# Information-Theoretic Causal Bounds under Unmeasured Confounding

*Working paper*

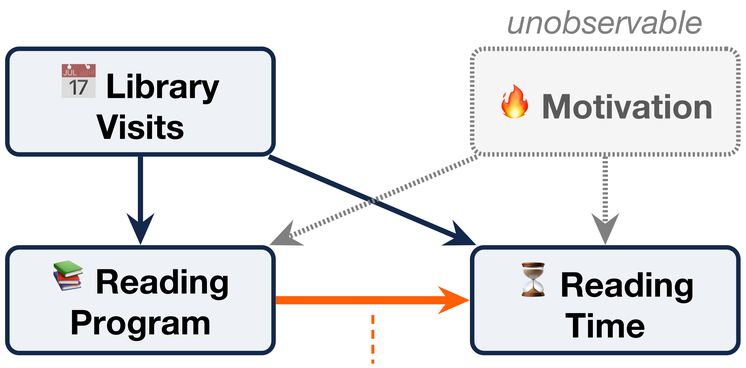
**Junhyung Ahn** · **Yonghan Jung**

School of Information Sciences, University of Illinois Urbana-Champaign

## 1 Problem

### Setting

A library runs a reading program.
Students choose to join or not.
We record weekly reading time.




### Our Interest

Effect of the reading program on reading time

### Challenge — Hidden Bias

Joined students: Avg Reading Time = 5 hr
Motivated students join more and read more
→ **5 hr is not the program's effect alone**

### Question

**What if** all students had joined?

**Observed World** ⟷ **What-If World**

**5 hr** ⟷ **? hr** *(no-data)*

**What-If World:** every student joins, so motivation no longer decides who joins (want to see the true effect of the program on reading time)

## 2 Solution

### Result

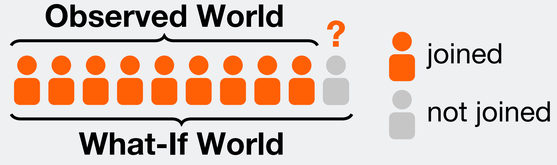
Novel

We can bound the difference between the Observed and What-If Worlds:

$$\mathrm{Diff}(\mathrm{Observed\ World}, \mathrm{What\text{-}If\ World}) \leq \underbrace{f(\Pr(\mathrm{Join} \mid \mathrm{Visits}))}_{\text{computable from the Observed World}}$$

### Intuition

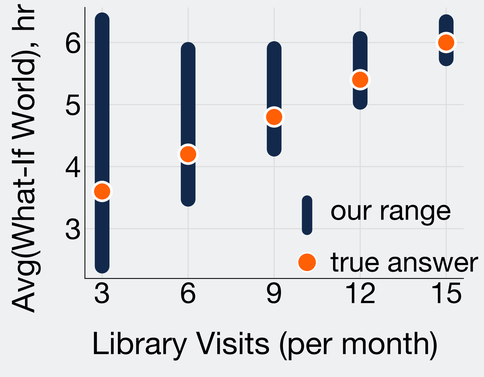
Students with 15 library visits a month:
90% of them join (computable from data).




Only 1 of 10 is unknown → Small Diff

**More joined → Smaller Diff**

### Simulation

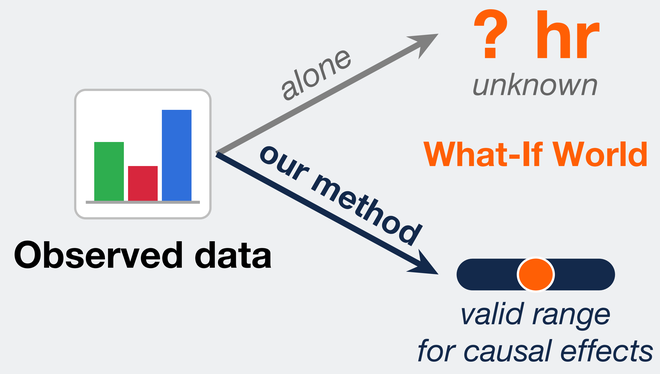



- The true answer stays inside our range
- More visits → more joined → narrower range

## 3 Conclusion

### Impact

Causal inference is not possible
from observation alone,
**yet we provide a valid range**
**for causal effects.**




**A new way of data analysis**
**for causal inference**
**without assumptions**

### Summary

- **New:** the first general approach connecting the Observed and What-If Worlds without assumptions
- **Validity:** the range surely holds the causal answer
- **Future Direction:** Assumption-lean Causal Generative AI using Data