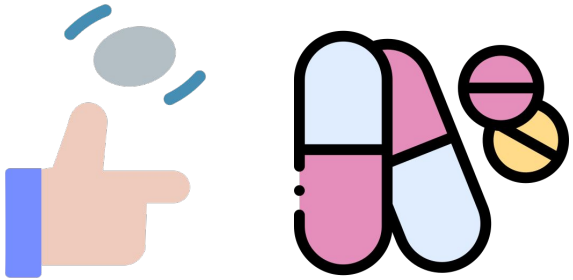
Falsification of Internal and External Validity in Observational Studies via Conditional Moment Restrictions

Zeshan Hussain*, Ming-Chieh Shih*, Michael Oberst, Ilker Demirel, David Sontag

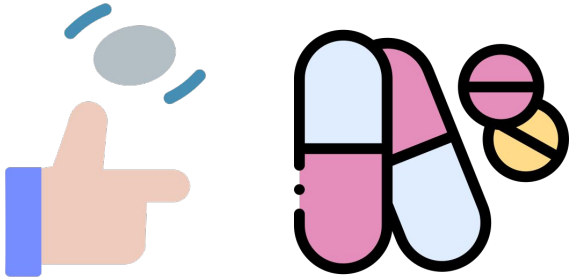
***equal contribution, by alphabetical order**

Motivation: Making use of experimental data



Randomized Controlled Trial (RCT)

Motivation: Making use of experimental data



Randomized Controlled Trial (RCT)

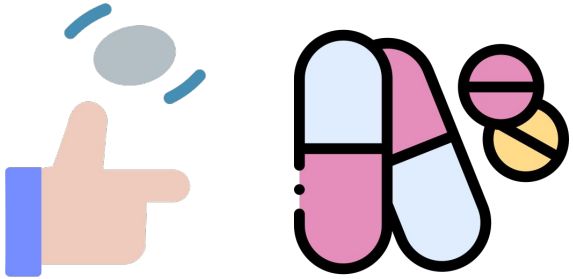
Randomization ⇒ Unbiased Estimates

Strict Inclusion Criteria

Small sample sizes

Motivation: Making use of experimental data

Internal validity: estimated causal effects are unbiased within the study population



Randomized Controlled Trial (RCT)

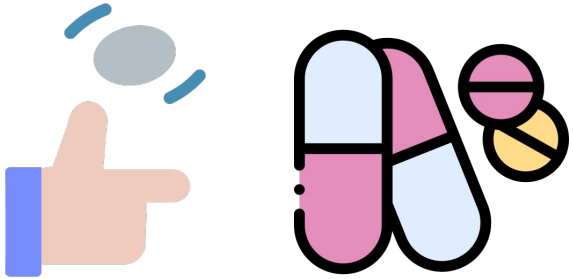
Randomization ⇒ Unbiased Estimates

Strict Inclusion Criteria

Small sample sizes

Motivation: Making use of experimental data

Internal validity: estimated causal effects are unbiased within the study population



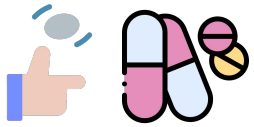
Randomized Controlled Trial (RCT)

Randomization ⇒ Unbiased Estimates

Strict Inclusion Criteria

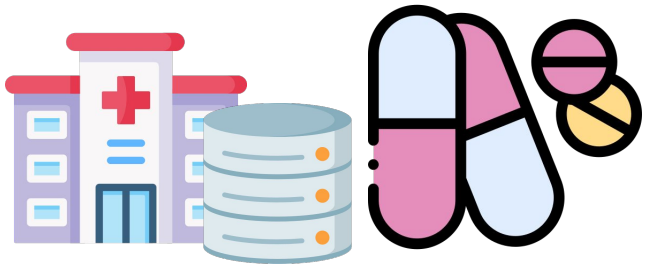
External validity: ability to generalize estimates across wider populations

Motivation: Making use of experimental data



RCTs

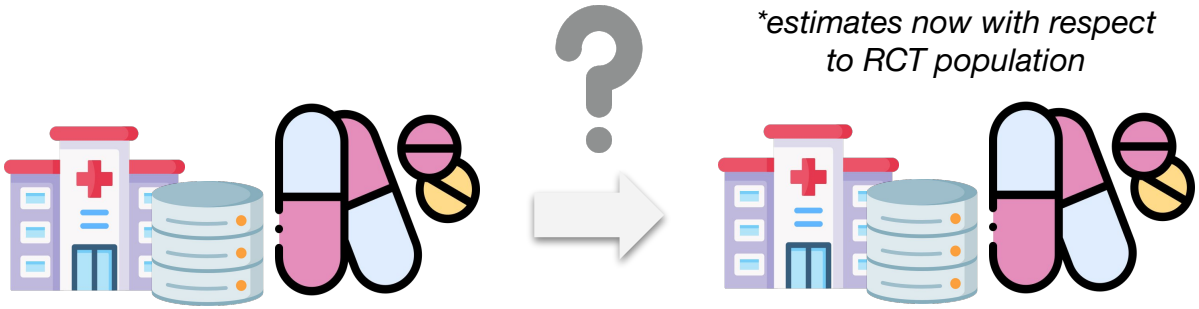
- Unbiased Estimates
- Strict Inclusion Criteria
- Small sample sizes



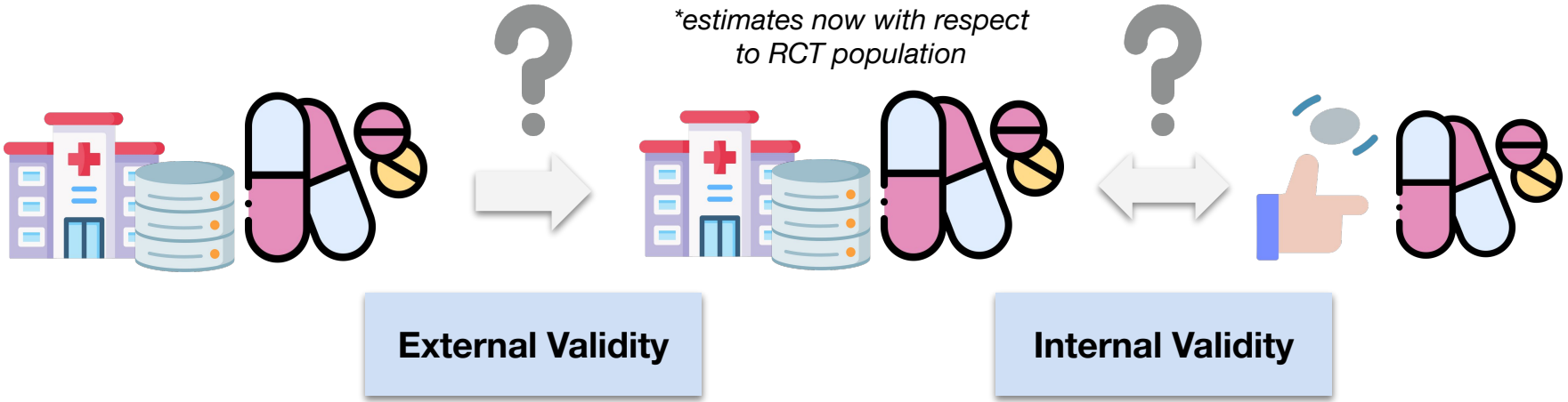
Observational Studies

- Broader, more diverse populations
- Larger sample sizes
- Potential (unobserved) confounding

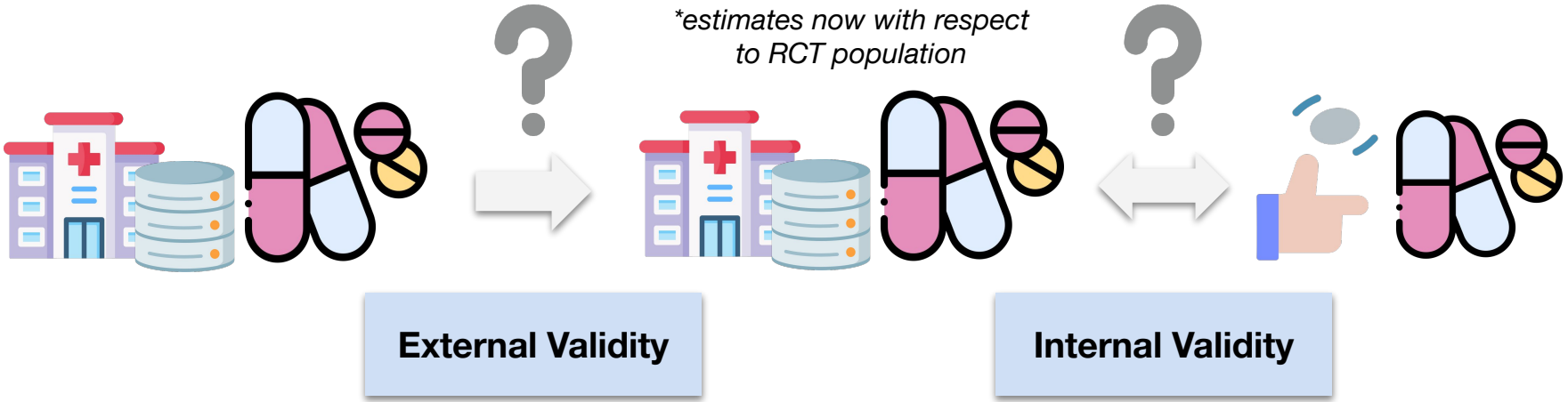
Goal: Falsification of observational estimates



Goal: Falsification of observational estimates

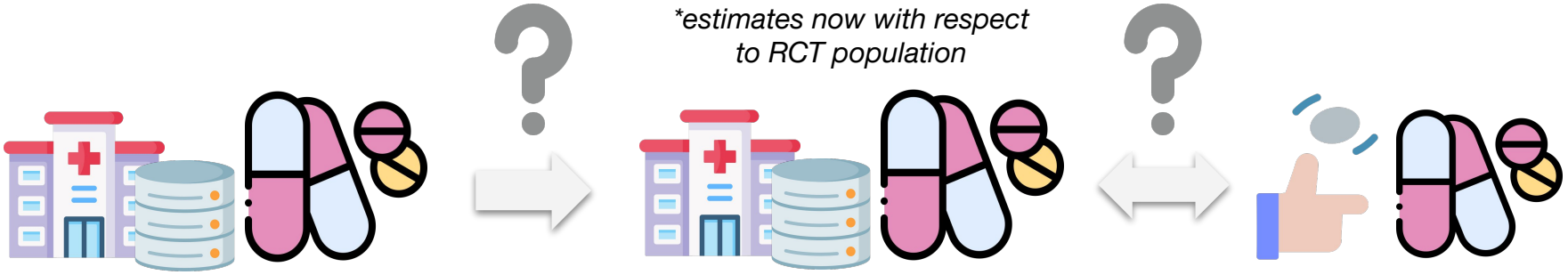


Goal: Falsification of observational estimates



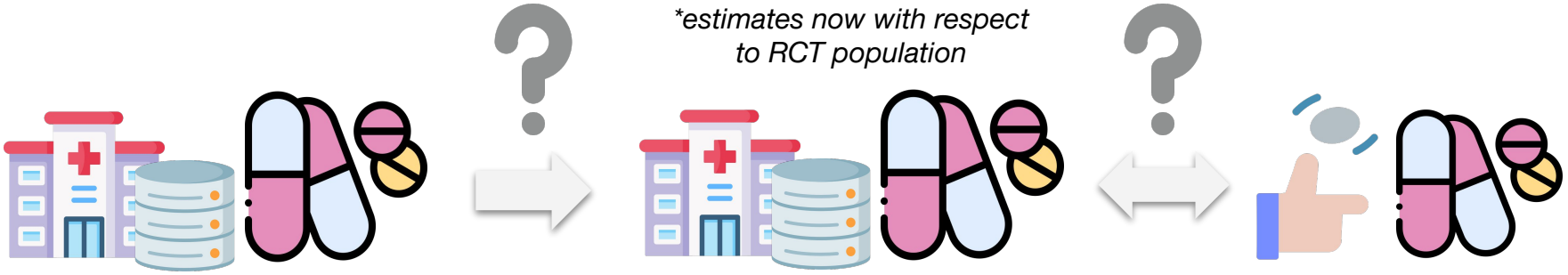
Goal: Use experimental data as a form of validation — *"falsify"* assumptions of external and internal validity in observational studies

Desired properties of a falsification algorithm



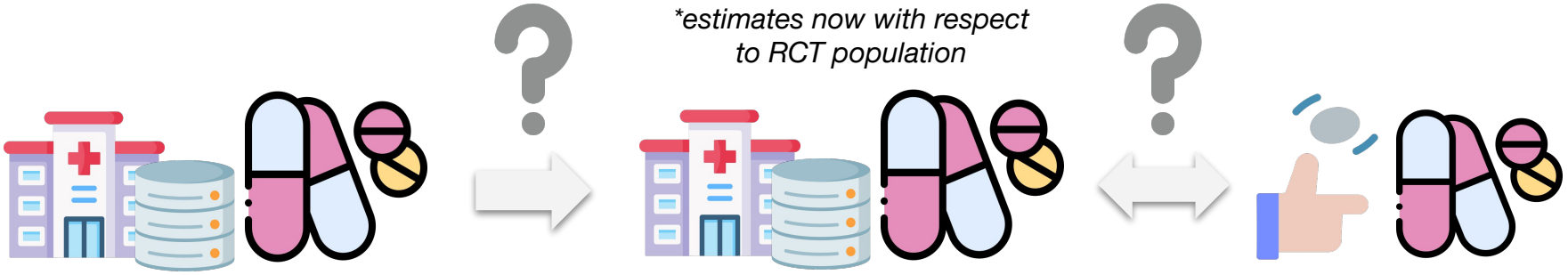
Property 1: High power

Desired properties of a falsification algorithm



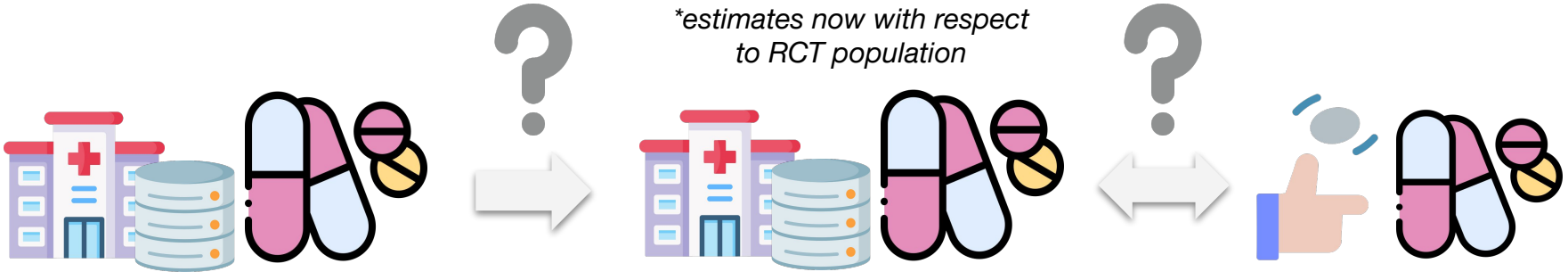
Property 1: High power

Desired properties of a falsification algorithm



Property 1: High power

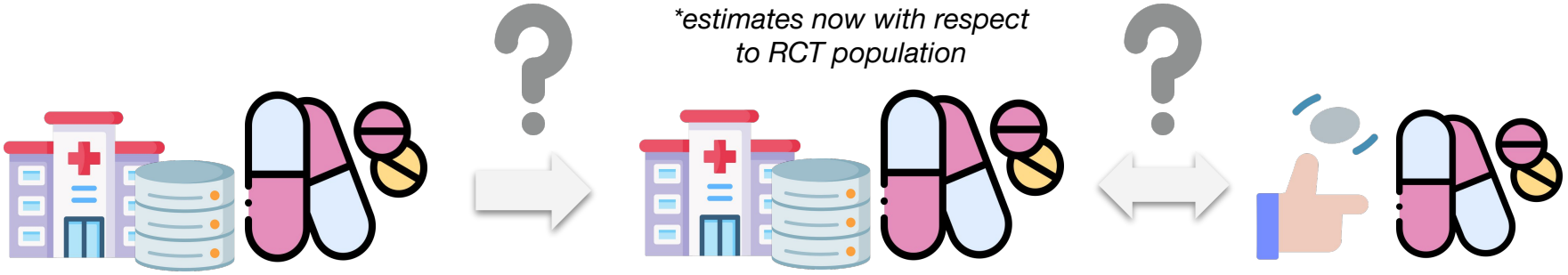
Desired properties of a falsification algorithm



