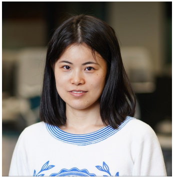
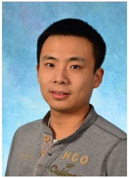
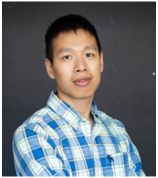
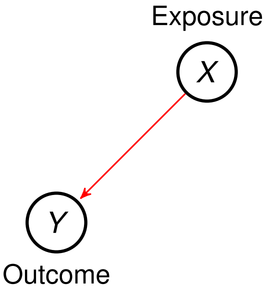
The Promises of Parallel Outcomes

Ying Zhou, Dingke Tang, Dehan Kong, **Linbo Wang**
University of Connecticut, University of Ottawa, and **University of Toronto**

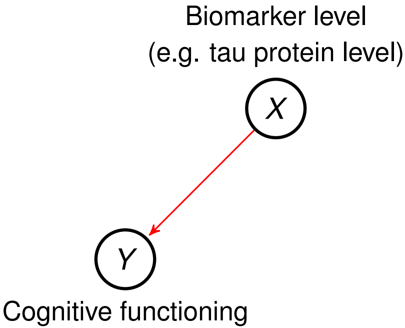
   

KDD 2025 Workshop – Causal Inference and Machine Learning in Practice
August 4, 2025

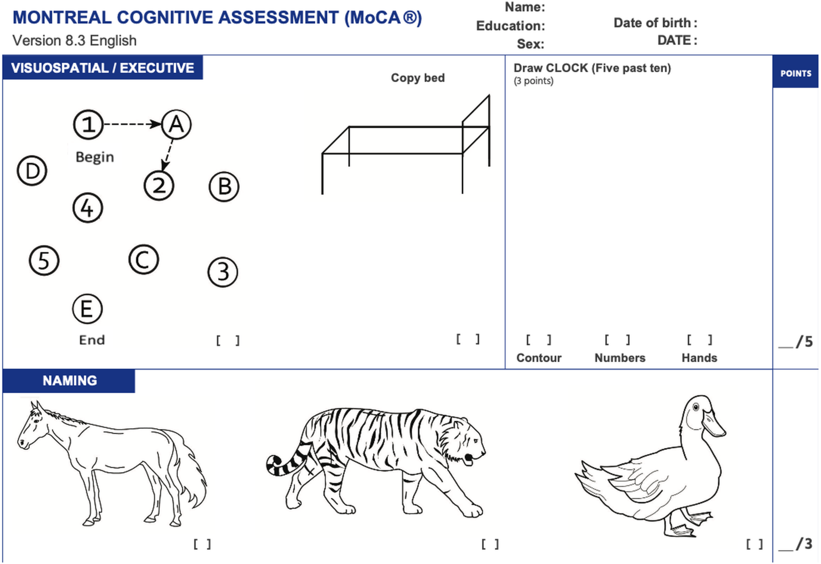
The Causal Inference Problem



The Causal Inference Problem



Cognitive test (e.g. MMSE)



MONTREAL COGNITIVE ASSESSMENT (MoCA®)
Version 8.3 English

Name:
Education:
Sex:
Date of birth :
DATE :

VISUOSPATIAL / EXECUTIVE

[ ]

Copy bed

[ ]

Draw CLOCK (Five past ten)
(3 points)

[ ] Contour
[ ] Numbers
[ ] Hands

POINTS

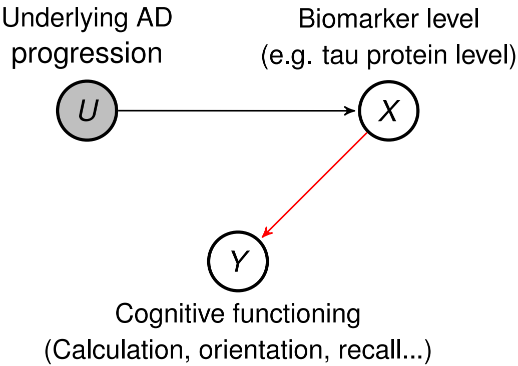
__/5

NAMING

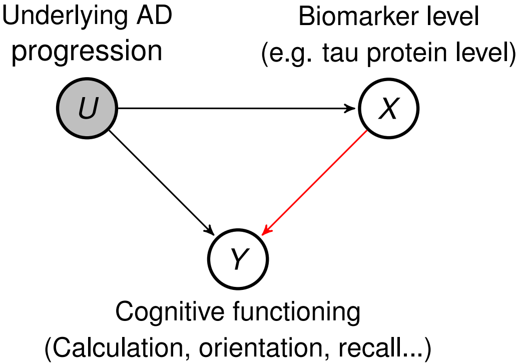
[ ]
[ ]
[ ]

