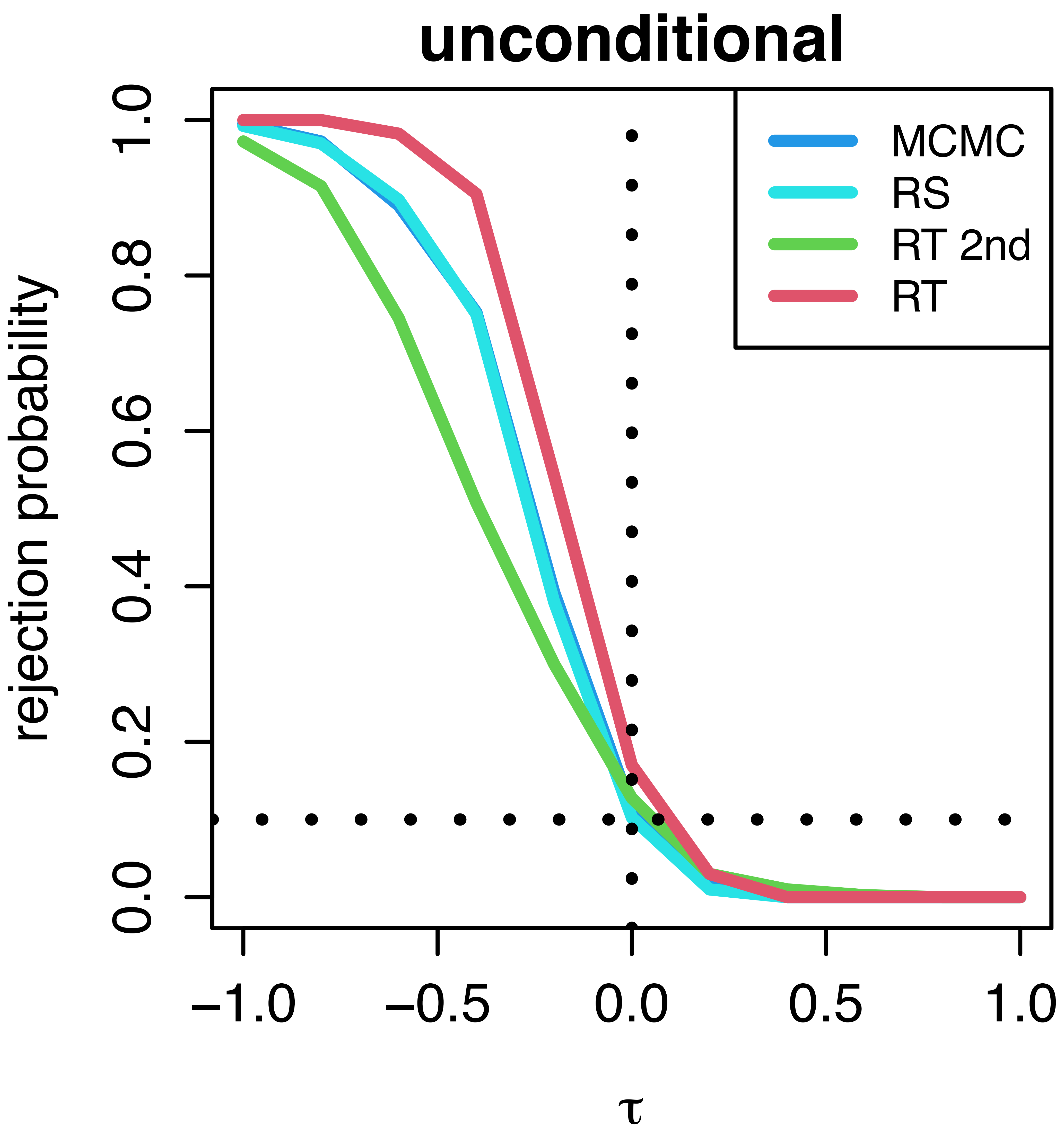
- 2 stages, 2 treatments $Z_i \in \{0,1\}$, 2 groups $X_i \in \{\text{low, high}\}$
- Potential outcomes: $Y_i(0) = Y_i(1) \sim N(0,1)$ i.i.d.
- First stage: 100 patients, Second stage: 40 patients
- $\Delta =$ standardized difference in SATEs between groups
- Selection variable:

$$S = \begin{cases} \text{only low}, & \Delta < \Phi^{-1}(0.2), & \text{recruit 40 from group } X_i = \text{low} \\ \text{only high}, & \Delta > \Phi^{-1}(0.8), & \text{recruit 40 from group } X_i = \text{high} \\ \text{both}, & \text{otherwise}, & \text{recruit 20 from each group} \end{cases}$$