Property 1: High power

Property 2: Controlled Type I Error

Desired properties of a falsification algorithm



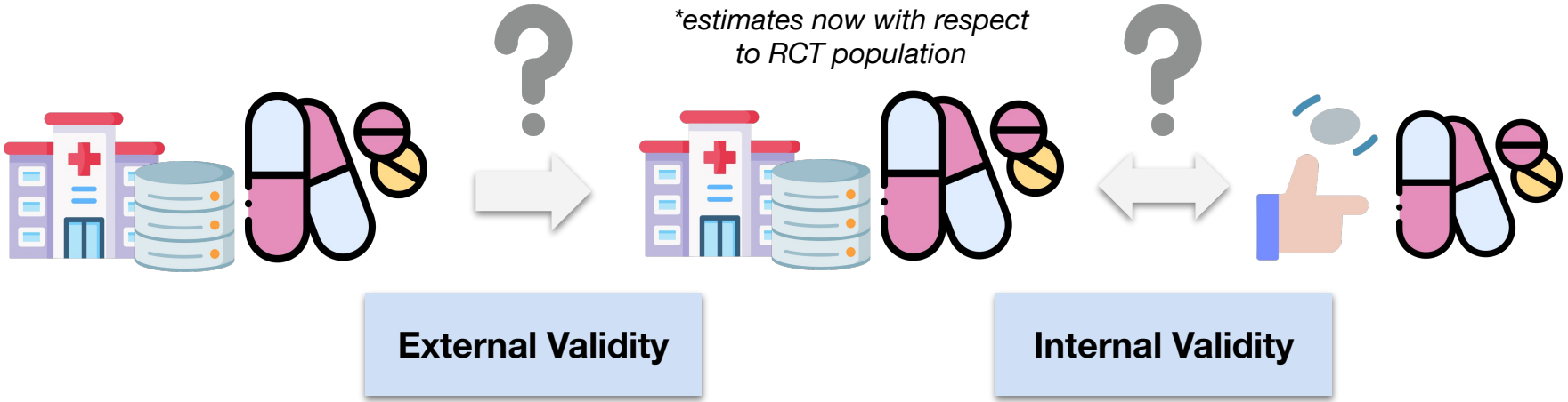
E.g., estimates for **older, male patients** tend to be **positively biased** compared to RCT estimates

Property 1: High power

Property 2: Controlled Type I Error

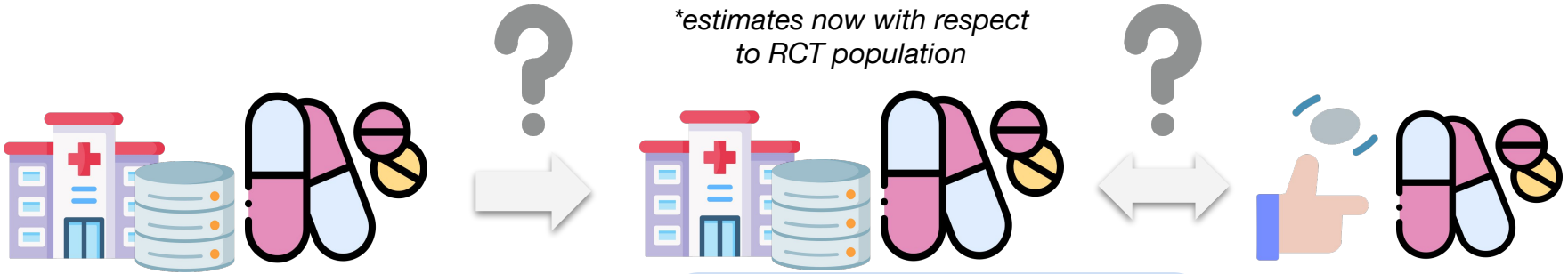
Property 3: Explanation of falsification

Goal: Falsification of observational estimates



Our Contribution: show how assumptions of external and internal validity can be converted into a set of **conditional moment restrictions (CMRs)**, which can be tested using existing techniques with theoretical guarantees

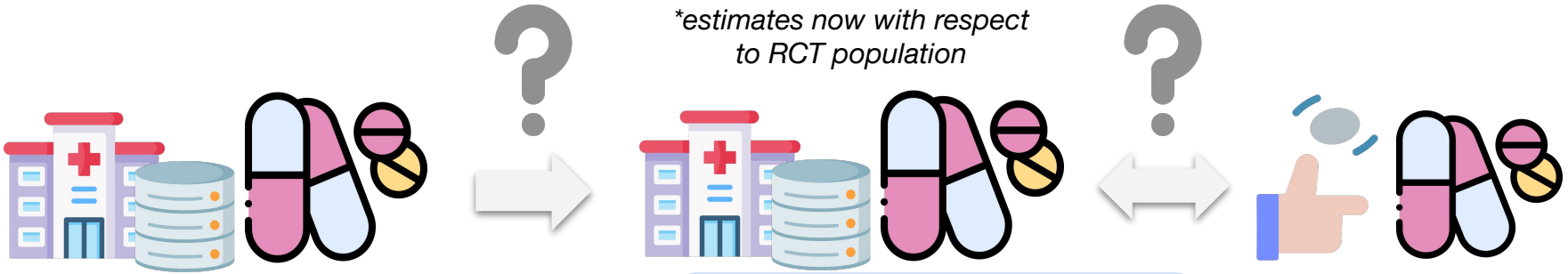
Goal: Falsification of observational estimates



CMR – conditional mean of certain functions of data is almost surely equal to zero (for correctly specified models)

Our Contribution: show how assumptions o xternal and internal validity can be converted into a set of **conditional moment restrictions (CMRs)**, which can be tested using existing techniques with theoretical guarantees

Goal: Falsification of observational estimates



Set of CMRs will encapsulate the **<u>difference in CATE estimates</u>** between the RCT and observational study.

Our Contribution: show how assumptions (xternal and internal validity can be converted into a set of **conditional moment restrictions (CMRs)**, which can be tested using existing techniques with theoretical guarantees

Context: Prior approaches to falsification

$$\text{ATE}_{OBS} \quad \text{ATE}_{RCT}$$

[1] Franklin, Jessica M., et al. "Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative." Circulation 143.10 (2021): 1002-1013.
[2] Dagan, Noa, et al. "BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting." New England Journal of Medicine (2021).

Context: Prior approaches to falsification

ATE_{OBS} ATE_{RCT}

Could lead to false negatives

[1] Franklin, Jessica M., et al. "Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative." Circulation 143.10 (2021): 1002-1013.
[2] Dagan, Noa, et al. "BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting." New England Journal of Medicine (2021).

Context: Prior approaches to falsification

$$\text{ATE}_{OBS} \quad \text{ATE}_{RCT}$$

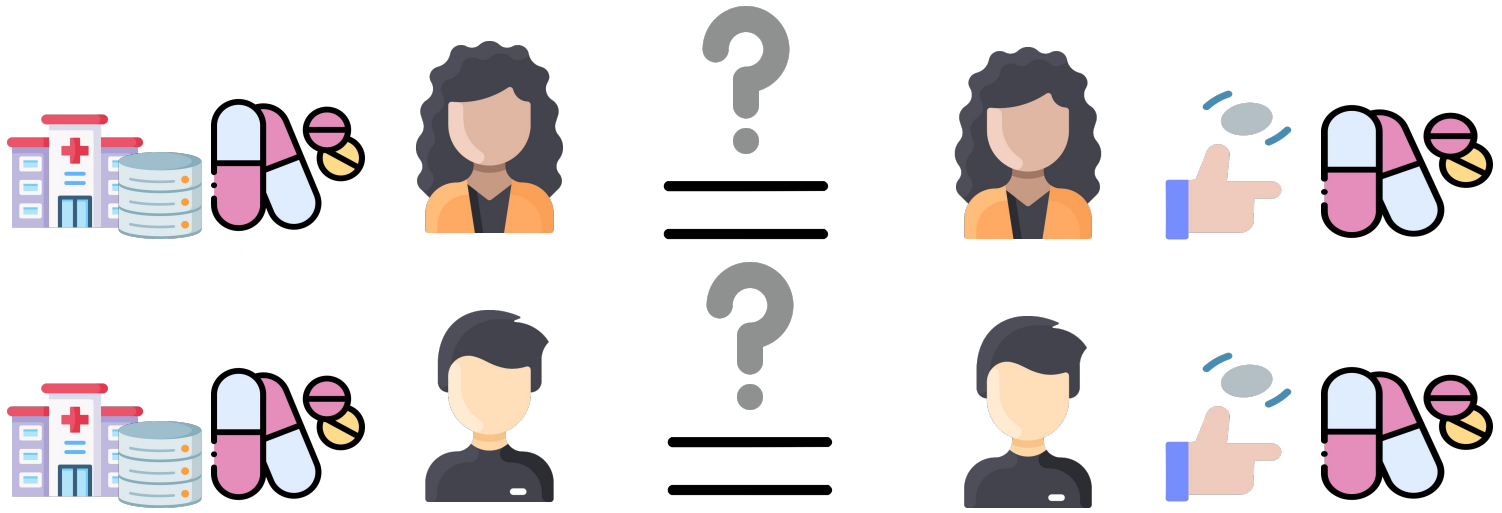
No explanation for rejection

[1] Franklin, Jessica M., et al. "Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative." Circulation 143.10 (2021): 1002-1013.
[2] Dagan, Noa, et al. "BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting." New England Journal of Medicine (2021).

Context: Prior approaches to falsification

[3] Hussain, Z. M.*, Oberst, M.*, Shih, M. C.*, & Sontag, D. (2022). Falsification before Extrapolation in Causal Effect Estimation. Advances in Neural Information Processing Systems, 35, 6161-6174., ***equal contribution**

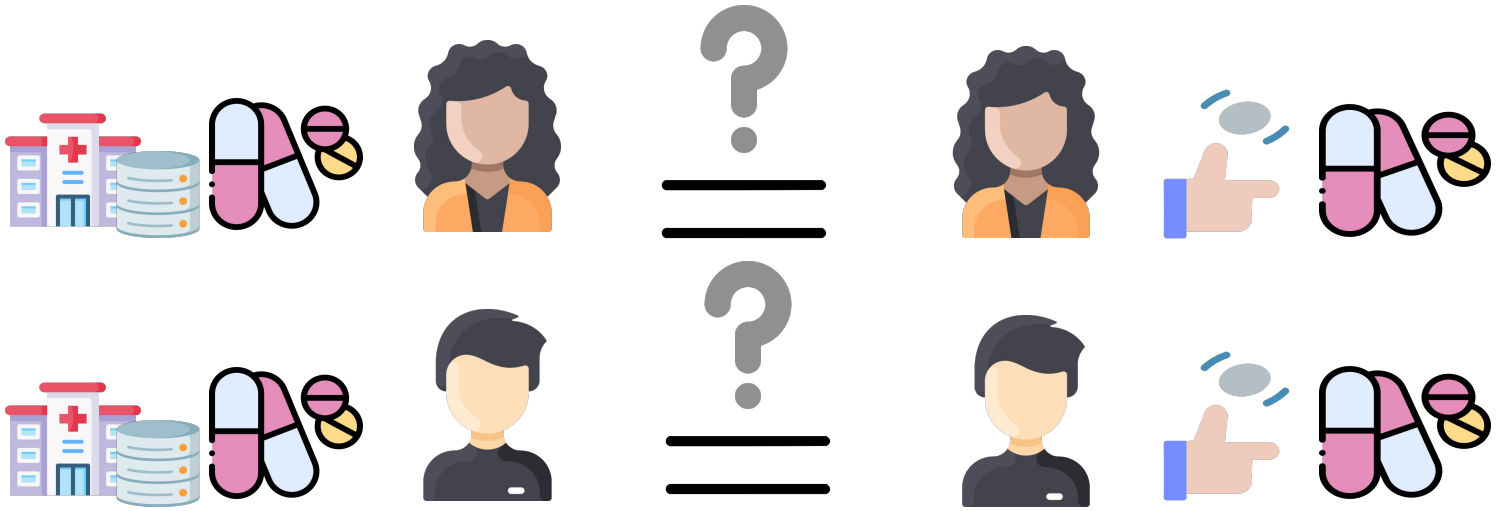
Context: Prior approaches to falsification



Need to correct for multiple hypothesis testing

[3] Hussain, Z. M.*, Oberst, M.*, Shih, M. C.*, & Sontag, D. (2022). Falsification before Extrapolation in Causal Effect Estimation. Advances in Neural Information Processing Systems, 35, 6161-6174., ***equal contribution**

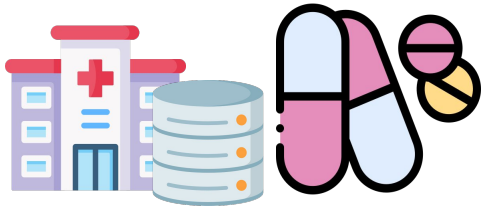
Context: Prior approaches to falsification



Need a-priori specification of the subgroups; would be nice to find these automatically

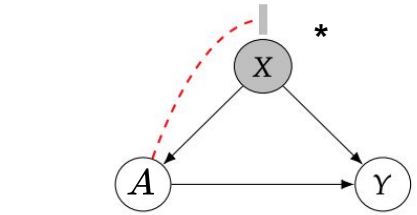
[3] Hussain, Z. M.*, Oberst, M.*, Shih, M. C.*, & Sontag, D. (2022). Falsification before Extrapolation in Causal Effect Estimation. Advances in Neural Information Processing Systems, 35, 6161-6174., ***equal contribution**

Approach: Formalization of Assumptions



Internal Validity

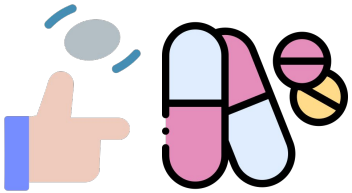
- Ignorability
- Consistency
- Positivity



$$Y^{} = Y$$

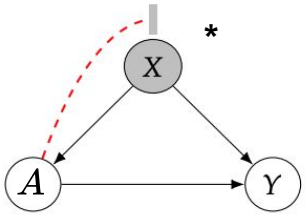
*taken from Brady Neal's: Introduction to Causal Inference

Approach: Formalization of Assumptions



Internal Validity

- **Ignorability**
- **Consistency**
- **Fixed Probability of Assignment**

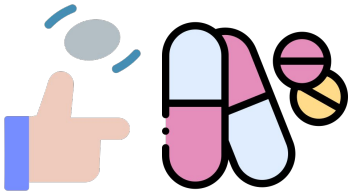


$$Y^{*} = Y$$

$$P(\text{pill}) = 0.5$$

*taken from Brady Neal's: Introduction to Causal Inference

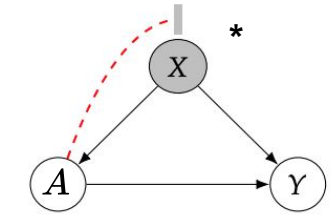
Approach: Formalization of Assumptions



Internal Validity

Ignorability

Consistency



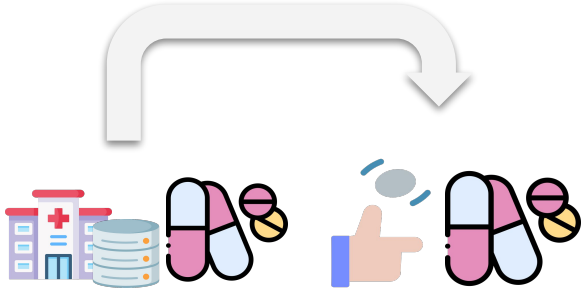
$$Y = Y$$

$$P(\text{pill}) = 0.5$$

But we still need to be able to **compare causal effects** between the RCT and observational study

