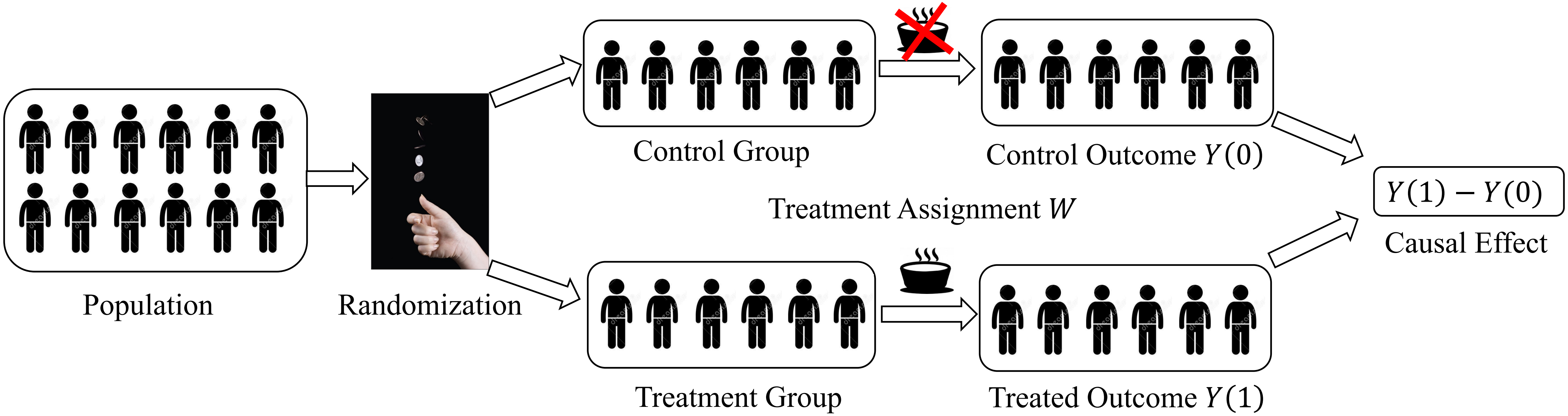
Learning Causal Effect for High-dimensional Observation Data with Unmeasured Confounding

Hedong YAN, 21481776
Computer Science Department
Hong Kong Baptist University
Supervisor: Yiu-ming Cheung

- Background
- Application
- Problem and Challenge
- Related Work
- Our Preliminary work
- Our Approach to Address Challenges
- Futural Plan

Background

- Causal Effect and Randomized Control Trail
- "Unknown potential yields" of Neyman's agriculture experiment

