Proximal Causal Inference with Text Data

NeurIPS 2024, Vancouver, Canada

Jacob M. Chen, Rohit Bhattacharya,
Katherine A. Keith

Williams College

Johns Hopkins University

Slides credit: Rohit Bhattacharya

Proximal Causal Inference with Text Data

Jacob M. Chen
Department of Computer Science
Johns Hopkins University
jchen459@jhu.edu

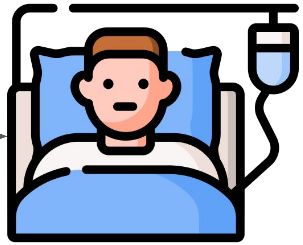
Rohit Bhattacharya
Department of Computer Science
Williams College
rb17@williams.edu

Katherine A. Keith
Department of Computer Science
Williams College
kak5@williams.edu

Answering a causal question

Does the treatment improve patient outcomes?

Treatment



Patient outcome

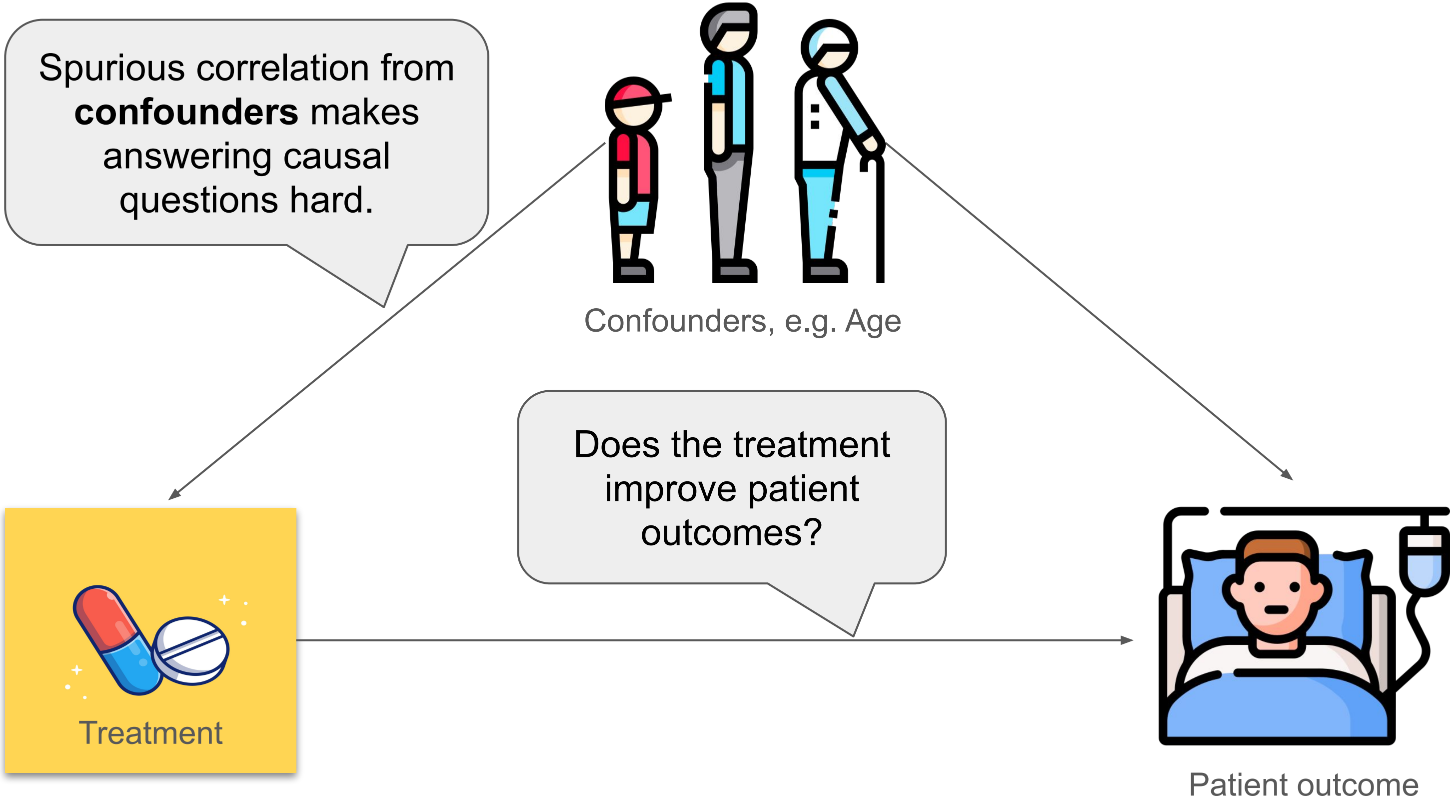
- We are interested in estimating the **average causal effect (ACE)** — quantifies the mean difference in outcomes under two different interventions