*taken from Brady Neal's: Introduction to Causal Inference

Approach: Formalization of Assumptions



External Validity

Mean Exchangeability of Contrast

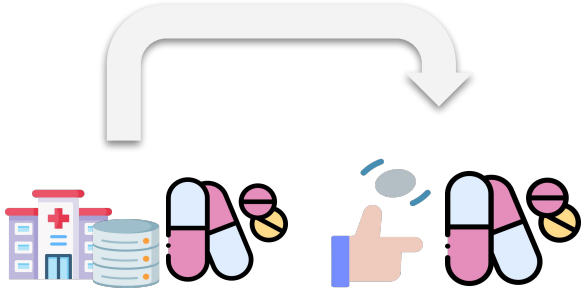
$$\text{CATE} := \mathbb{E}[Y_1 - Y_0 | X = x]$$

Outcome under treatment

Outcome under control

Assignment of covariates

Approach: Formalization of Assumptions



External Validity

Mean Exchangeability of Contrast

$$\text{CATE} := \mathbb{E}[Y_1 - Y_0 | X = x]$$

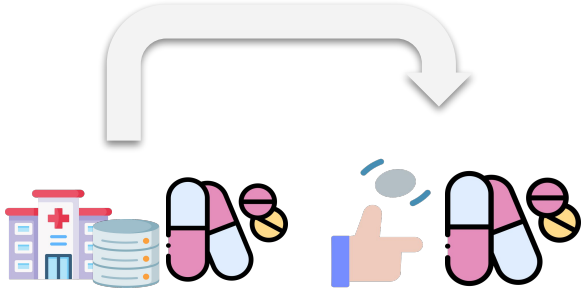
Outcome under treatment

Outcome under control

Assignment of covariates

$$\text{CATE} = \text{CATE}$$

Approach: Formalization of Assumptions



External Validity

Mean Exchangeability of Contrast

$$\text{CATE} := \mathbb{E}[Y_1 - Y_0 | X = x]$$

Outcome under treatment

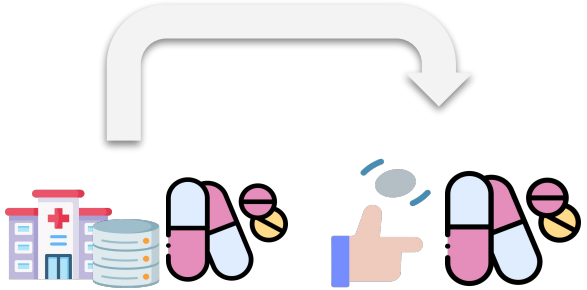
Outcome under control

Assignment of covariates

$$\mathbb{E}[Y_1 - Y_0 | X = x] = \mathbb{E}[Y_1 - Y_0 | X = x, S = s]$$

$$\forall s \in \{0, 1\}$$

Approach: Formalization of Assumptions



External Validity

Mean Exchangeability of Contrast

$$\text{CATE} := \mathbb{E}[Y_1 - Y_0 | X = x]$$

Outcome under treatment

Outcome under control

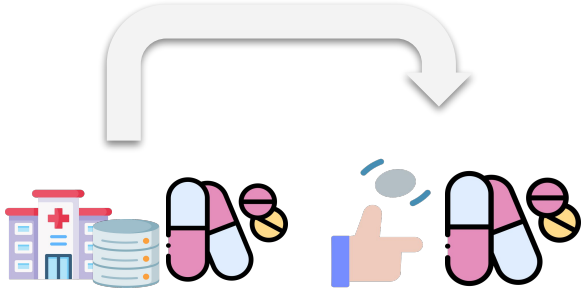
Assignment of covariates

$$\mathbb{E}[Y_1 - Y_0 | X = x] = \mathbb{E}[Y_1 - Y_0 | X = x, S = s]$$

Indicator variable for RCT or OBS

$$\forall s \in \{0, 1\}$$

Approach: Formalization of Assumptions



External Validity

Positivity of Selection

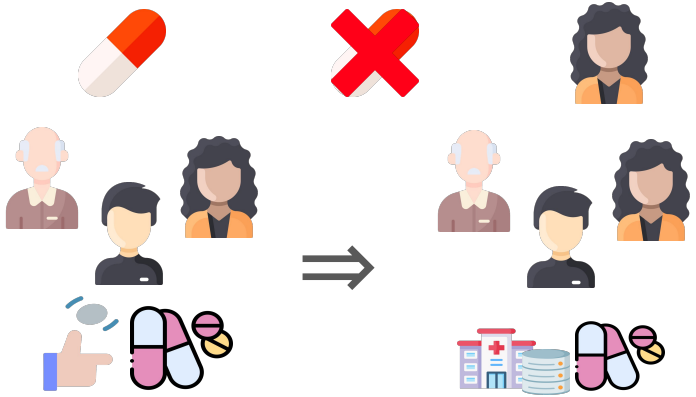
Mean Exchangeability of Contrast

$$\text{CATE} := \mathbb{E}[Y_1 - Y_0 | X = x]$$

Outcome under treatment

Outcome under control

Assignment of covariates



Approach: Formalization of Assumptions

External V

Mean Exchangeability of Contrast

$$\text{CATE} \quad - Y_0 \mid X = x]$$

Aside: the assumption on causal contrasts is weaker than the standard assumption of exchangeability/ transportability of **counterfactual means**.

er

Assignment of covariates

Approach: Formalization of Assumptions

External V

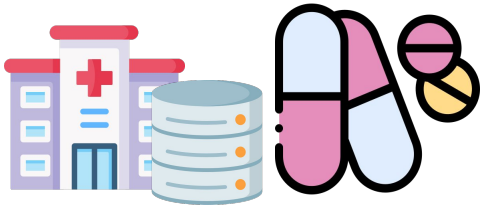
Mean Exchangeability of Contrast

$$\text{CATE} := \mathbb{E}[Y_1 - Y_0 \mid X = x]$$

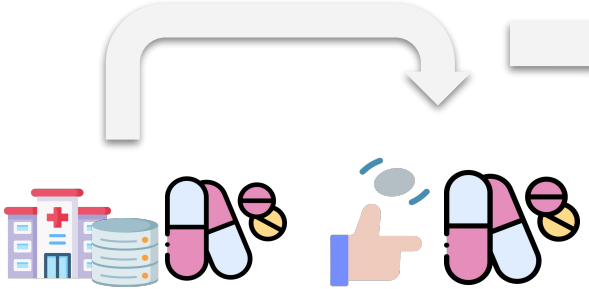
See paper for empirical and theoretical motivations for this assumption, and examples for when **counterfactual means in the RCT are not identifiable** from observational data, but the **causal contrast is identified**.

ssignment of covariates

Approach: Framing as a set of CMRs



Internal Validity

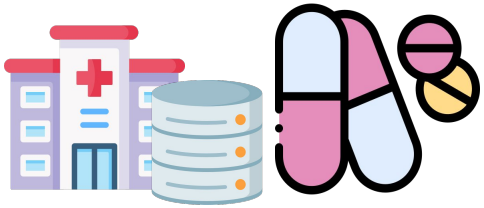


External Validity

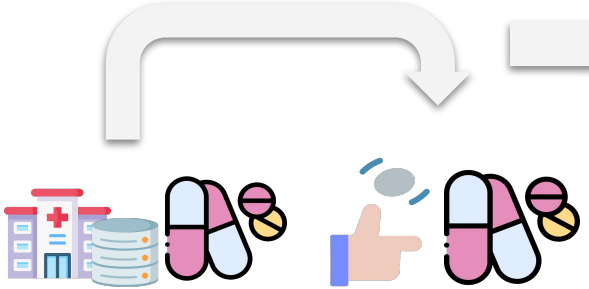
$\psi_1 \quad \psi_0$

Unbiased instance-wise CATE signals from observational study and RCT

Approach: Framing as a set of CMRs



Internal Validity



External Validity

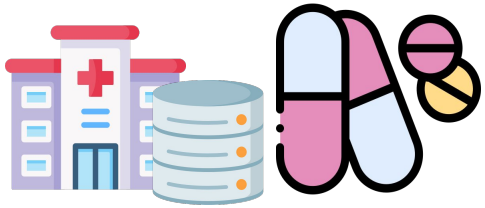
$\psi_1 \quad \psi_0$

Unbiased instance-wise CATE signals from observational study and RCT

$$H_0 : \mathbb{E}[\psi_1 - \psi_0 | X] = 0 \quad P_X - a.s.$$

CMRs

Approach: Framing as a set of CMRs



Internal Validity

External Validity

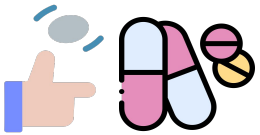
$\psi_1 \quad \psi_0$

Unbiased instance-wise CATE signals from observational study and RCT

$$H_0 : \mathbb{E}[\psi_1 - \psi_0 | X] = 0 \quad P_X - a.s.$$

CMRs

Approach: Getting unbiased CATE signals



IPW-style estimator

ψ_0

Internal Validity of RCT

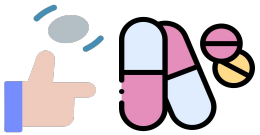
CATE signal from RCT

Unbiased estimator (for CATE in RCT population)

$$\mathbb{E}[\psi_0(Y, S, A, X)|X] = \mathbb{E}[Y_1 - Y_0|X, S = 0]$$

A denotes treatment assignment (**0** or **1**)
S denotes study index (**0** for RCT, **1** for OBS.)
X denotes patient covariates
Y denotes outcome
$\mathbf{Y_a}$ denotes potential outcome under treatment **A = a**

Approach: Getting unbiased CATE signals



ψ_0

Internal Validity of RCT

CATE signal from RCT

Unbiased estimator (for CATE in RCT population)

Doubly robust transported estimator

$$\mathbb{E}[\psi_0(Y, S, A, X)|X] = \mathbb{E}[Y_1 - Y_0|X, S = 0]$$

ψ_1

Internal and External Validity of OBS

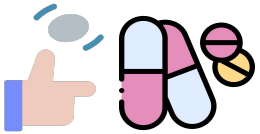
CATE signal from OBS

Unbiased estimator (for CATE in RCT population)

$$\mathbb{E}[\psi_1(Y, S, A, X)|X] = \mathbb{E}[Y_1 - Y_0|X, S = 0]$$

A denotes treatment assignment (**0** or **1**)
S denotes study index (**0** for RCT, **1** for OBS.)
X denotes patient covariates
Y denotes outcome
$\mathbf{Y_a}$ denotes potential outcome under treatment **A = a**

Approach: Getting unbiased CATE signals



CATE signal from RCT

ψ_0

Internal Validity of RCT

Unbiased estimator (for CATE in RCT population)

CATE signal from OBS

ψ_1

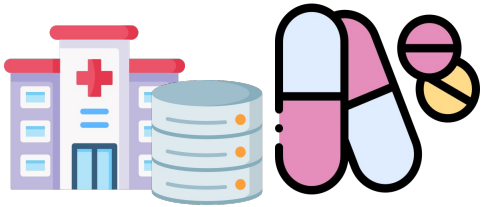
Internal and External Validity of OBS

Unbiased estimator (for CATE in RCT population)

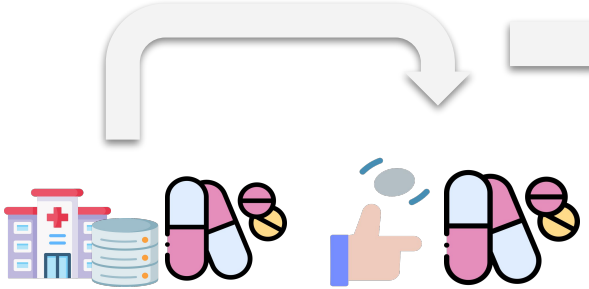
$$\psi := \psi_1 - \psi_0$$

$$\mathbb{E}[\psi|X] = 0$$

Approach



Internal Validity



External Validity

$\psi_1 \quad \psi_0$

Unbiased instance-wise CATE signals from observational study and RCT

$$H_0 : \mathbb{E}[\psi_1 - \psi_0 | X] = 0 \quad P_X - a.s.$$

CMRs

Approach: Testing the set of CMRs

$$H_0 : \mathbb{E}[\psi|X] = 0, P_X - a.s.$$

Law of Iterated Expectations

$$\mathbb{E}[\psi f(X)] = 0, \forall f \in \mathcal{F}$$

Approach: Testing the set of CMRs

$$H_0 : \mathbb{E}[\psi|X] = 0, P_X - a.s.$$

Law of Iterated Expectations

Some set of measurable functions on *X*

$$\mathbb{E}[\psi f(X)] = 0, \forall f \in \mathcal{F}$$

Approach: Testing the set of CMRs

$$H_0 : \mathbb{E}[\psi|X] = 0, P_X - a.s.$$

Law of Iterated Expectations

Muandet et al, 2020 – constrain this to be a RKHS

$$\mathbb{E}[\psi f(X)] = 0, \forall f \in \mathcal{F}$$

Approach: Testing the set of CMRs

$$H_0 : \mathbb{E}[\psi|X] = 0, P_X - a.s.$$

Law of Iterated Expectations

Muandet et al, 2020 – use the MMR within unit ball of RKHS as test statistic!

$$\mathbb{E}[\psi f(X)] = 0, \forall f \in \mathcal{F}$$

Approach: Testing the set of CMRs

$$H_0 : \mathbb{E}[\psi|X] = 0, P_X - a.s.$$

Law of Iterated Expectations

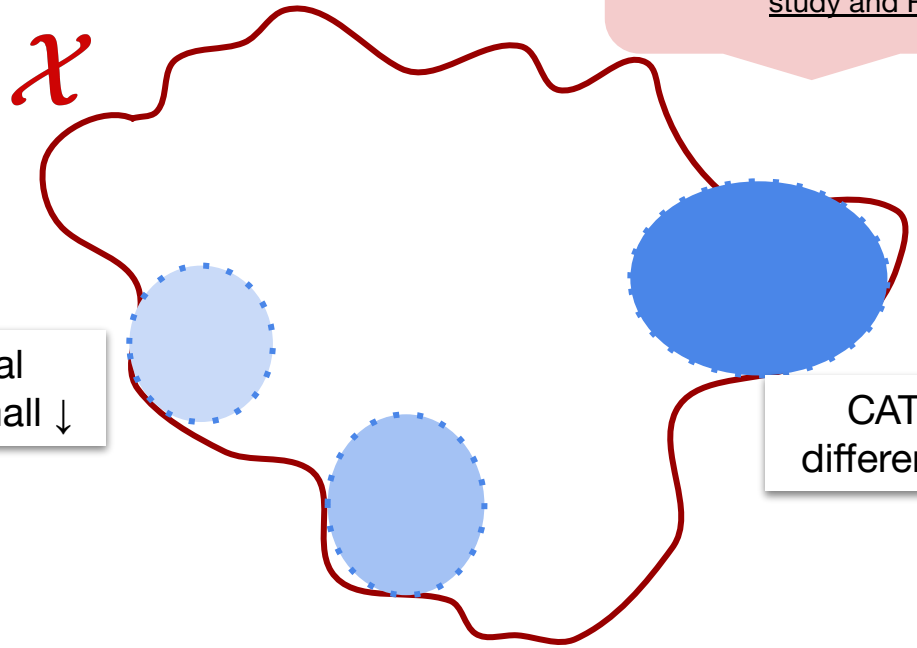
Muandet et al, 2020 – this test statistic fully captures original set of CMRs and has a nice closed-form expression!

$$\mathbb{E}[\psi f(X)] = 0, \forall f \in \mathcal{F}$$

Approach: Intuition behind MMR test statistic

$$\mathbb{M}^2 = \sup_{f \in \mathcal{F}, ||f|| \leq 1} (\mathbb{E}[\psi f(X)])^2$$

The MMR test statistic is trying to find regions in *X* where the signal difference (i.e. the difference in the CATE estimates between the observational study and RCT) is maximized