__/3

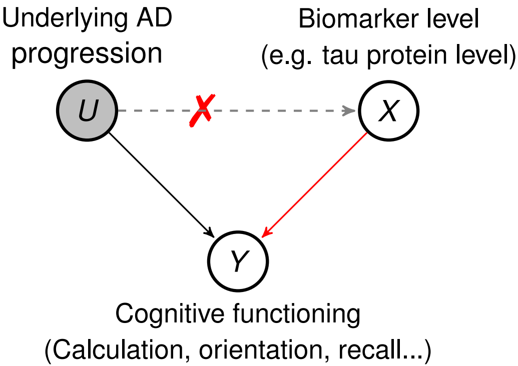
Causal Inference with Unmeasured Confounding



Causal Inference with Unmeasured Confounding



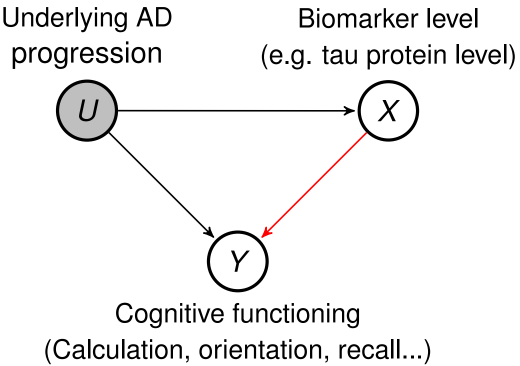
Causal Inference with Unmeasured Confounding



Classical solutions:

- Randomized experiment

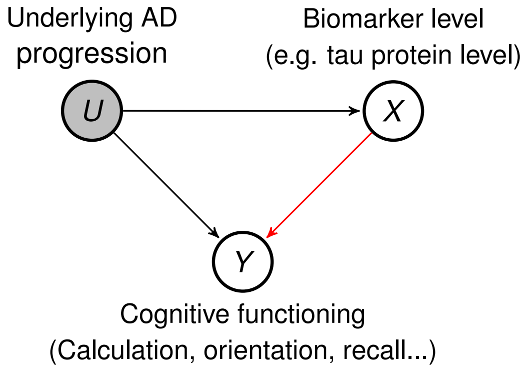
Causal Inference with Unmeasured Confounding



Classical solutions:

- Randomized experiment
- Measure all the confounders

Causal Inference with Unmeasured Confounding



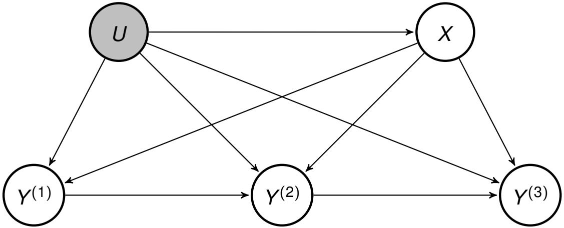
Classical solutions:

- Randomized experiment
- Measure all the confounders

Our solution:

- Leverage the information in a multivariate outcome

Set-up: Causal Inference with Multiple Outcomes



A multiple-outcomes set-up

If U were observed, then

$$E[\boldsymbol{Y}(x)] = E_U E[\boldsymbol{Y} \mid X = x, U]$$

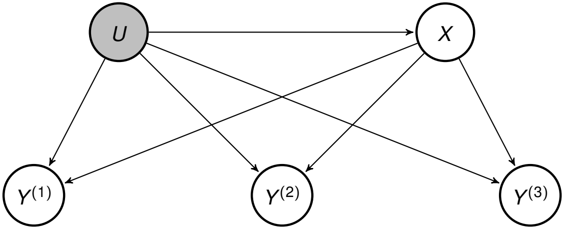
However, if U were not fully observed, then $E[\boldsymbol{Y}(x)]$ is not identifiable