Power Analysis

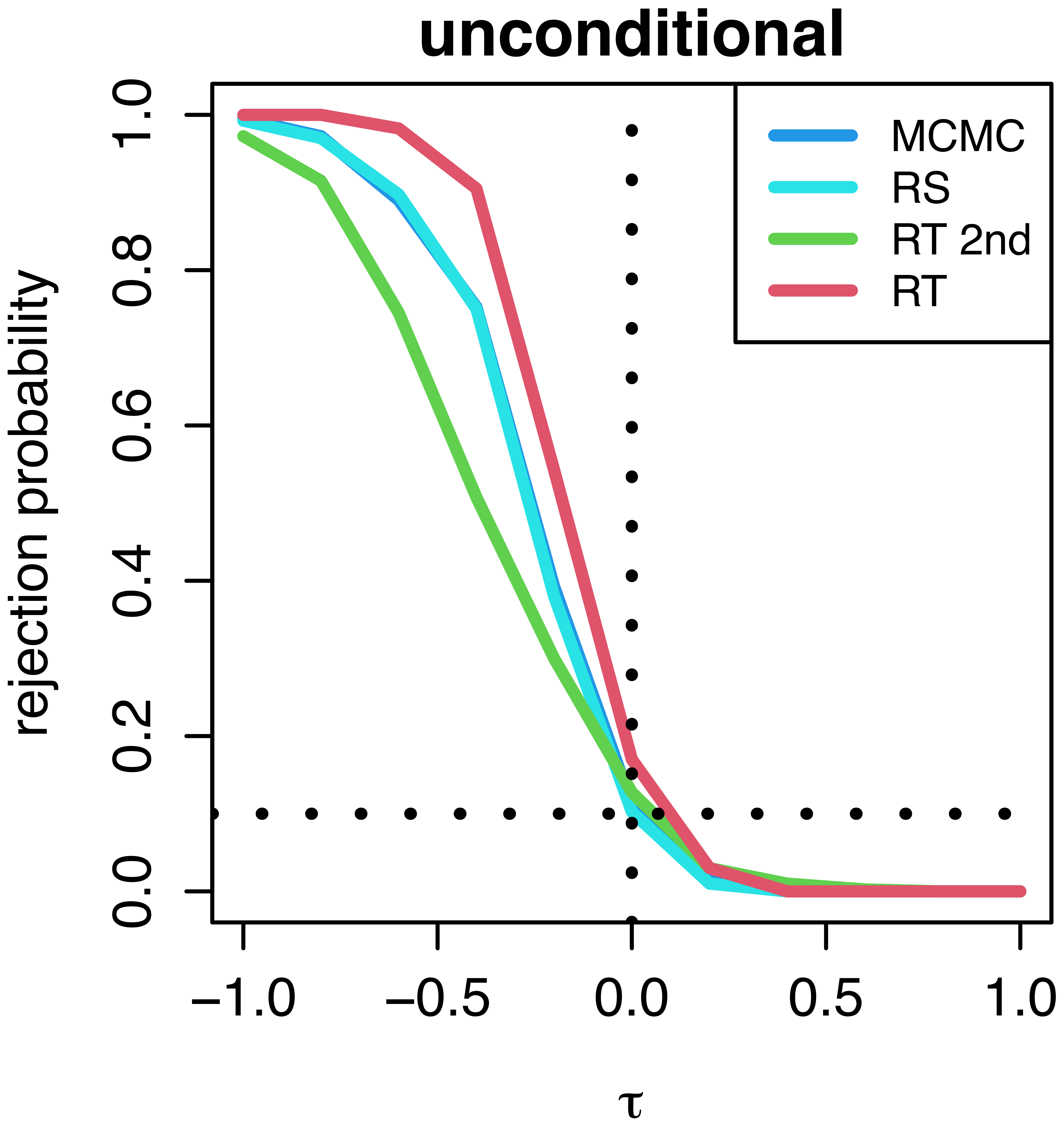


Power Analysis