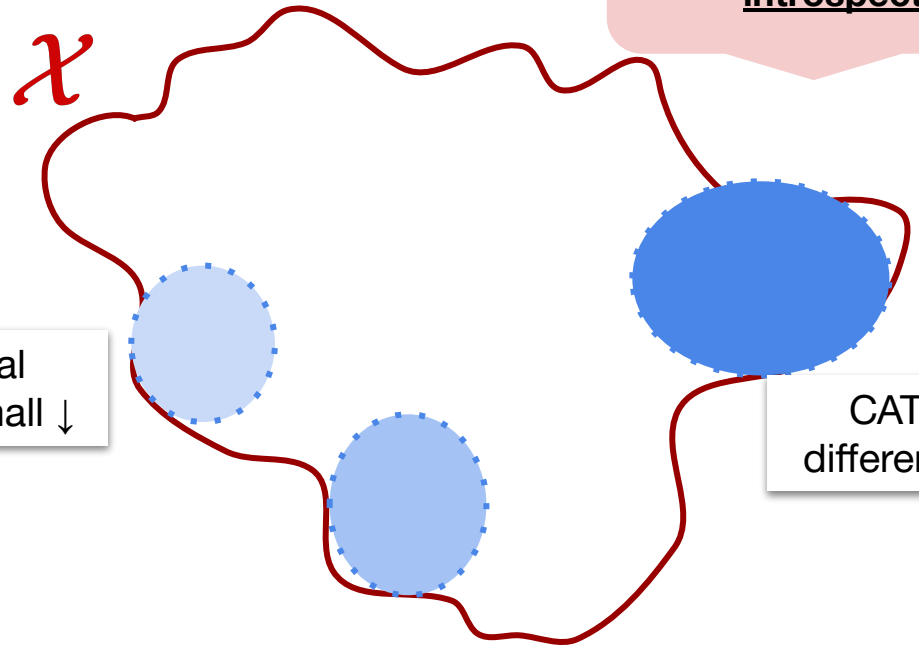
CATE signal difference small ↓

CATE signal difference large ↑

Approach: Intuition behind MMR test statistic

$$\mathbb{M}^2 = \sup_{f \in \mathcal{F}, ||f|| \leq 1} (\mathbb{E}[\psi f(X)])^2$$

Can also get the **maximizer, *f**,** which we can estimate and visualize to **introspect** into this region



CATE signal difference small ↓

CATE signal difference large ↑

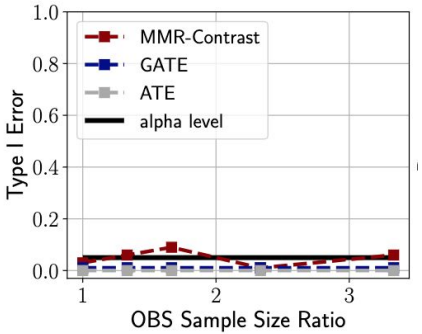
Experiments: Semi-synthetic Results w/ IHDP

RCT run on premature infants studying treatment effect of professional home visits on future cognitive function.

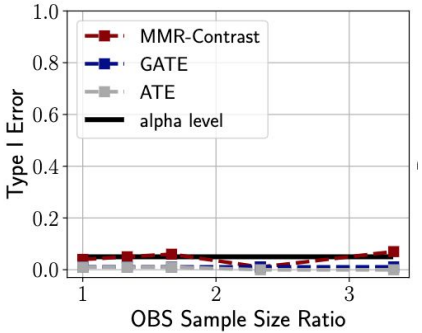
We simulate confounding to generate an OBS dataset. We also induce a difference in the covariate distributions of the RCT and OBS.

We simulate the outcomes.

Experiments: Semi-synthetic Results w/ IHDP



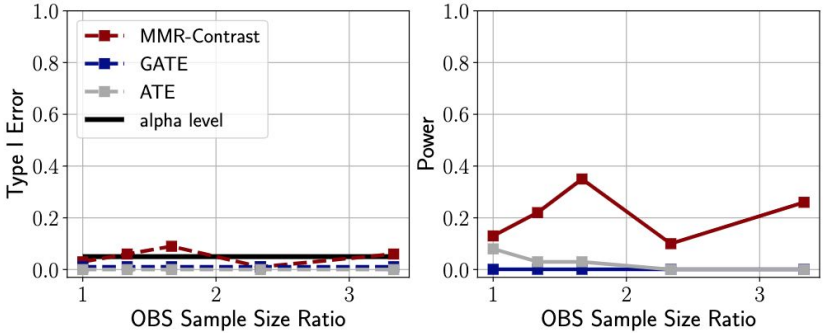
(a) Low confounder strength ($\max(\gamma) = 1.$). (left) no unobserved confounders; (right): one confounder concealed



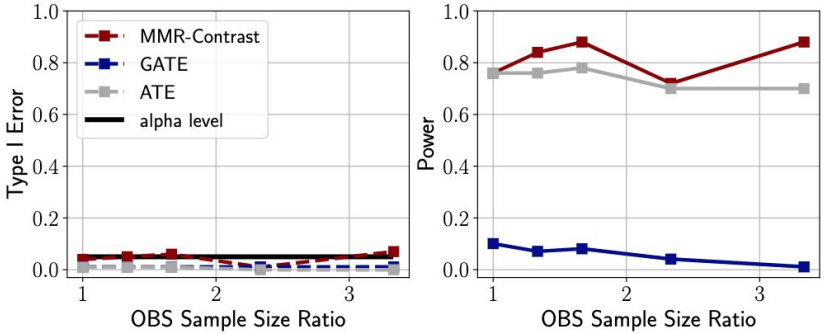
(b) High confounder strength ($\max(\gamma) = 2.75$). (left) no unobserved confounders; (right): one confounder concealed

All three approaches largely retain the nominal level of 0.05

Experiments: Semi-synthetic Results w/ IHDP



(a) Low confounder strength ($\max(\gamma) = 1.$). (left) no unobserved confounders; (right): one confounder concealed

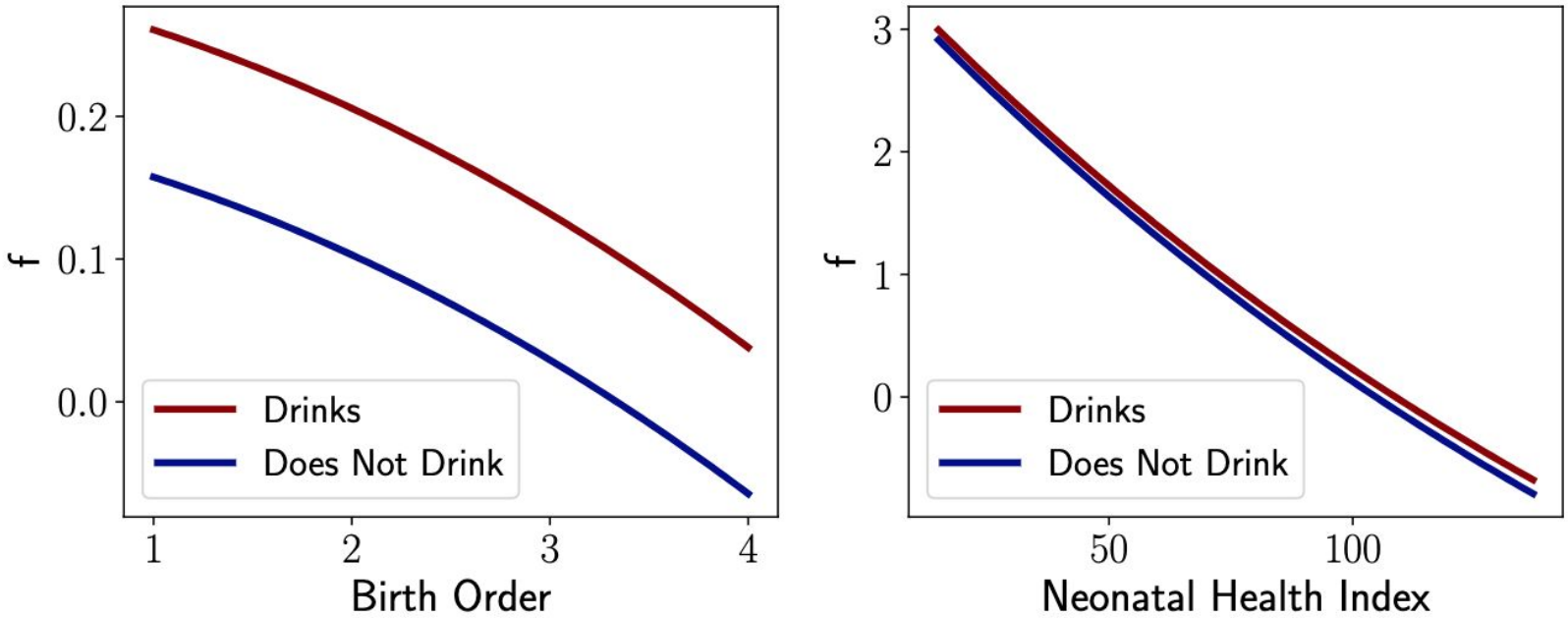


(b) High confounder strength ($\max(\gamma) = 2.75$). (left) no unobserved confounders; (right): one confounder concealed

MMR-Contrast (our method) shows superior power, particularly when confounding strength is lower

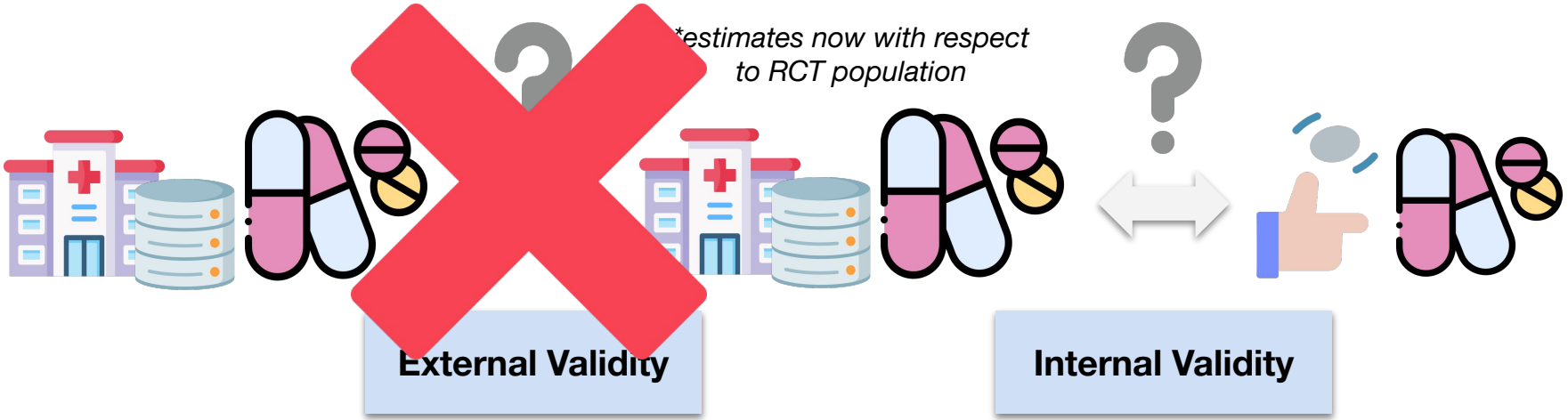
Experiments: Semi-synthetic Results w/ IHDP

Visualization of changes in witness function for pairs of covariates



Overflow

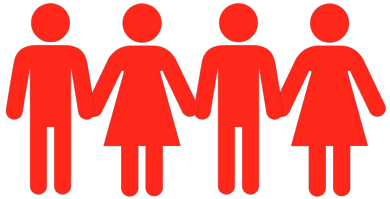
Goal: Falsification of observational estimates



Result: Use experimental data as a form of validation — "falsify" assumptions of external and internal validity in observational studies

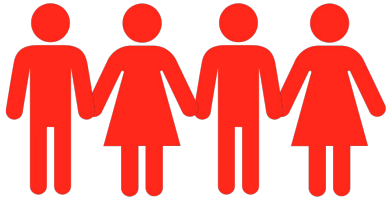
Motivating Example

Randomized Controlled Trial **(RCT)**



Motivating Example

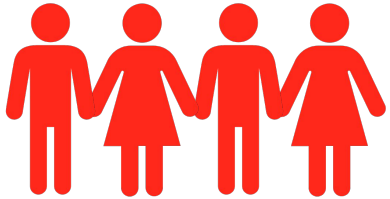
Randomized Controlled Trial **(RCT)**



RCTs often fail to include all types of patients (e.g.)

Motivating Example

Randomized Controlled Trial **(RCT)**



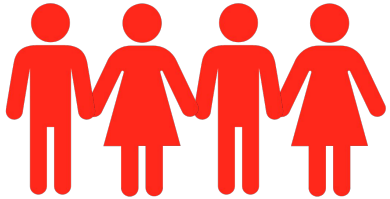
Real-world example: pregnant women were not included in initial COVID-19 trials[1]

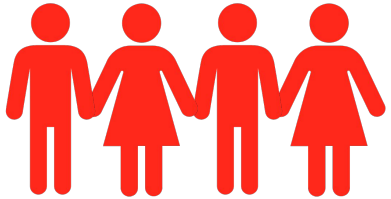
1] Dagan, Noa, et al. "Effectiveness of the BNT162b2 mRNA COVID-19 vaccine in pregnancy." *Nature medicine* 27.10 (2021): 1693-1695.

Motivating Example

Randomized Controlled Trial **(RCT)**

Observational Study **(OS)**





Observational studies contain a more **diverse cohort**, but may suffer from e.g. **unobserved confounding**.

Motivating Example

RCT

Motivating Example

Observational Study #1	Observational Study #2	RCT

Motivating Example

	Observational Study #1	Observational Study #2	RCT
			
			

Motivating Example

	Observational Study #1	Observational Study #2	RCT
			
			

Motivating Example

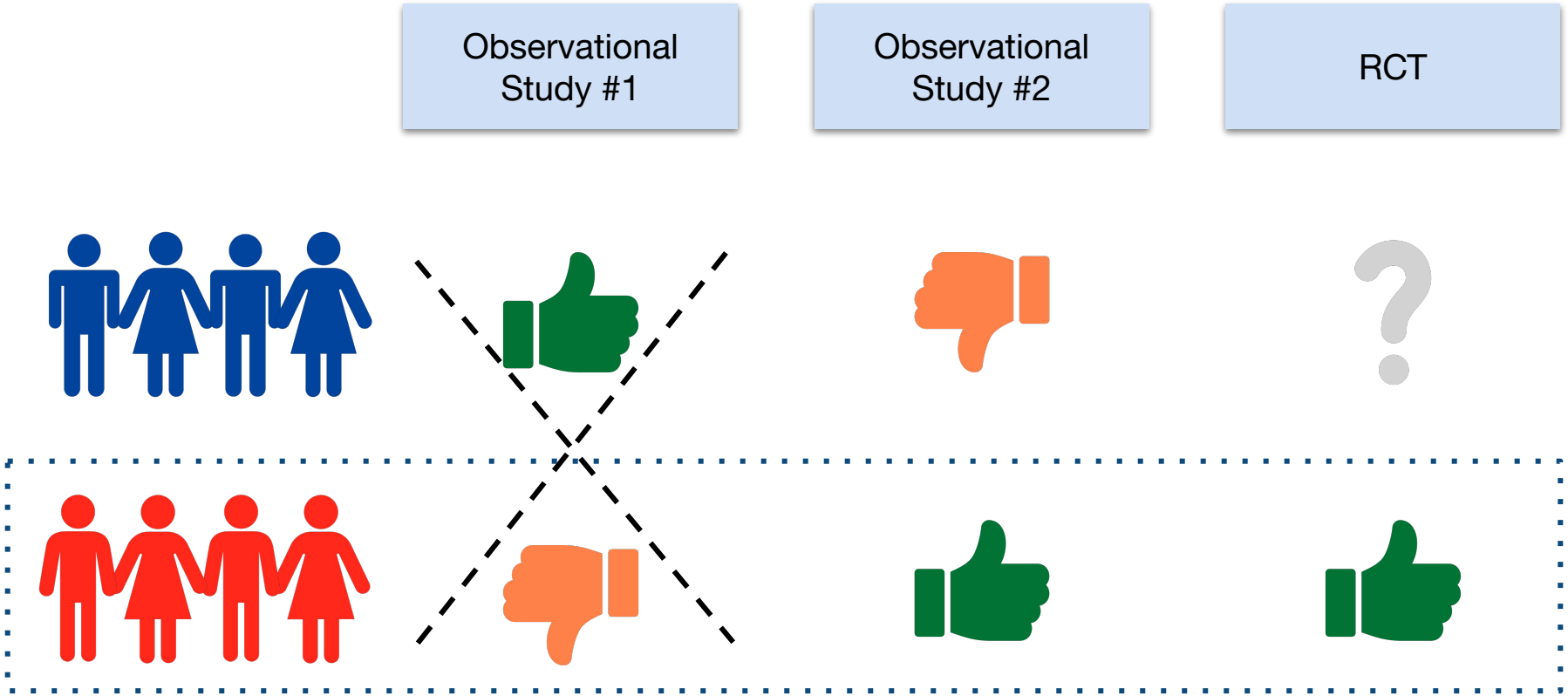
Observational Study #1	Observational Study #2	RCT