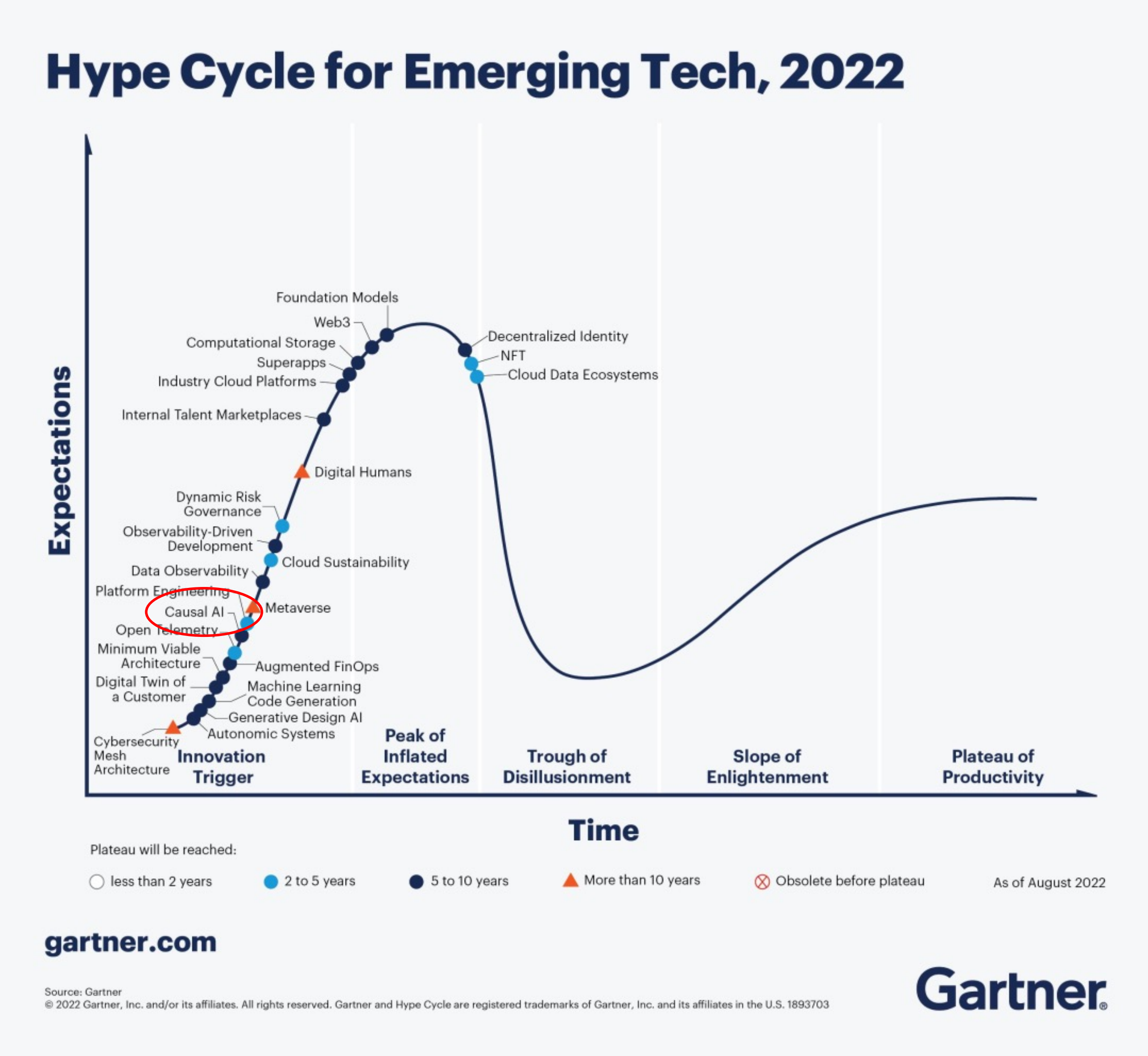


Background

- Limitation and Opportunity
  - RCT can only give a ***population-level*** conclusion.
  - RCT can not be performed due to ***immorality*** and ***high cost***.
  - Observational ***data*** upsurges.
- As an alternative, learning causal effects from the observational dataset is not totally impossible.

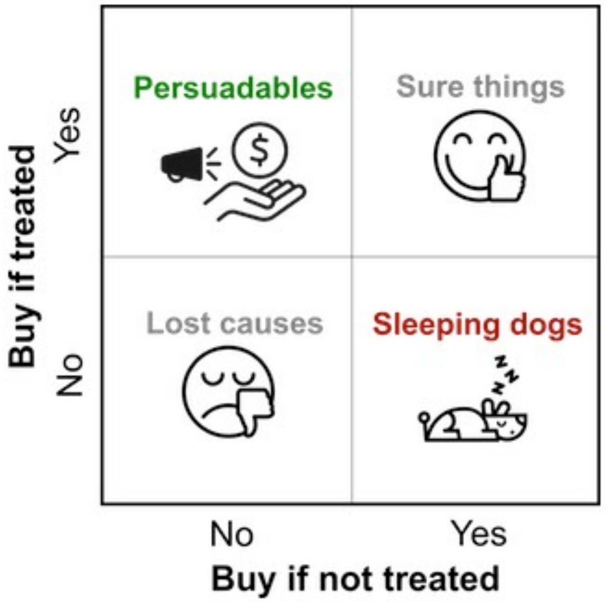
Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945-960.