Intervene to give patient treatment

Intervene to not give patient treatment

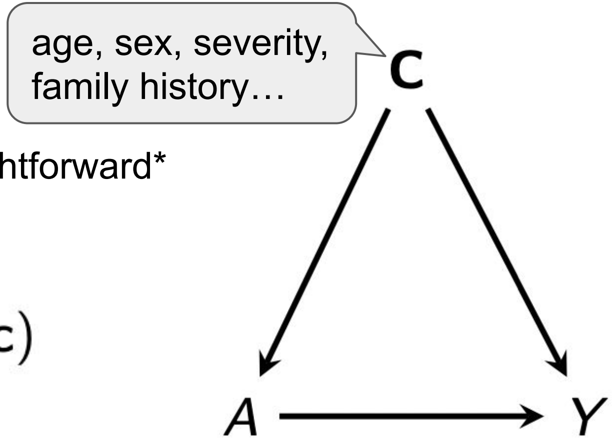
$$\text{ACE} = \mathbb{E}[Y \mid \text{do}(A = 1)] - \mathbb{E}[Y \mid \text{do}(A = 0)]$$

Casual questions are difficult to answer!



Identification when all confounders are observed

- Say we observe all relevant confounding variables ***C*** in our data
- Then identification and estimation of the ACE is (relatively) straightforward*



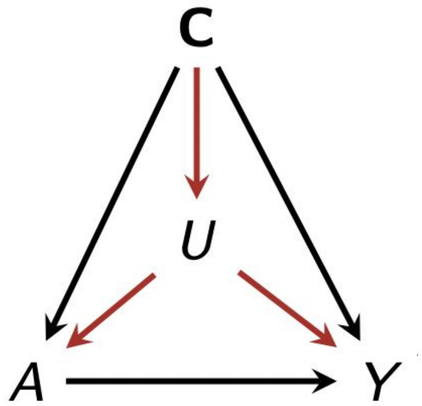
$$\mathbb{E}[Y \mid \mathrm{do}(A = a)] = \sum_{\mathbf{c}} \mathbb{E}[Y \mid A = a, \mathbf{c}] \times p(\mathbf{c})$$

Backdoor adjustment aka g-formula [Robins (1986), Pearl (1995)]

* Also requires *p(A|C) > 0* and consistency + estimation can be tricky in high-dimensional settings

But unmeasured confounding poses serious issues

- Say we observe all relevant confounding variables ***C*** in our data
- Then identification and estimation of the ACE is (relatively) straightforward*



$$\mathbb{E}[Y \mid \mathrm{do}(A = a)] = \sum_{\mathbf{c}} \mathbb{E}[Y \mid A = a, \mathbf{c}] \times p(\mathbf{c})$$

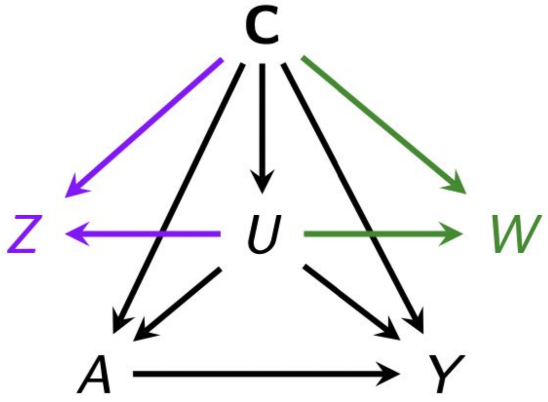
Backdoor adjustment aka g-formula [Robins (1986), Pearl (1995)]

No longer works, gives biased estimates

* Also requires *p(A|C) > 0* and consistency + estimation can be tricky in high-dimensional settings