Overview of the parallel-outcomes framework

Non-parametric identification

Parametric Modeling

Key Assumption: Parallel Outcomes

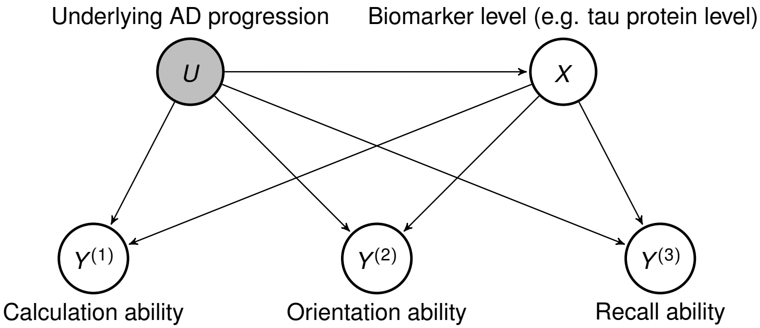


The simplest parallel-outcomes model

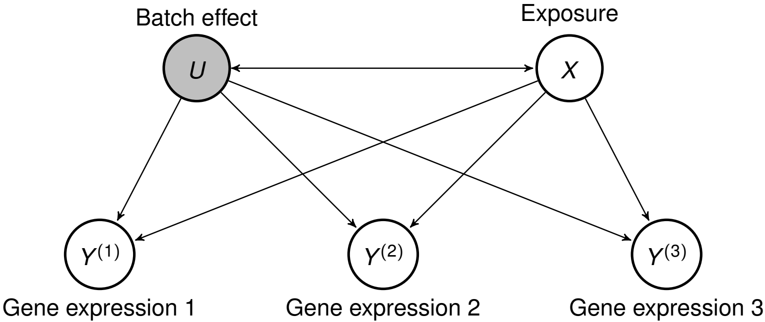
Parallel-outcomes Model:

$$Y^{(1)} \perp\!\!\!\perp Y^{(2)} \perp\!\!\!\perp \cdots \perp\!\!\!\perp Y^{(p)} \mid (U, X).$$

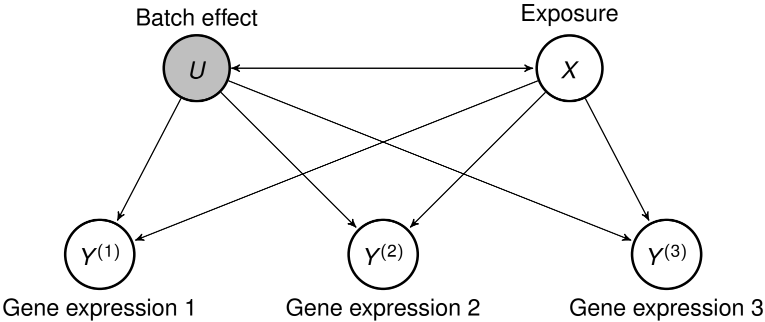
Example: Alzheimer's Disease



Another example: Gene expression analysis



Another example: Gene expression analysis



Often assumed in **Surrogate Variable Analysis** in genomics (Leek and Storey, 2008; Gagnon-Bartsch et al., 2013; Sun et al., 2012; Wang et al., 2017)

$$\boldsymbol{Y} = X\beta + U\gamma + \epsilon$$

- Independent noise

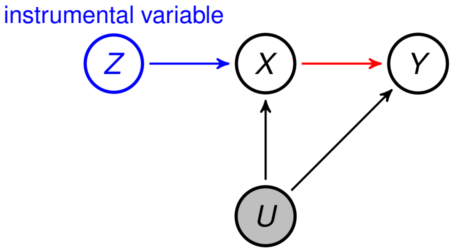
Main Results

1. Causal effects are **non-parametrically identifiable** with three parallel outcomes

2. Causal effects are identifiable with two parallel outcomes under linear structural equation models (with additional conditions)

Comparison to Auxiliary Variables Approaches

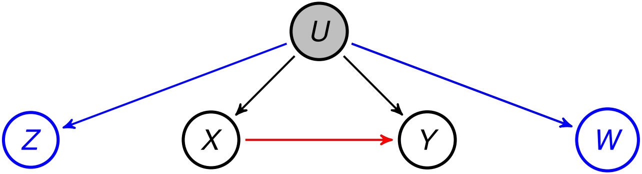
- Instrumental variable (IV) methods (Wright and Wright (1928); Goldberger (1972); Hernán and Robins (2006); LW and Tchetgen Tchetgen (2018))
  - Find an instrumental variable Z that acts like a randomization