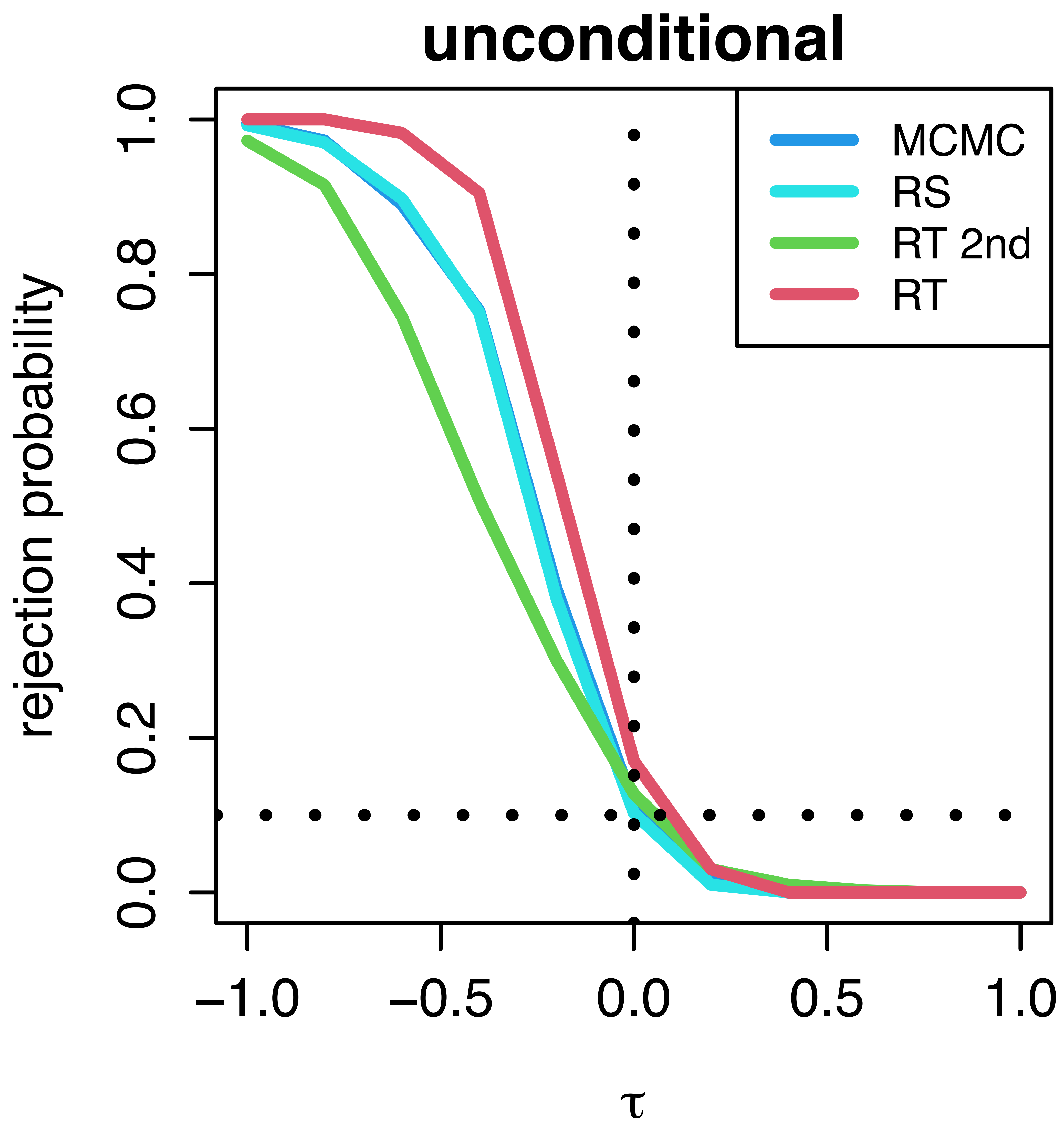
- RT: **no type-I error control**
- RT 2nd: valid but has **low power**