Identification with proxies: Proximal causal inference

- Say we observe two proxies *W* and *Z* of *U* such that
  - **(P1)** They are conditionally independent: $W \perp Z \mid U, \mathbf{C}$
    - "*W*, *Z* are independent sources of information about *U*"
  - **(P2)** *W* is independent of the treatment: $W \perp A \mid U, \mathbf{C}$
    - "One proxy is independent of the treatment."
  - **(P3)** *Z* is independent of the outcome: $Z \perp Y \mid A, U, \mathbf{C}$
    - "One proxy is independent of the outcome."
  - **(P4)** Completeness
    - "*W*, *Z* encode sufficient info about *U*"

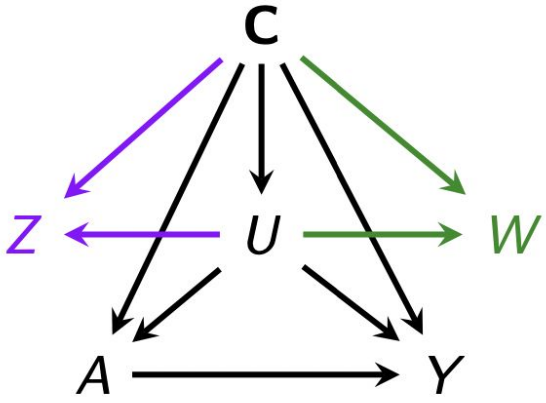


$$\mathbb{E}[Y \mid \mathrm{do}(A = a)] = \sum_{w,\mathbf{c}} h(a, w, \mathbf{c}) \times p(w, \mathbf{c})$$

- Then identification is possible by a more complex functional*

Proximal g-formula [Tchetgen Tchetgen et al (2020), Kuroki & Pearl (2014)]

Finding proxies in structured data is difficult



$Z \rightarrow Y$ and $W \rightarrow A$ are likely to exist in many cases in real-world data, violating P2, P3 of proximal

Goal: Construct Z and W in such a way that proximal assumptions hold ***by design***

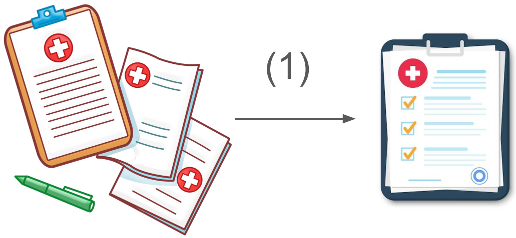
Our proposed solution: zero-shot prediction from text

Using text data indicative of U, use large language models to make predictions for U and use those predictions as the proxies in proximal causal inference.

Also check using a **heuristic** whether the proxies satisfy P1-P4.

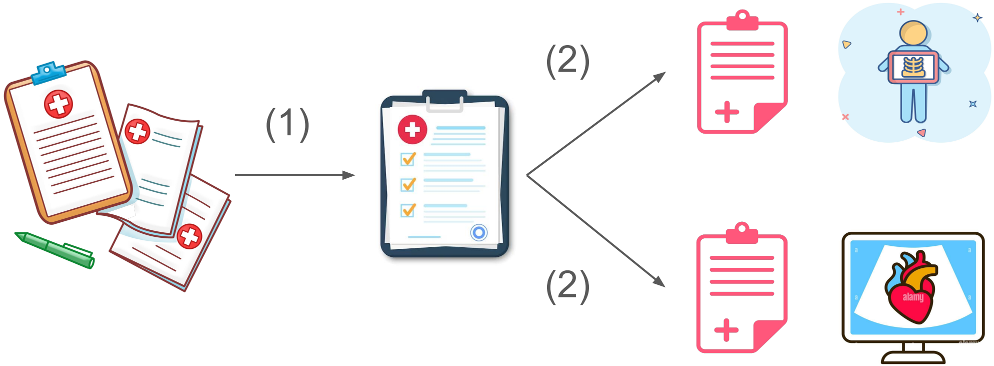
Our procedure for generating proxies for each patient