Background

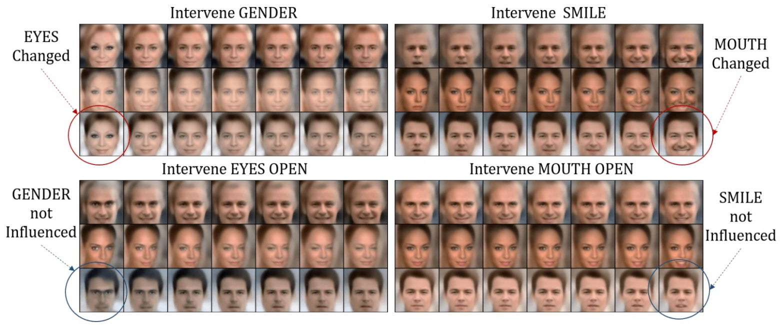


- Causal AI is still in the innovation trigger stage.

Application



Individuals	$E(Y_i(1))$	$E(Y_i(0))$	Recommendation
u_1	Good	Good	No
u_2	Bad	Bad	No
u_3	Good	Bad	Yes
u_4	Bad	Good	No
u_5	Good	Good	No



Uplift Marketing

Individual Drug Recommendation

Causal Feature

https://www.ambiata.com/blog/2020-07-07-uplift-modeling/

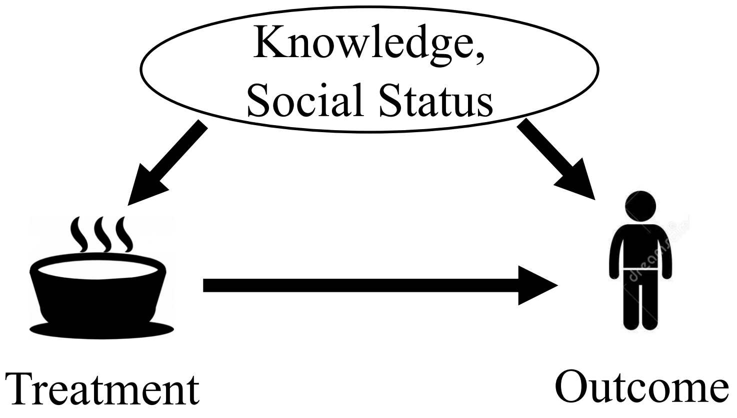
Yang, Mengyue, et al. "Causalvae: Disentangled representation learning via neural structural causal models." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

Problem and Challenge

- Open Problem: Learning causal effects from observational data
- Challenge:
  - Hidden confounding
  - High dimensionality
  - Robustness

Challenge 1: Hidden confounding

- Treatment assignment is unknown and not randomized in observation data. We can not identify causal effects from observational data.



Challenge 2: High dimensionality

- Potential dependency is exponential.
  - For example, the number of acyclic-directed mixed graphs is $O\left(2^{n^2-n} * n! * 1.3^{n^2}\right)$ where n is the number of variables.