Power Analysis



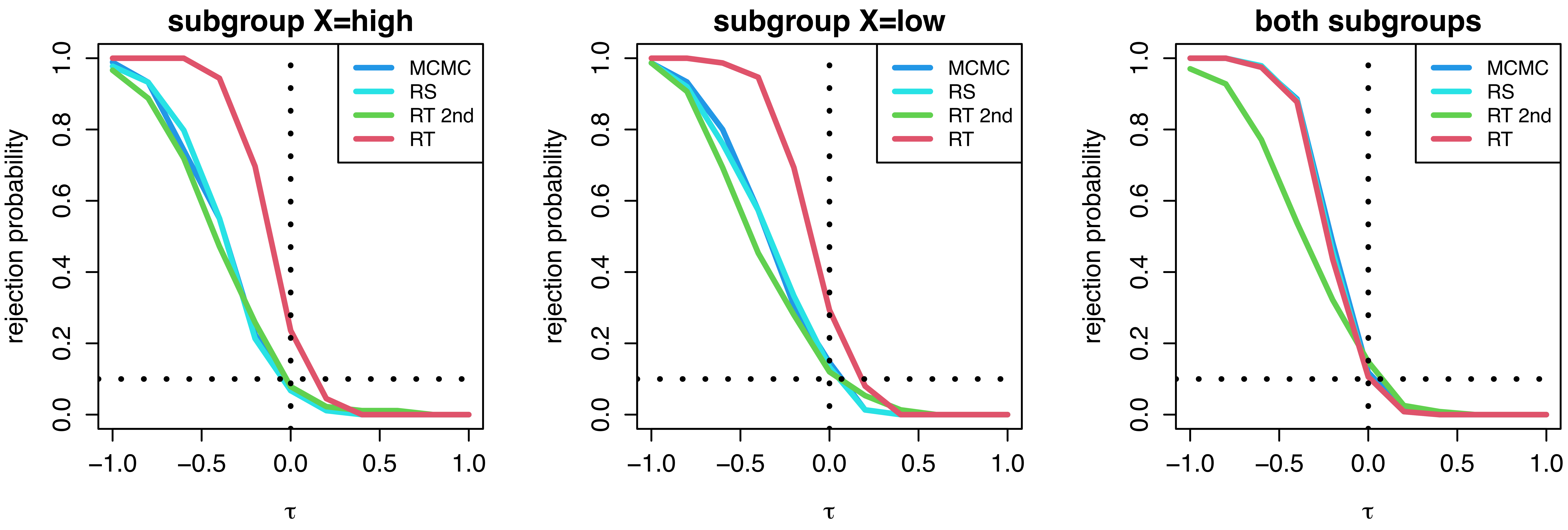
- RT: **no type-I error control**
- RT 2nd: valid but has **low power**
- Selective RT: **valid and more powerful.**