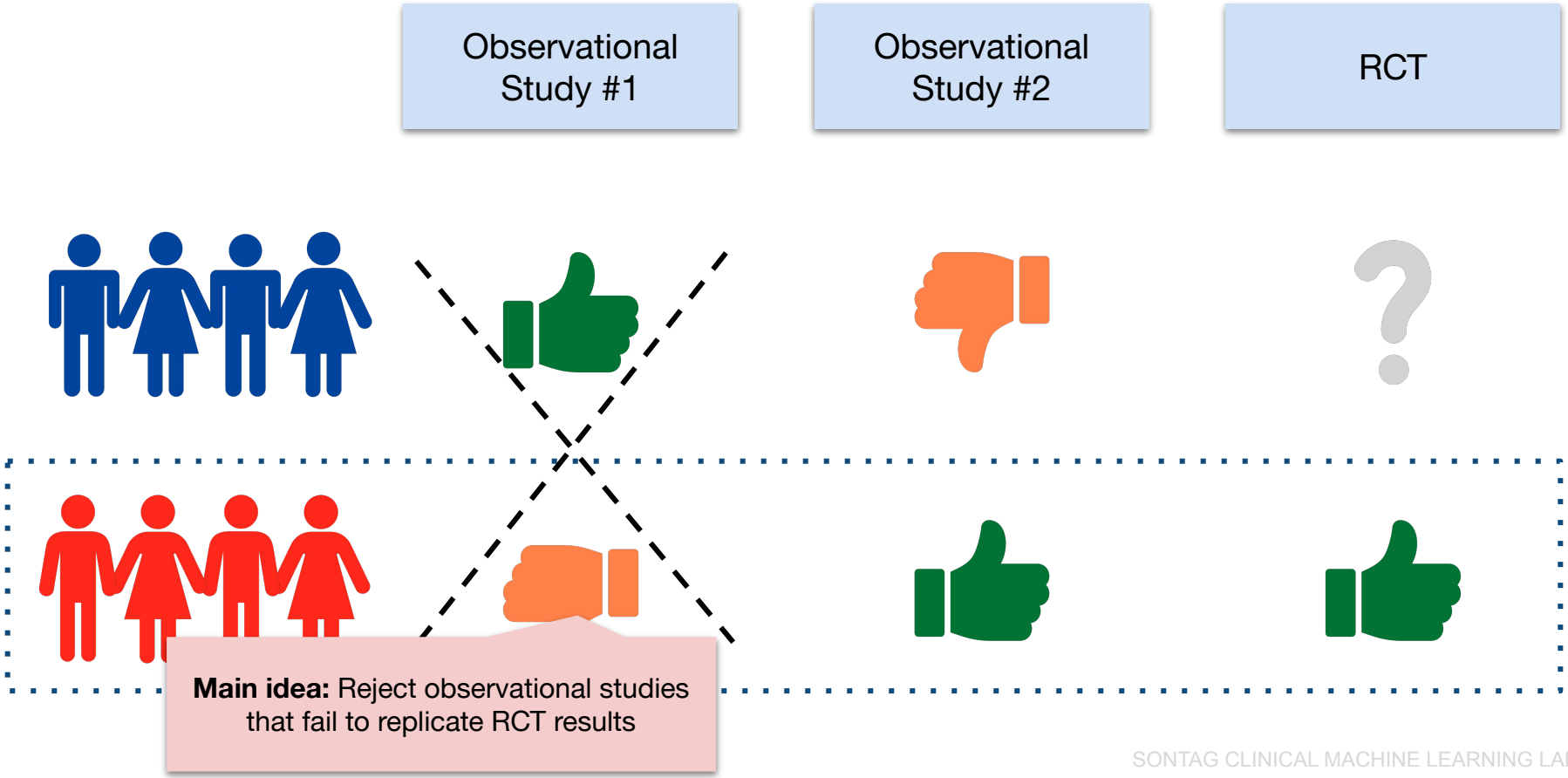




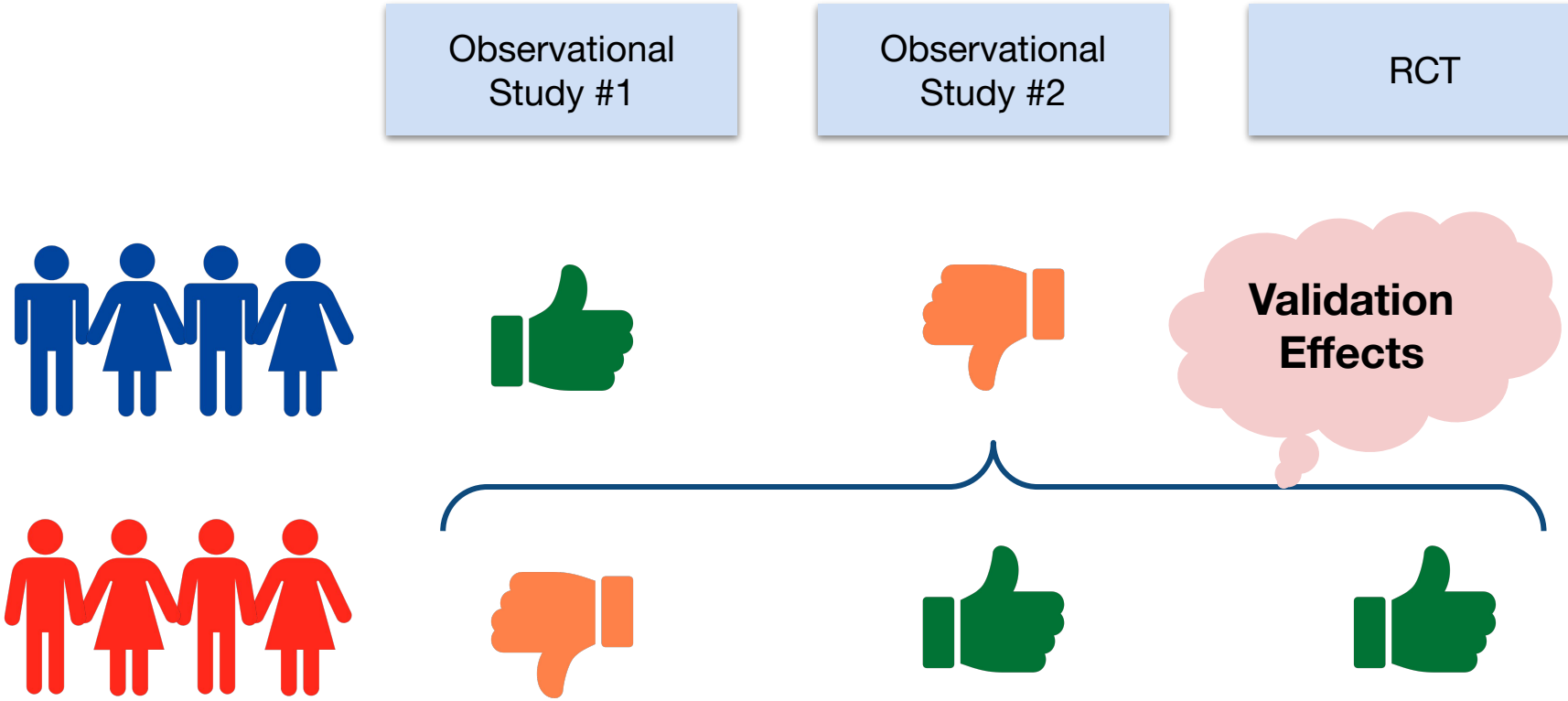
Motivating Example



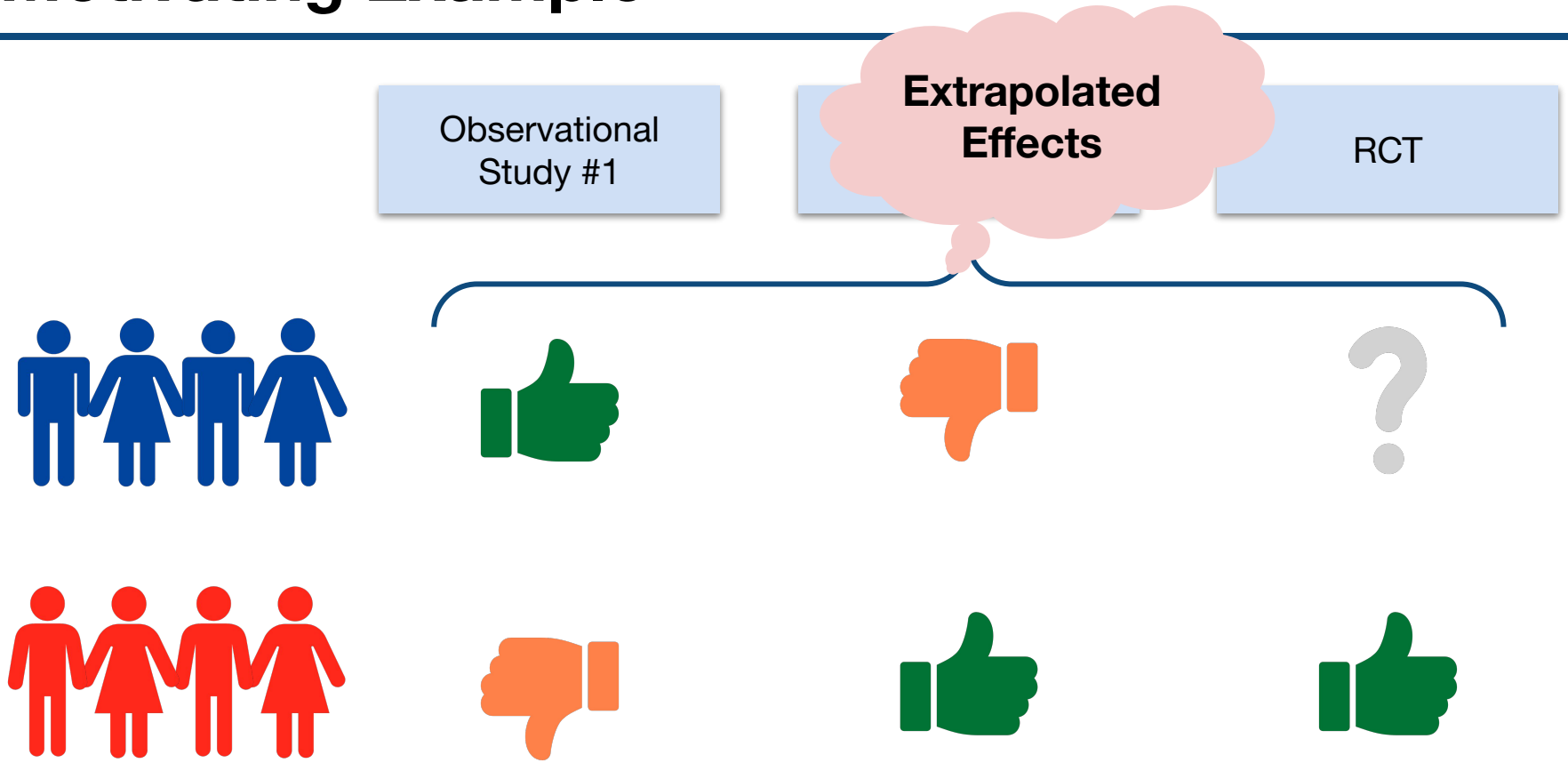
Motivating Example



Motivating Example



Motivating Example



Contributions

Our Approach

1 Falsification of observational estimates

Contributions

Our Approach

1

Falsification of observational estimates

Use framework of **hypothesis testing**

Contributions

Our Approach

1

Falsification of observational estimates

Use framework of **hypothesis testing**

Reject estimators that do not replicate **RCT estimates**

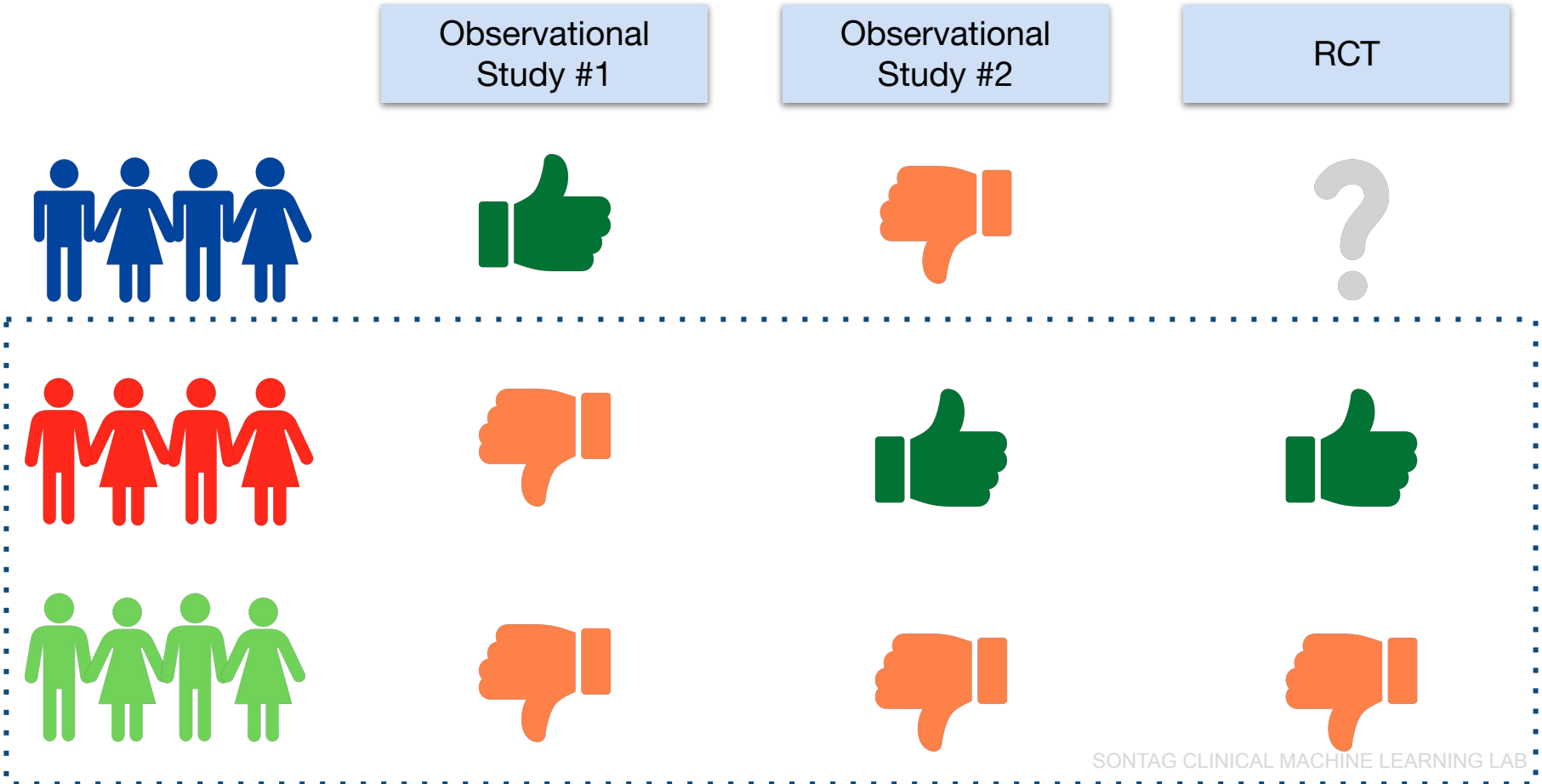
Contributions

Our Approach

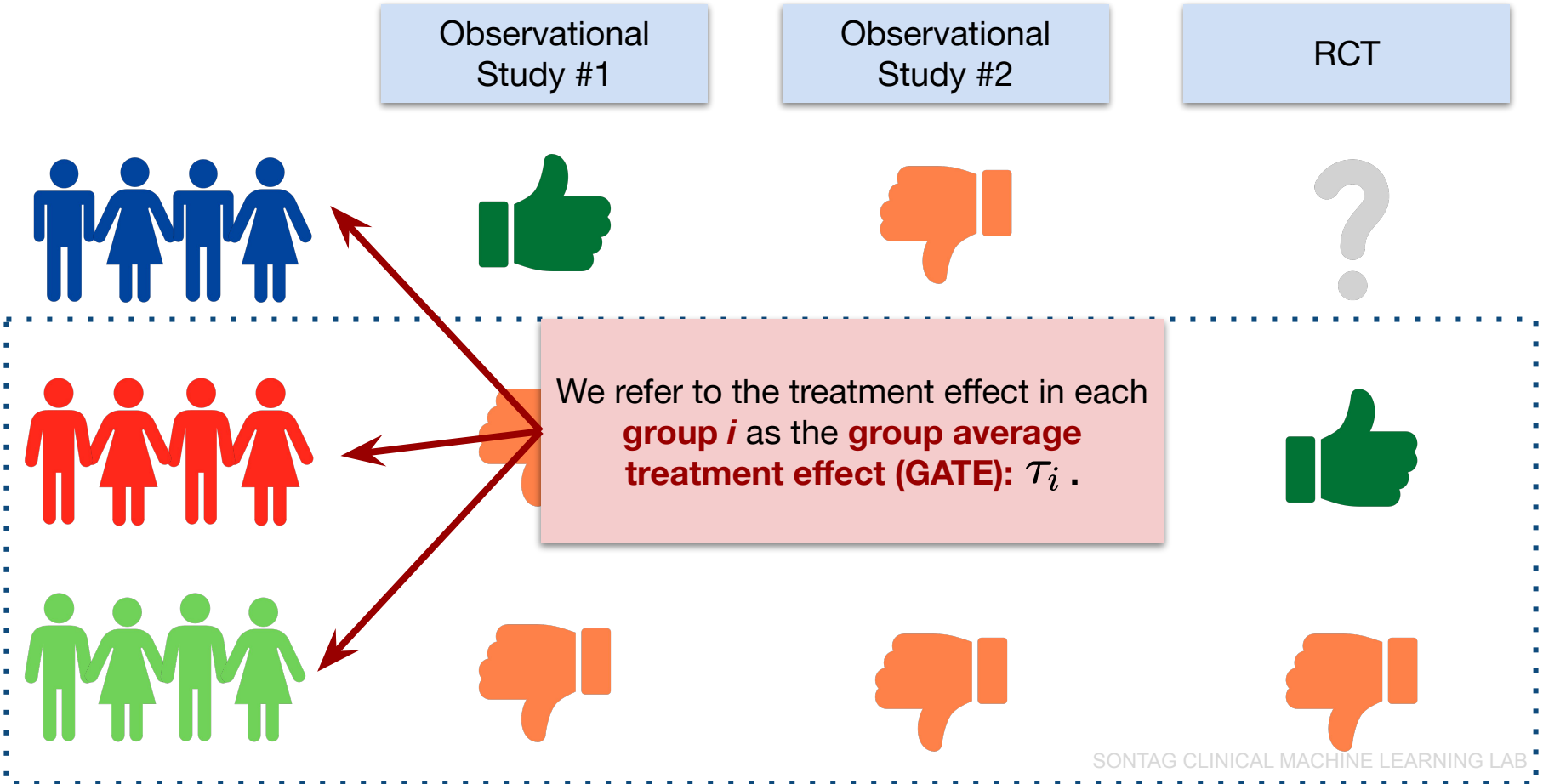
1. Falsification of observational estimates
2. Pessimistic Combination of Confidence Intervals

Take the **union** over all the intervals of the **accepted estimators**.