Comparison to Auxiliary Variables Approaches

- Instrumental variable (IV) methods (Wright and Wright (1928); Goldberger (1972); Hernán and Robins (2006); LW and Tchetgen Tchetgen (2018))
  - Find an instrumental variable Z that acts like a randomization

- Negative controls (Miao et al., 2018)
  - Find a negative control exposure Z and a negative control outcome W



Comparison to Auxiliary Variables Approaches

- Instrumental variable (IV) methods (Wright and Wright (1928); Goldberger (1972); Hernán and Robins (2006); LW and Tchetgen Tchetgen (2018))
  - Find an instrumental variable Z that acts like a randomization

- Negative controls (Miao et al., 2018)
  - Find a negative control exposure Z and a negative control outcome W

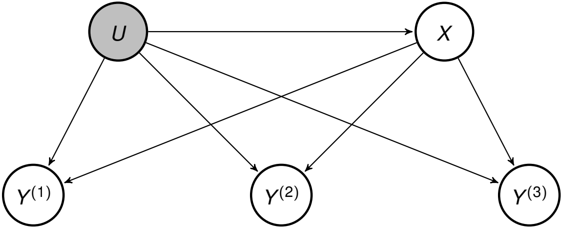
The parallel-outcomes framework requires no external data

Overview of the parallel-outcomes framework

Non-parametric identification

Parametric Modeling

Start from a binary model...



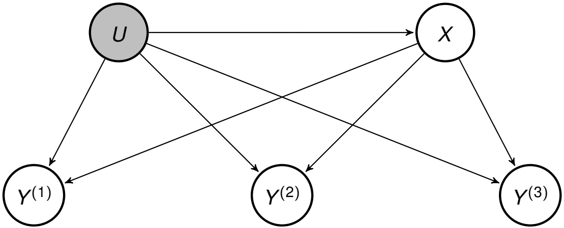
The simplest parallel-outcomes model

For each $x = 1, 2$,

$$\mathrm{pr}(y^{(1)}, y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(1)} \mid u, x)\mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

$2^3 - 1 = 7$ *eqns*   *2 parameters*   2   2   1

Start from a binary model...



The simplest parallel-outcomes model

For each $x = 1, 2,$

$$\mathrm{pr}(y^{(1)}, y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(1)} \mid u, x)\mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

$2^3 - 1 = 7$ *eqns*   *2 parameters*   2   2   1

Promising, but these are non-linear equations...

Toward identifiability

For any $y^{(1)}, y^{(2)}, y^{(3)}, x$, we have

$$\mathrm{pr}(y^{(1)}, y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(1)} \mid u, x)\mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

Toward identifiability

For any $y^{(1)}, y^{(2)}, y^{(3)}, x$, we have

$$\mathrm{pr}(y^{(1)}, y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(1)} \mid u, x)\mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

$$\mathrm{pr}(y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

Toward identifiability

For any $y^{(1)}, y^{(2)}, y^{(3)}, x$, we have

$$\mathrm{pr}(y^{(1)}, y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(1)} \mid u, x)\mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

$$\mathrm{pr}(y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

Idea: if only we could take the ratio between these two equations...

The matrix adjustment method (Rothman et al., 2008; Hu, 2008)

For any $y^{(1)}, y^{(2)}, y^{(3)}, x$, we have

$$\mathrm{pr}(y^{(1)}, y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(1)} \mid u, x)\mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

$$\mathrm{pr}(y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

Fix $y^{(1)}$, and **write summation in terms of matrix multiplication**

$$P(y^{(1)}, Y^{(2)}, Y^{(3)} \mid x) = P(Y^{(2)} \mid U, x)P_D(y^{(1)} \mid U, x)P_D(U \mid x)P(Y^{(3)} \mid U, x)^T$$

$$P(Y^{(2)}, Y^{(3)} \mid x) = P(Y^{(2)} \mid U, x)P_D(U \mid x)P(Y^{(3)} \mid U, x)^T$$

where $P_D(\cdot)$ are diagonal matrices.