Our procedure for generating proxies for each patient



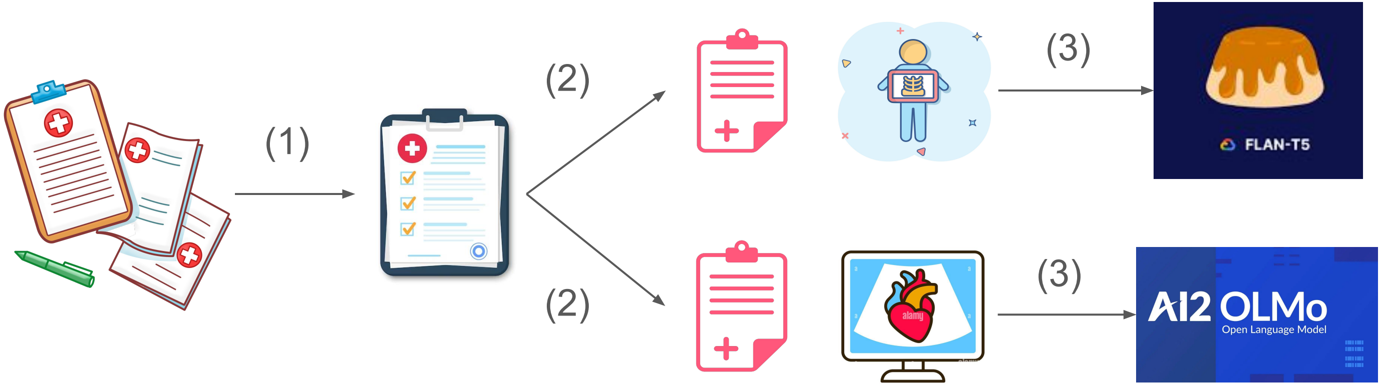
1. **Filter** text **T** to pre-treatment text $\mathbf{T}^{\mathbf{pre}}$

Our procedure for generating proxies for each patient



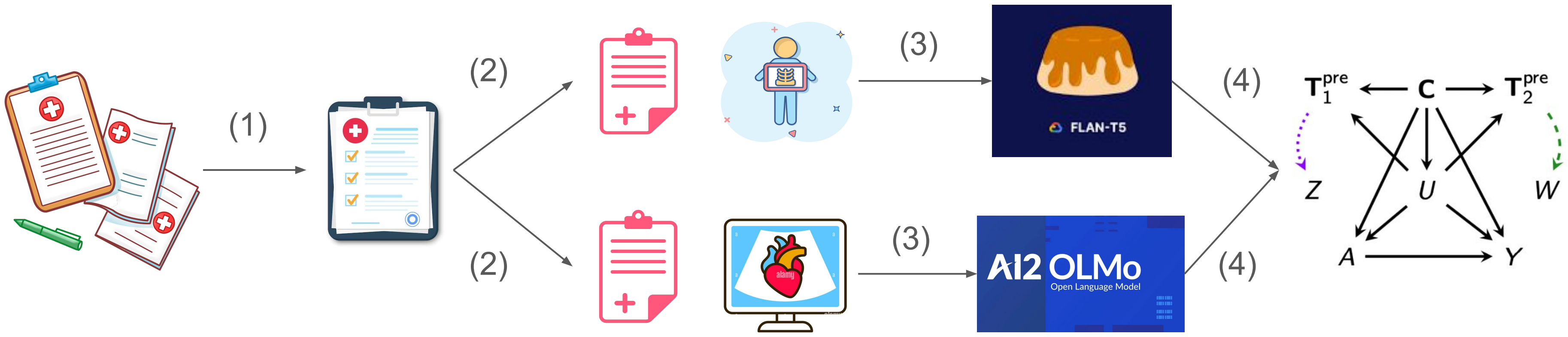
1. **Filter** text **T** to pre-treatment text $\mathbf{T}^{\mathbf{pre}}$
2. **Split** text $\mathbf{T}^{\mathbf{pre}}$ into independent instances $\mathbf{T}^{\mathbf{pre1}}$ and $\mathbf{T}^{\mathbf{pre2}}$ — e.g., radiology and echocardiogram notes

Our procedure for generating proxies for each patient



1. **Filter** text **T** to pre-treatment text $\mathbf{T^{pre}}$
2. **Split** text $\mathbf{T^{pre}}$ into independent instances $\mathbf{T^{pre1}}$ and $\mathbf{T^{pre2}}$ — e.g., radiology and echocardiogram notes
3. **Zero-shot prediction** for *Z* and *W* using two different LLMs e.g., Flan-T5 and OLMo