Power Analysis

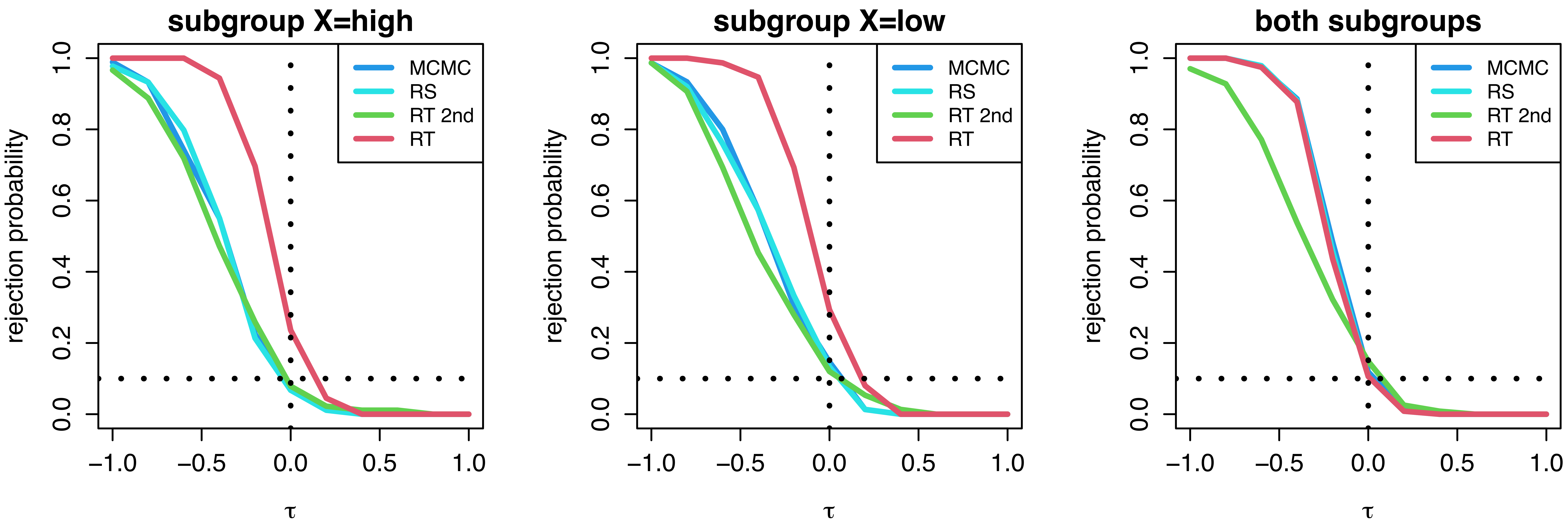


- RT: **no type-I error control**
- RT 2nd: valid but has **low power**
- Selective RT: **valid and more powerful.**
- Rejection sampling and MCMC lead to very similar approximations.

Power Analysis

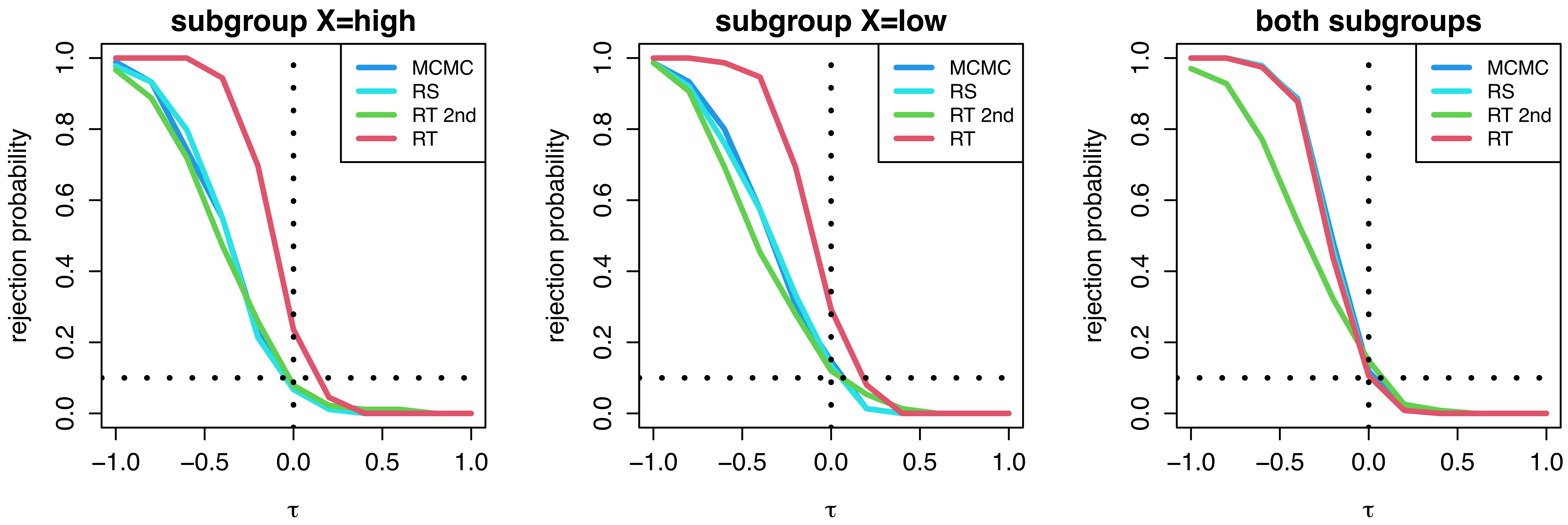


Power Analysis



- 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”

Conclusion

- Robust inference in randomized controlled trials
- Experiments with adaptive treatments, recruitment and null hypothesis
- Visualization via graphs
- **Key idea: Conditioning randomization p-value on the selection information**
- Approximation via rejection sampling or MCMC

Thanks for your attention!

tobias.freidling@epfl.ch

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

Imbens, G.W. and Rubin, D.B. (2015) Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge: Cambridge University Press.

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).

Marston, N.A. et al. (2020) ‘Predicting Benefit From Evolocumab Therapy in Patients With Atherosclerotic Disease Using a Genetic Risk Score’, Circulation, 141(8), pp. 616–623.

Pearl, J. (2009) Causality: Models, Reasoning, and Inference. 2nd edn. Cambridge University Press.

Pitman, E.J.G. (1937) ‘Significance Tests Which May be Applied to Samples From any Populations’, Supplement to the Journal of the Royal Statistical Society, 4(1), pp. 119–130.

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

Hold-out Units