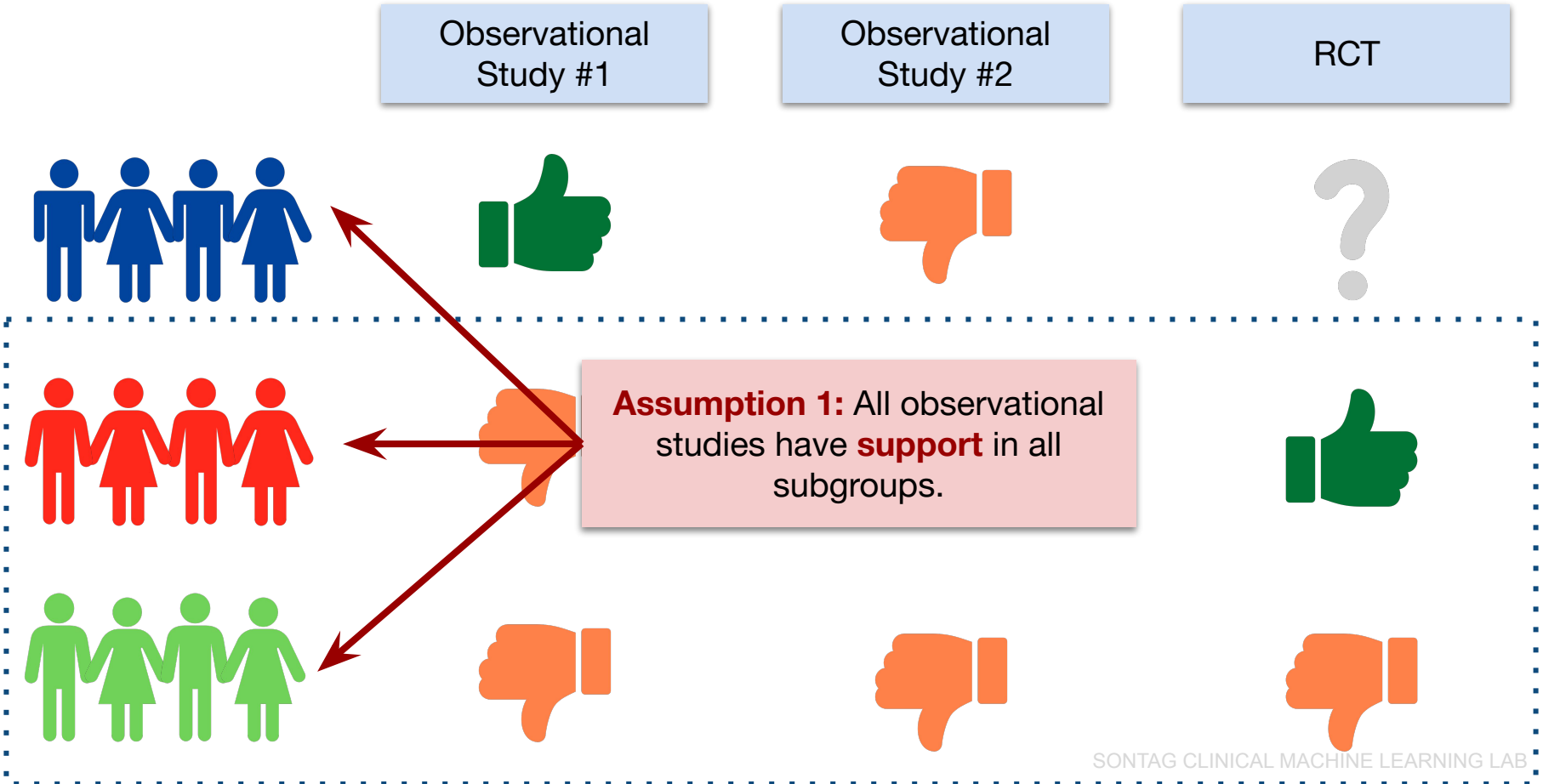
Formalizing Falsification



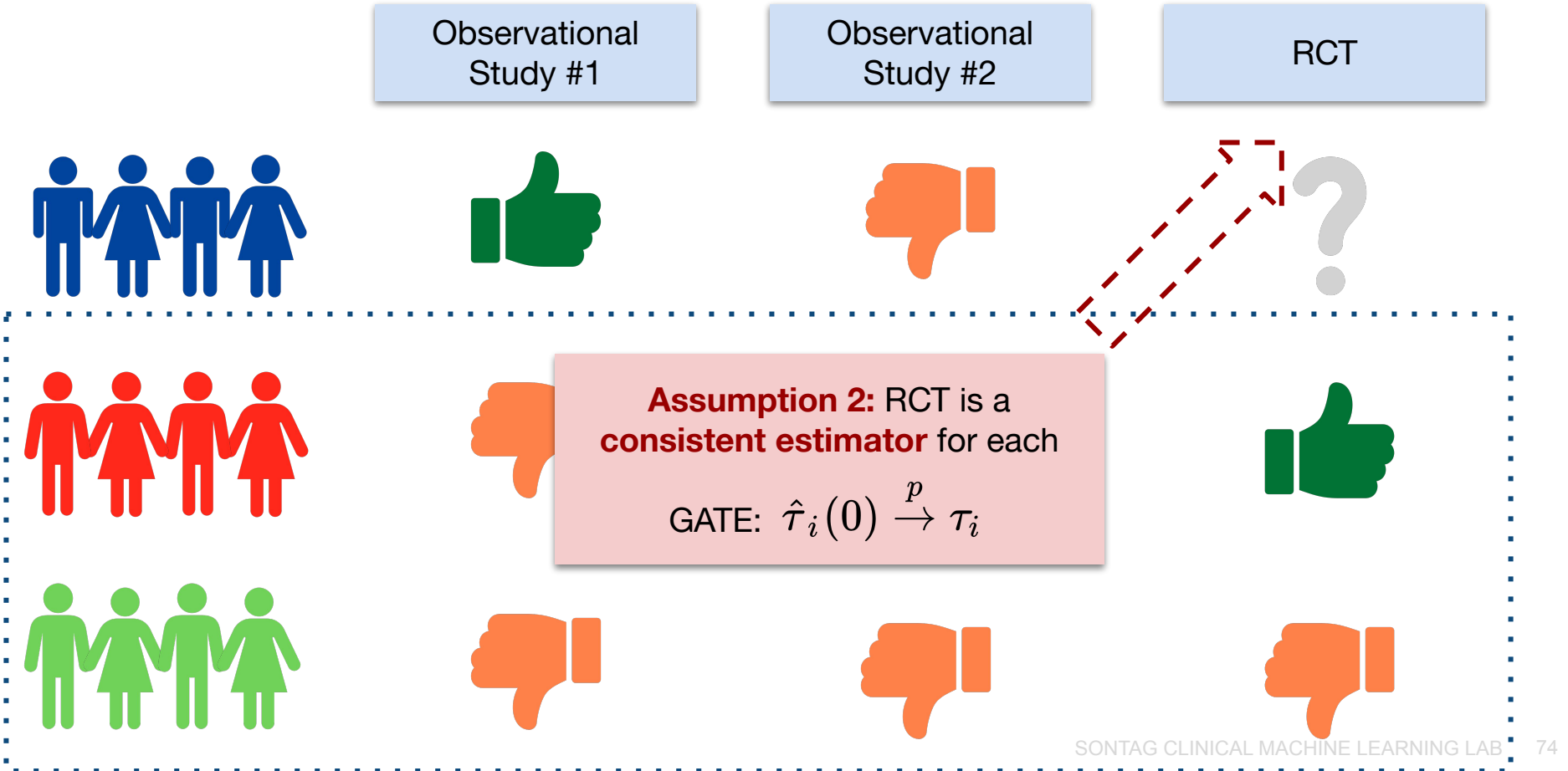
Formalizing Falsification



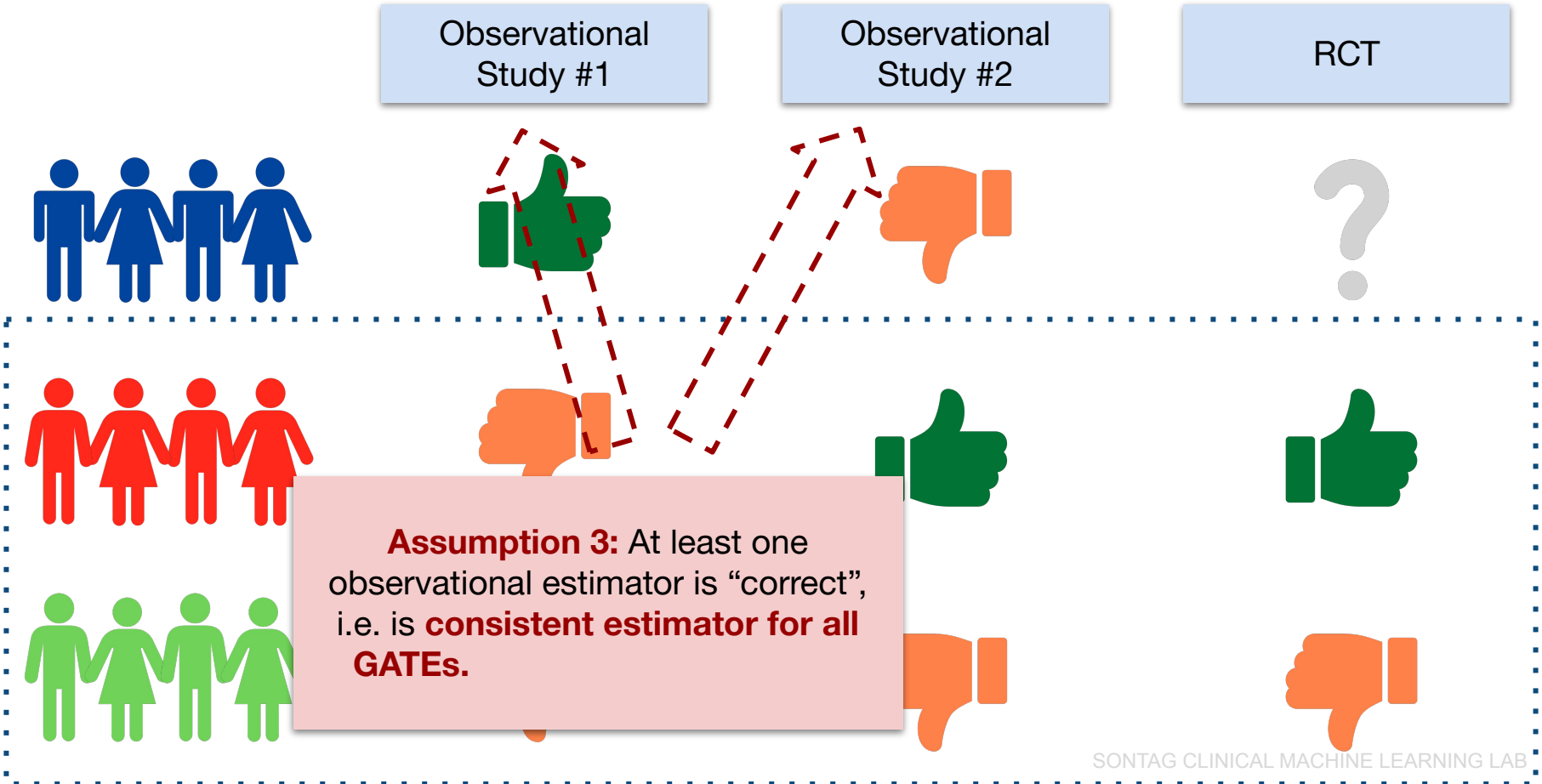
Formalizing Falsification



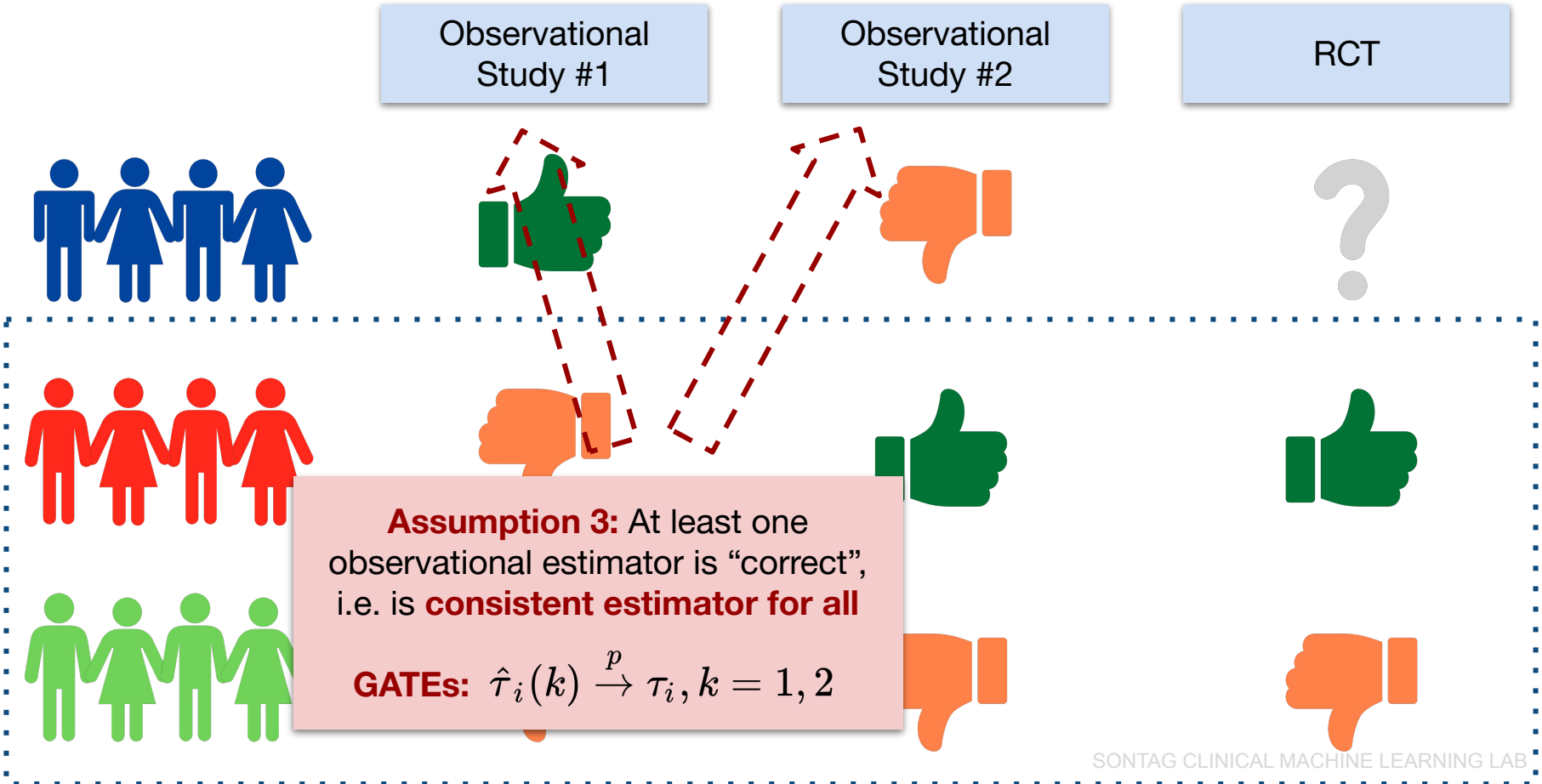
Formalizing Falsification



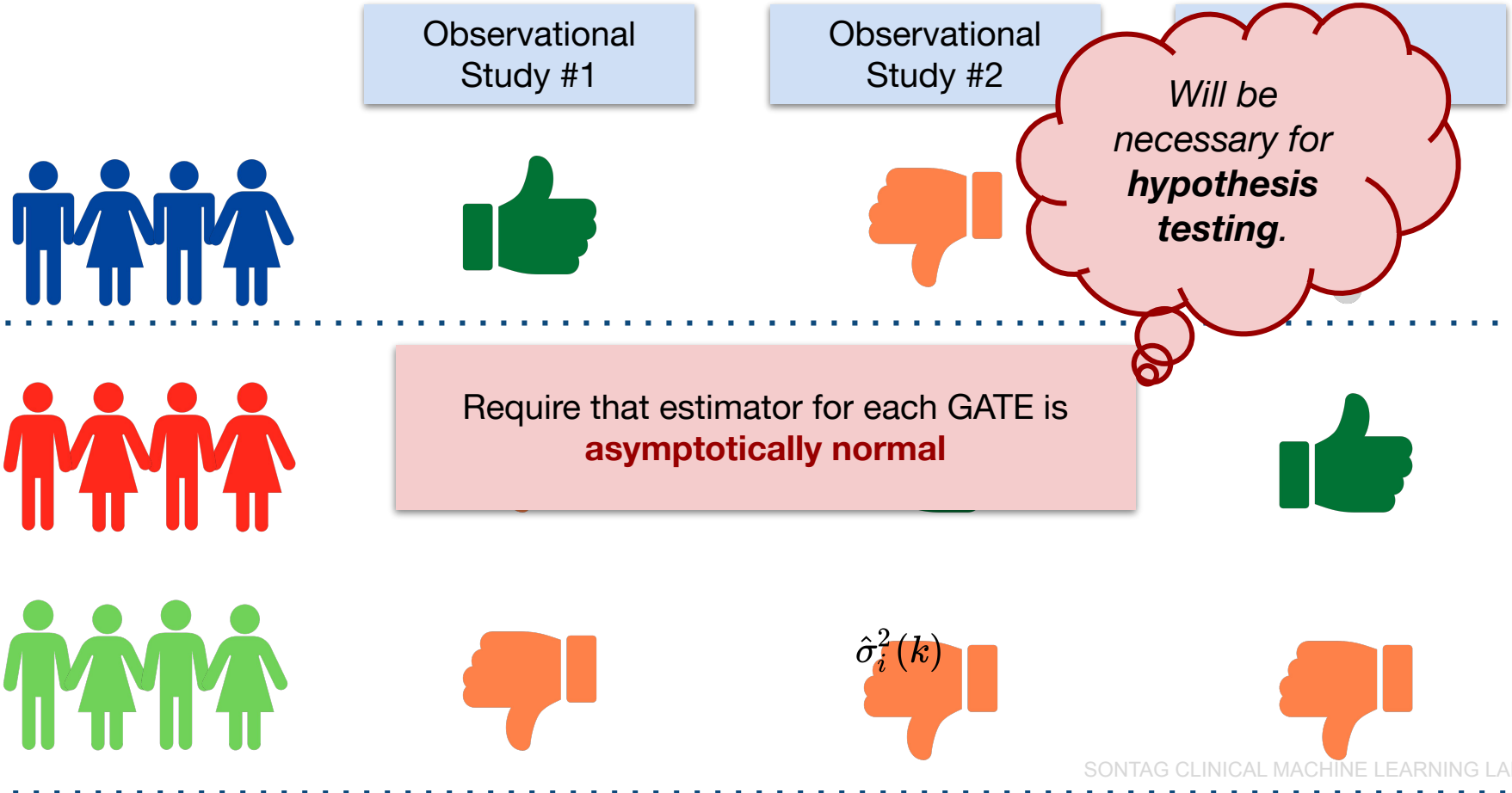
Formalizing Falsification



Formalizing Falsification



Formalizing Falsification



Formalizing Falsification

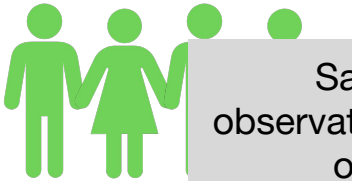
Observational Study #1

Obser

*Will be necessary for **hypothesis testing**, and we give examples where this is reasonable.*

Require that estimator for each GATE is **asymptotically normal**

$$\sqrt{N_k}(\hat{\tau}_i(k) - \tau_i(k))/\hat{\sigma}_i(k) \xrightarrow{d} \mathcal{N}(0,1)$$



Sample size of observational study (k=1,2) or RCT (k=0)

$\hat{\sigma}_i^2(k)$ is estimate of variance, converges in probability to asymptotic variance

Formalizing Falsification

Observational Study #1

Obser

*We demonstrate asymptotic normality of GATE estimators with **transportation.***

Require that estimator for each GATE is **asymptotically normal**

Hypothesis Test Construction

Observational Study #1

Observational Study #2

RCT

$$H_0 : \tau_{\text{red}}(1) = \tau_{\text{red}}$$

Want to perform above **hypothesis test** with **asymptotic level**, α

Hypothesis Test Construction

Observational Study #1

Observational Study #2

RCT

$$H_0 : \tau_{\text{red}}(1) = \tau_{\text{red}}$$