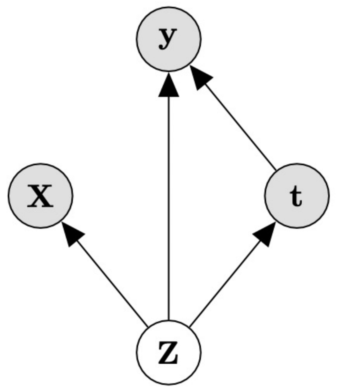
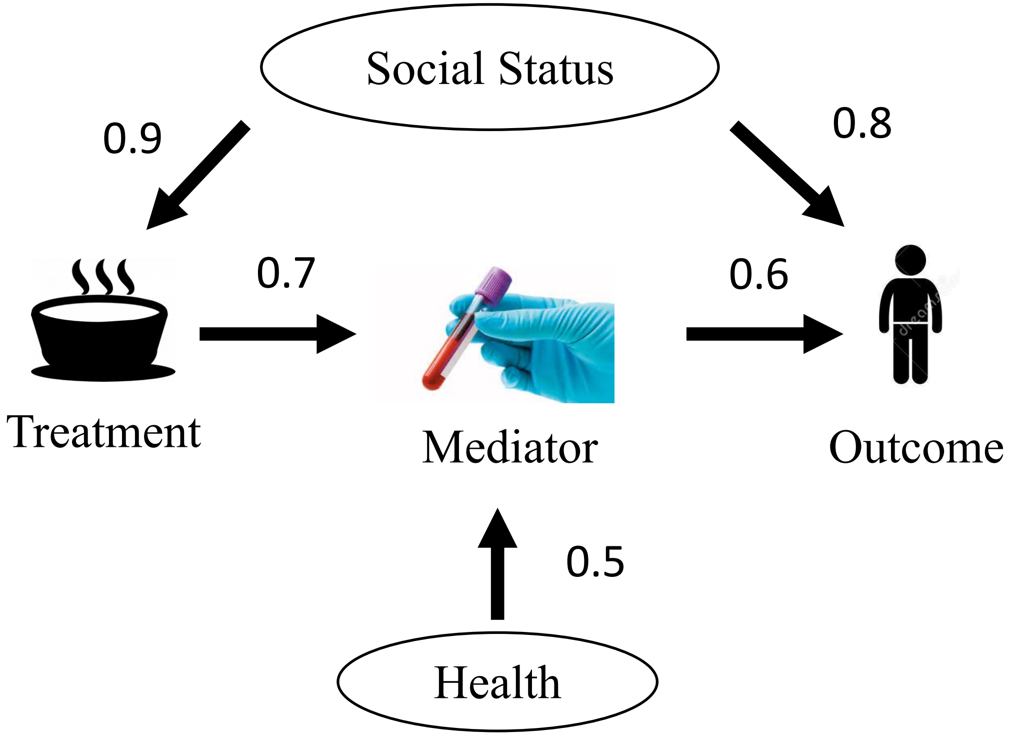


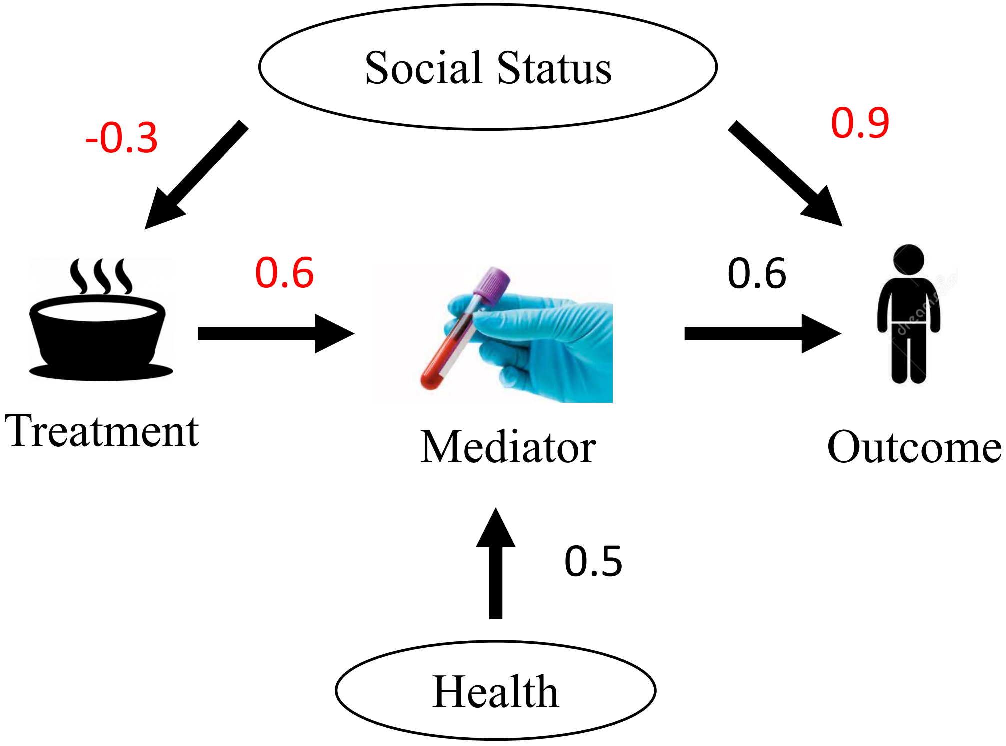
Figure 1: Example of a proxy variable. **t** is a treatment, e.g. medication; **y** is an outcome, e.g. mortality. **Z** is an unobserved confounder, e.g. socio-economic status; and **X** is noisy views on the hidden confounder **Z**, say income in the last year and place of residence.