Our procedure for generating proxies for each patient



1. **Filter** text **T** to pre-treatment text $\mathbf{T}^{\mathbf{pre}}$
2. **Split** text $\mathbf{T}^{\mathbf{pre}}$ into independent instances $\mathbf{T}^{\mathbf{pre1}}$ and $\mathbf{T}^{\mathbf{pre2}}$ — e.g., radiology and echocardiogram notes
3. **Zero-shot prediction** for *Z* and *W* using two different LLMs e.g., Flan-T5 and OLMo
4. **Check** an *odds ratio* heuristic and **plug-in** to proximal g-formula if the proxies pass

Results semi-synthetic

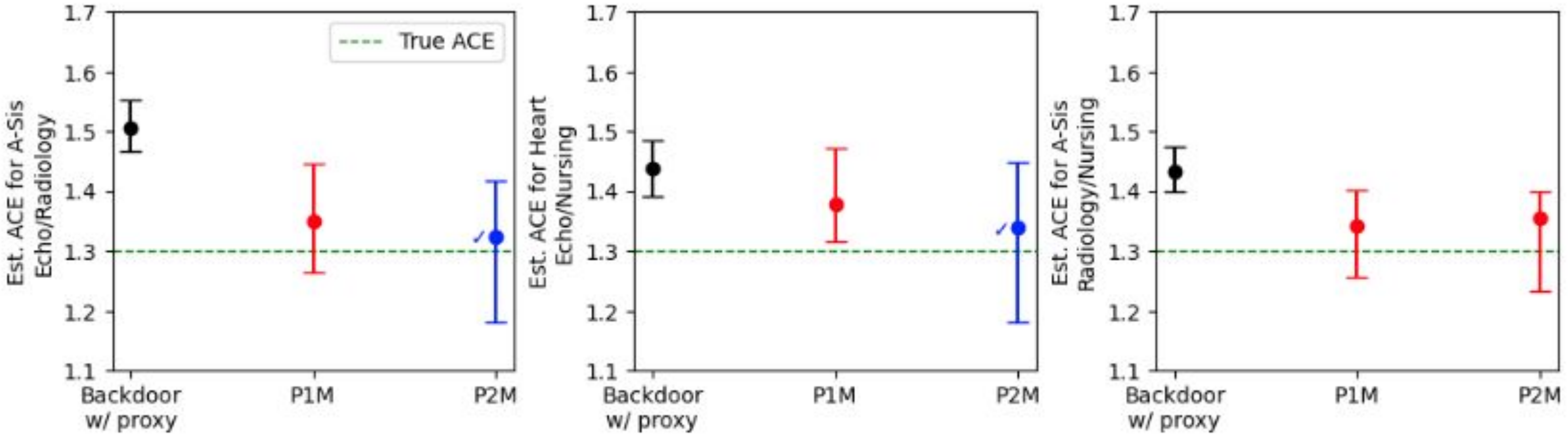
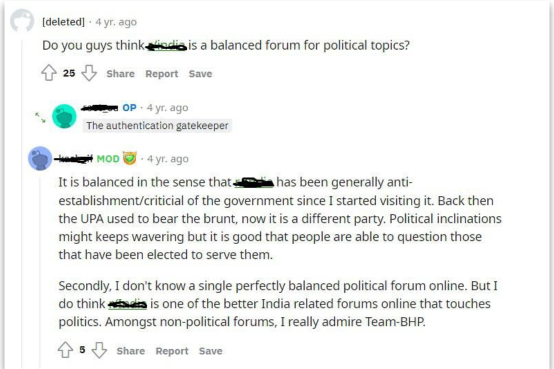


Figure 4: **Semi-synthetic results** for ACE point estimates (dots) and 95% CIs (bars). We distinguish settings that passed the odds ratio heuristic (✓) from those that failed, with $\gamma_{\text{high}} = 2$.

You can use our approach with other kinds of text



e.g., political speeches, reddit posts, etc.

Links and Contact Information

- Poster link: https://neurips.cc/virtual/2024/poster/95623
- Camera-ready paper: https://openreview.net/pdf?id=L4RwA0qyUd
- Code: https://github.com/jacobmchen/proximal_w_text

- Jacob M. Chen jchen459@jhu.edu
- Rohit Bhattacharya rb17@williams.edu
- Katherine A. Keith kak5@williams.edu