Set equal to 0

We can use the following test statistic, which we show **converges in distribution** to a standard **normal distribution**

$$\hat{T}_N(k = 1, i = \text{red people}) := \frac{(\hat{\tau}_i(1) - \hat{\tau}_i(0)) - (\tau_i(1) - \tau_i)}{\frac{\hat{\sigma}_i^2(1)}{N_1} + \frac{\hat{\sigma}_i^2(0)}{N_0}}$$

Estimated variance

Hypothesis Test Construction

Observational Study #1

Observational Study #2

RCT

$$H_0 : \tau_{\text{red}}(1) = \tau_{\text{red}}$$

Set equal to 0

Reject the observational study if

$$|\hat{T}_N(k=1, i=\text{red people})| > z_{\alpha/2}$$

$$\hat{T}_N(k=1, i=\text{red people}) := \frac{(\hat{\tau}_i(1) - \hat{\tau}_i(0)) - (\tau_i(1) - \tau_i)}{\frac{\hat{\sigma}_i^2(1)}{N_1} + \frac{\hat{\sigma}_i^2(0)}{N_0}}$$

Estimated variance

Hypothesis Test Construction

Observational Study #1

Observational Study #2

RCT

$$H_0 : \tau_{\text{red}}(1) = \tau_{\text{red}}$$

Set equal to 0

Note that we use **Bonferroni correction to control FPR of test**, since we test many subgroups (e.g. red people, blue people, etc.)

$$\hat{T}_N(k = 1, i = \text{red people}) := \frac{(\hat{\tau}_i(1) - \hat{\tau}_i(0)) - (\tau_i(1) - \tau_i)}{\frac{\hat{\sigma}_i^2(1)}{N_1} + \frac{\hat{\sigma}_i^2(0)}{N_0}}$$

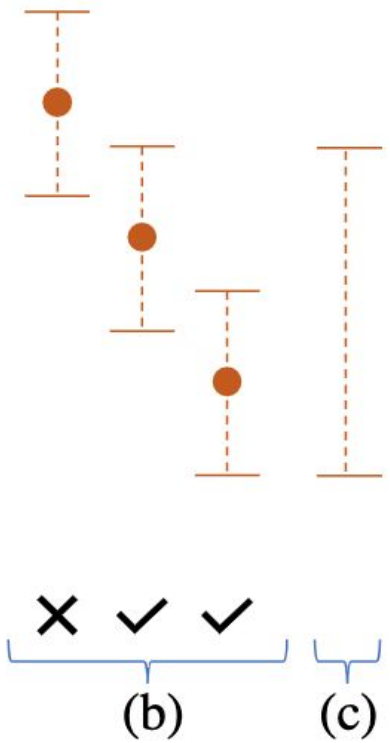
Estimated variance

Our Approach

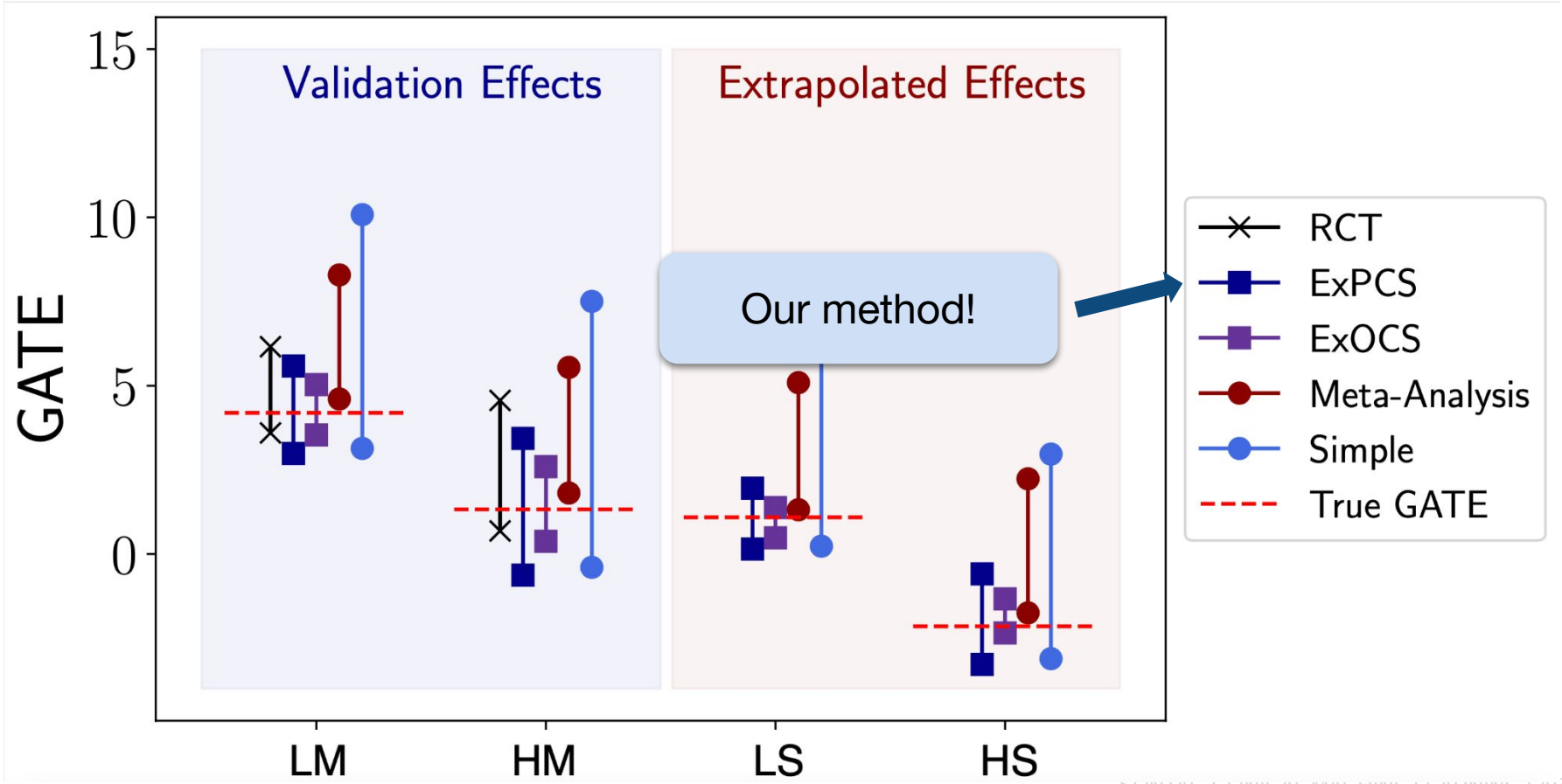
1. Falsification of observational estimates
2. Pessimistic Combination of Confidence Intervals