The matrix adjustment method (Rothman et al., 2008; Hu, 2008)

For any $y^{(1)}, y^{(2)}, y^{(3)}, x$, we have

$$\mathrm{pr}(y^{(1)}, y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(1)} \mid u, x)\mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

$$\mathrm{pr}(y^{(2)}, y^{(3)} \mid x) = \sum_u \mathrm{pr}(y^{(2)} \mid u, x)\mathrm{pr}(y^{(3)} \mid u, x)\mathrm{pr}(u \mid x)$$

Fix $y^{(1)}$, and **write summation in terms of matrix multiplication**

$$P(y^{(1)}, Y^{(2)}, Y^{(3)} \mid x) = P(Y^{(2)} \mid U, x)P_D(y^{(1)} \mid U, x)P_D(U \mid x)P(Y^{(3)} \mid U, x)^T$$

$$P(Y^{(2)}, Y^{(3)} \mid x) = P(Y^{(2)} \mid U, x)P_D(U \mid x)P(Y^{(3)} \mid U, x)^T$$

where $P_D(\cdot)$ are diagonal matrices.

"Take the ratio":

$$P(y^{(1)}, Y^{(2)}, Y^{(3)} \mid x)P(Y^{(2)}, Y^{(3)} \mid x)^{-1} = P(Y^{(2)} \mid U, x)P_D(y^{(1)} \mid U, x)P(Y^{(2)} \mid U, x)^{-1}.$$

Identification

Theorem (Zhou, Tang, Kong and LW, 2024)

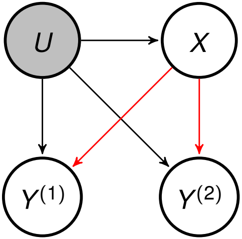
Under the parallel-outcomes model and some additional regularity conditions, for all x, the potential outcome distributions $pr(y^{(j)}(x)), j = 1, 2, 3$ are identifiable in a discrete parallel-outcome model in which all of $Y^{(1)}, Y^{(2)}, Y^{(3)}, U$ have k levels.

Overview of the parallel-outcomes framework

Non-parametric identification

Parametric Modeling

Linear structural equation modeling



Linear structural equation models:

$$X = \alpha_X U + \gamma_X^\top V + \epsilon_X,$$
$$Y^{(1)} = \alpha_1 U + \beta_1 X + \gamma_1^\top V + \epsilon_1,$$
$$Y^{(2)} = \alpha_2 U + \beta_2 X + \gamma_2^\top V + \epsilon_2.$$

Here V represents measured covariates

Theorem (Zhou, Tang, Kong & LW, 2024)

Assume a linear structural model and the following conditions:

Condition

X, $Y^{(1)}$, $Y^{(2)}$ have finite second moments.

Condition

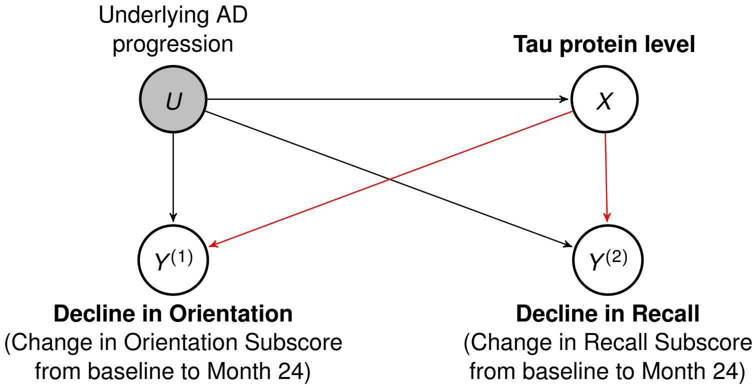
The distributions of ϵ_X, ϵ_1 and ϵ_2 are symmetric;

Condition

The distribution of U is asymmetric.

Then the causal effects β_1 and β_2 are identifiable.

Data Example



- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- 925 subjects with complete measurements
- measured covariates: age, gender and education length

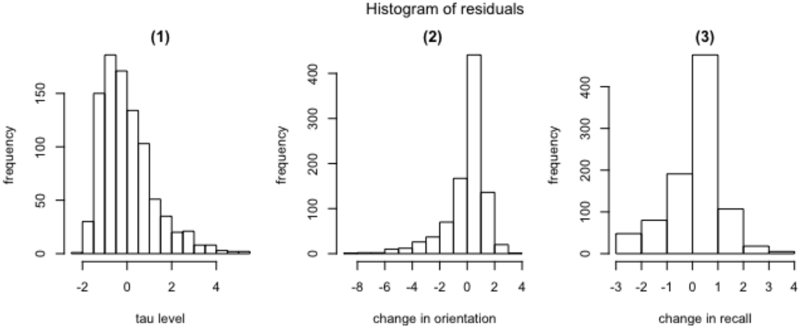
Testable implication

$$X = \alpha_X U + \gamma_X^\top V + \epsilon_X,$$
$$Y^{(1)} = \alpha_1 U + \beta_1 X + \gamma_1^\top V + \epsilon_1,$$
$$Y^{(2)} = \alpha_2 U + \beta_2 X + \gamma_2^\top V + \epsilon_2.$$