2

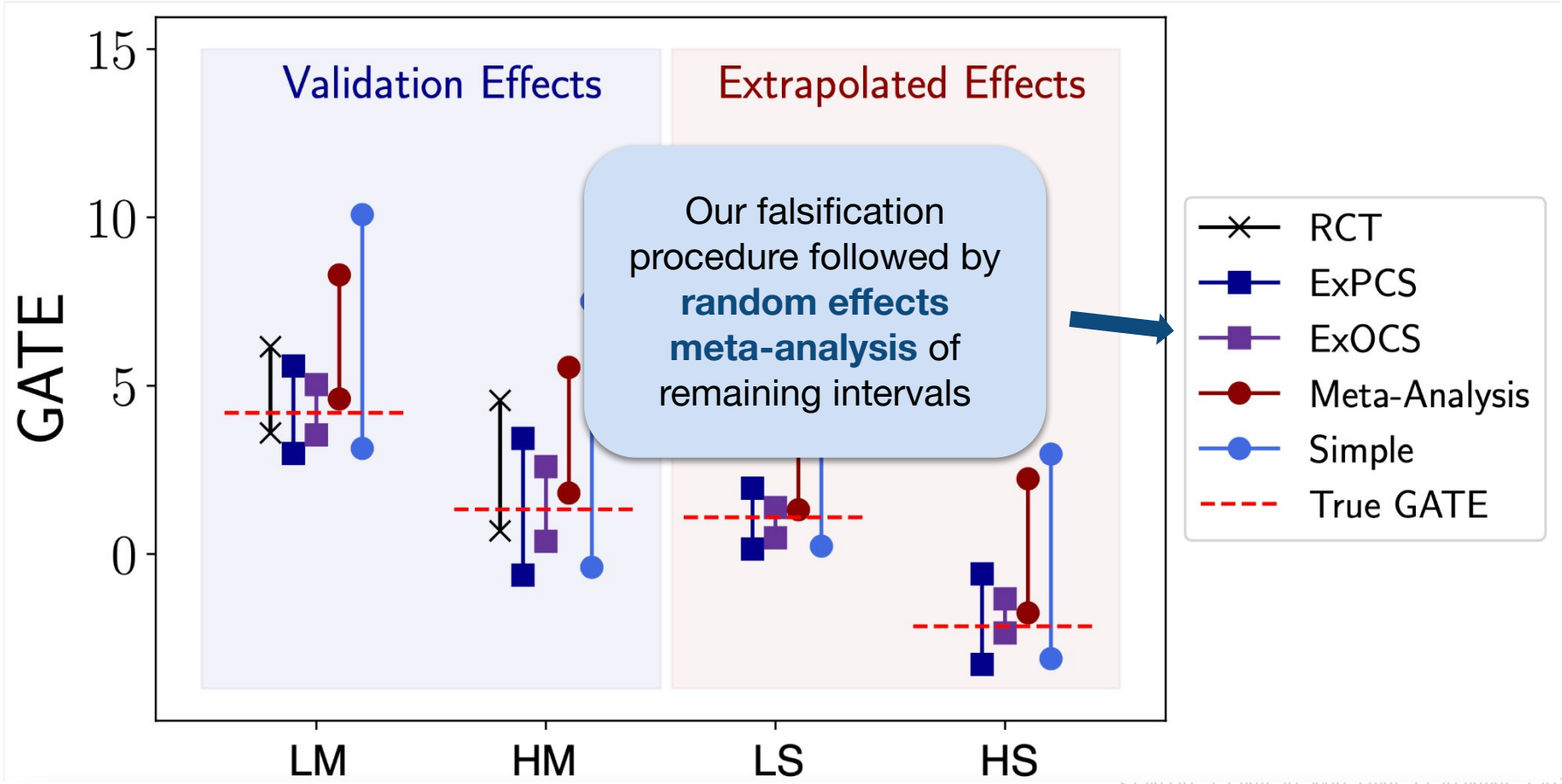
Pessimistic Combination of Confidence Intervals



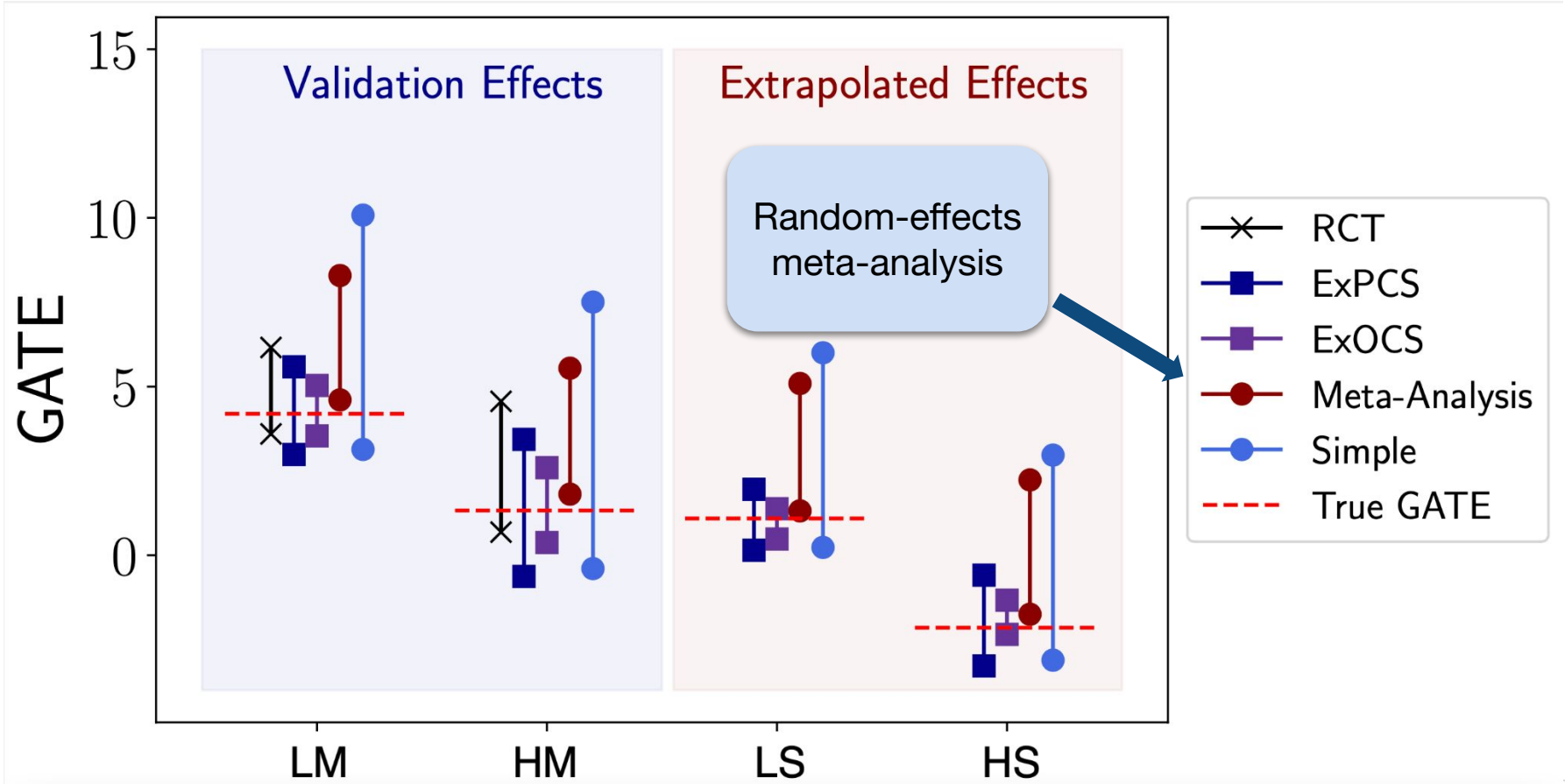
Results on Semi-Synthetic



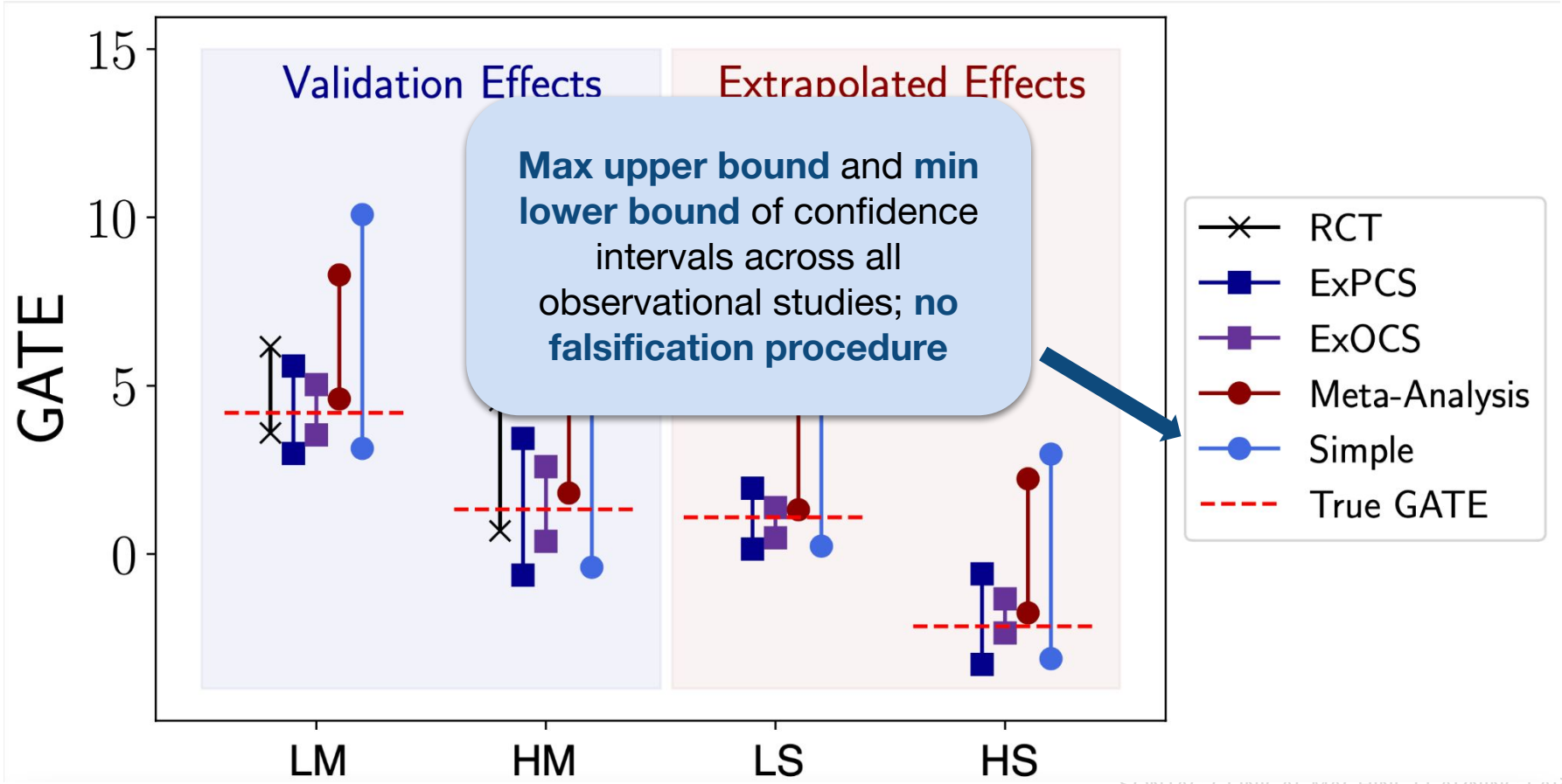
Results on Semi-Synthetic



Results on Semi-Synthetic

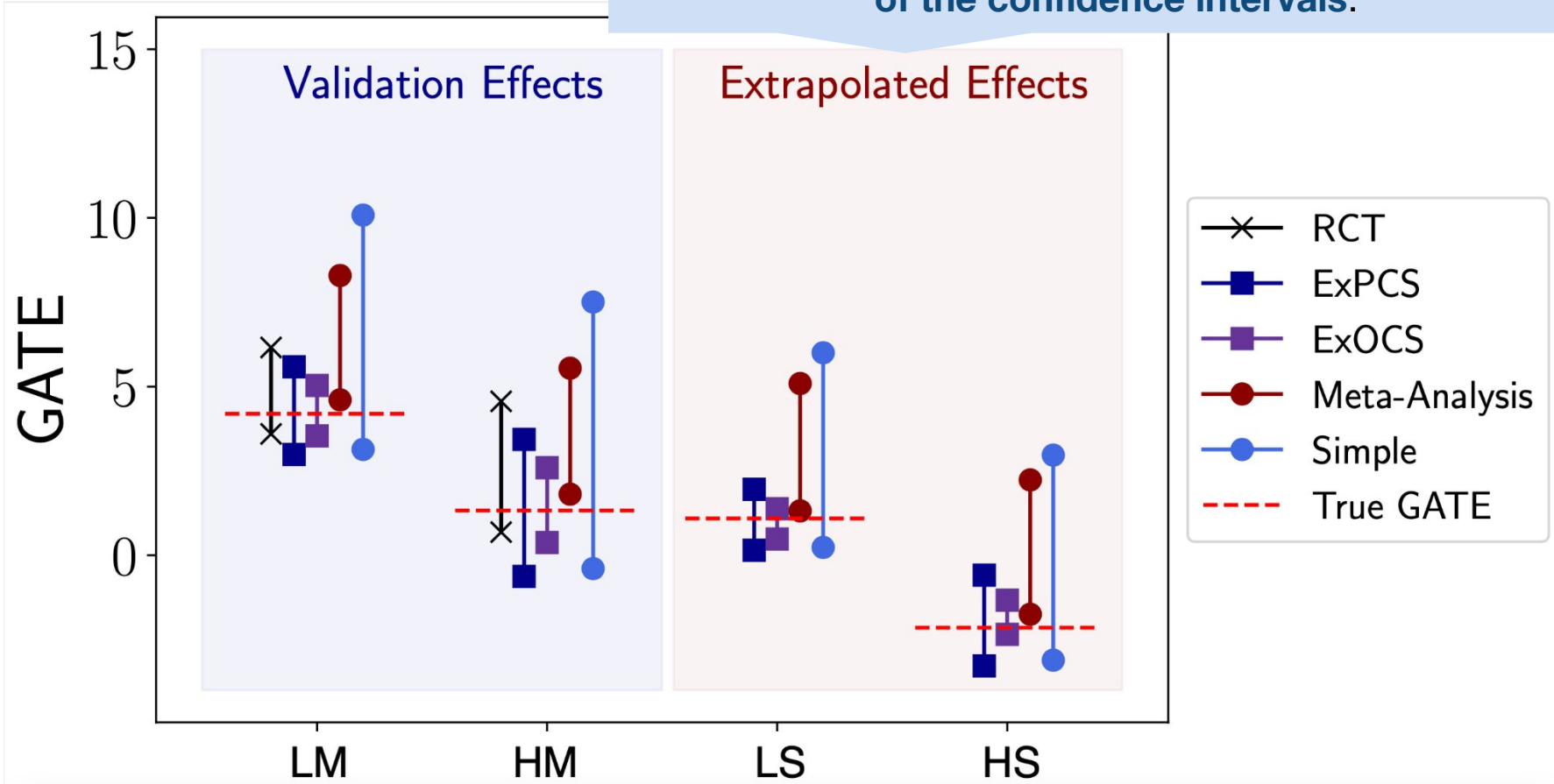


Results on Semi-Synthetic



Results on Semi-S

Compared to baselines, our approach has the best balance between **coverage of the true GATE** and **width of the confidence intervals**.



For more results and discussion, visit us at poster ID 54677!

Thank you!

Results on Women's Health Initiative Data

	Coverage	Length	OS %
Simple	0.39	0.416	–
Meta-Analysis	0.03	0.260	–
ExOCS	0.28	0.058	–
ExPCS (ours)	0.45	0.081	0.99
Oracle	0.44	0.068	–

Table 1: Coverage, length, and unbiased OS % of ExPCS and baselines. ExPCS achieves comparable coverage to the oracle method with highly efficient intervals. Additionally, we do not reject the unbiased OS in 99% of the tasks.