Conditions:

- $\epsilon_X, \epsilon_1, \epsilon_2$ are symmetric
- U is asymmetric

Implies that the residuals of X, $Y^{(1)}$, $Y^{(2)}$ are asymmetric



(A)symmetry check of residual distributions.

Estimation Results

Subscore category	Crude	Adjusted
Orientation	-0.35	-0.30
	(95% CI: [-0.43, -0.27])	(95% CI: [-0.42, -0.21])
Recall	-0.11	-0.08
	(95% CI: [-0.17,-0.06])	(95% CI: [-0.15, 0.004])

Unit: Change in subscore per 100 pg/mL

- A higher tau level may lead to acceleration in cognitive decline in both orientation and recall abilities
- Effect attenuated compared to crude estimates

Parallel outcomes: Summary

- The key challenge to causal inference from observation studies is unmeasured confounding
- The parallel-outcomes framework provides a solution
- Leverage condition independence structure among multiple parallel outcomes
- Promising for analyzing high-dimensional response data
  - Can use the extra outcomes to relax the conditional independence assumption

Thank you!

Paper: Zhou Y., Tang D., Kong D., and Wang L.. Promises of Parallel Outcomes. *Biometrika*, 111.2 (2024): 537-550.

Slides: Available on my personal website https://sites.google.com/site/linbowangpku/papers

Contact: Linbo Wang (University of Toronto), linbo.wang@utoronto.ca

References I

Gagnon-Bartsch, J. A., Jacob, L., and Speed, T. P. (2013). Removing unwanted variation from high dimensional data with negative controls. *Berkeley: Tech Reports from Dep Stat Univ California*, pages 1–112.

Goldberger, A. S. (1972). Structural equation methods in the social sciences. *Econometrica*, 40(6):979–1001.

Hernán, M. A. and Robins, J. M. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology*, 17(4):360–372.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, 144(1):27–61.

Leek, J. T. and Storey, J. D. (2008). A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 105(48):18718–18723.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008). *Modern Epidemiology*. Lippincott Williams & Wilkins.

Sun, Y., Zhang, N. R., Owen, A. B., et al. (2012). Multiple hypothesis testing adjusted for latent variables, with an application to the agemap gene expression data. *The Annals of Applied Statistics*, 6(4):1664–1688.

References II

Wang, J., Zhao, Q., Hastie, T., and Owen, A. B. (2017). Confounder adjustment in multiple hypothesis testing. *Annals of statistics*, 45(5):1863–1894.

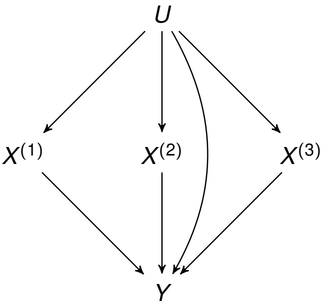
Wang, Y. and Blei, D. M. (2019). The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528):1574–1596.

Wright, P. G. and Wright, S. (1928). *The tariff on animal and vegetable oils*. Macmillan.

Comparison to Multi-cause causal inference (Wang and Blei, 2019)

- Multiple exposures, One outcome
- Shared confounding among exposures

$$X^{(1)} \perp\!\!\!\perp X^{(2)} \perp\!\!\!\perp \ldots X^{(p)} \mid U$$



Comparison to Multi-cause causal inference (Wang and Blei, 2019)

- Multiple exposures, One outcome
- Shared confounding among exposures

$$X^{(1)} \perp\!\!\!\perp X^{(2)} \perp\!\!\!\perp \dots X^{(p)} \mid U$$

Causal effects are identifiable under additional parametric assumptions (Kong, Yang & LW, 2019)

- Linear treatment effect model
- An additional parametric binary choice model model for the outcome Y

Cox and Donnelly (2011, p.96). *Principles of Applied Statistics*:

> *If an issue can be addressed nonparametrically then it will often be better to tackle it parametrically; however, if it cannot be resolved nonparametrically then it is usually dangerous to resolve it parametrically.*