Challenge 3: Robustness

- What if the dependency relationship (structure and parameters) changed?



$M_1 : E(Y(W)) = E_{X|w=0}(E_W(E(Y|W,X))$



M_1 still work well?

Related works: Algorithm

Four Components

Counterfactual Imputation: impute the influence, such as loss value, of counterfactual data on our model.
Balancing Regularization: treatment group and control group are sampled from the same distribution.
Potential Outcome Prediction: learning potential outcomes prediction function for causal effect estimation.
Estimand Modeling: learning a function for the specific causal quantity that we want.

Table 1: Algorithms of causal effect learning from observation data. BLR/BNN: Shalit et al. (2017);TARNet/CFR-MMD/CFR-Wasserstein: Johansson et al. (2016);Dargonet: Shi et al. (2019);X-learner: Künzel et al. (2019);CEVAE: Louizos et al. (2017);Deconfounder: Wang & Blei (2019);GANITE: Yoon et al. (2018);SITE: Yao et al. (2018);DRNets: Schwab et al. (2020);VCNets: Nie et al. (2021).

Algorithms	Learning Stage	Counterfactual Imputation	Balancing Regularization	Potential Outcome Prediction	Estimand Modeling	Hidden Confounding
BLR BNN	Two-stage	Nearest Neighbor	Moment's Difference	Linear Neural Network	None	None
TARNet CFR-MMD CFR-Wasserstein Dargonnet	End-to-end	Perfect Counterfactual	None MMD Wasserstein CrossEntropy	Twin Neural Networks	None	None
X-Learner	Three-stage	Perfect Counterfactual	None	Twin BARTs	Yes	None
CEVAE	End-to-End	Perfect Counterfactual	Bayesian Variational Inference Network	Model Network	None	Proxy variables
Deconfounder	Two-stage	Perfect Counterfactual	Posterior Predictive Check of Factor Model	Linear	None	Proxy variables
GANITE	Two-stage	Counterfactual GAN	None	ITE GAN	None	None
SITE	End-to-end	PDDM Similarity	Middle Point Distance	Neural Network	None	None
DRNets VCNets	End-to-end	Nearest Neighbor	None	Treatment-Dose Networks Varying Coefficient Network	None	None

Related works: Benchmark

Table 3: Causal Dataset. Causeme: 202; JustCause: Hawkins & Kim (2021); e-CARE: Du et al. (2022); IHDP: Hill (2011); News: Johansson et al. (2016); Twins: Louizos et al. (2017); Jobs: Shalit et al. (2017); Movies: Wang & Blei (2019); GWAS: Song et al. (2015).

Type	Name	Introduction	Website
Benchmark	Causeme	time-series	https://causeme.uv.es/
Benchmark	JustCause	support IHDP, ACIC etc.	https://justcause.readthedocs.io/en/latest/
Benchmark	e-CARE	reasoning and explanation for NLP	https://scir-sp.github.io
Dataset	IHDP	home visits and IQ testing	https://github.com/vdorie/npci
Dataset	News	New York Times corpus	https://archive.ics.uci.edu/ml/datasets/Bag+of+Words
Dataset	Twins	birth weight and mortality	http://www.nber.org/data/linked-birth-infant-death-data-vital-statistics-data.html
Dataset	Jobs	labor earnings	https://users.nber.org/ rdehejia/data/.nswdata3.html
Dataset	Movies	Movie income and stars	https://www.kaggle.com/tmdb
Dataset	GWAS	genome-wide association studies	https://github.com/StoreyLab/gcatest
Competition	ACIC 2022	conference challenge	https://acic2022.mathematica.org/data
Competition	PCIC 2022	conference challenge	https://pattern.swarma.org/pcic/competition.html

Related works: Dimensionality Reduction

Table 2: Dimensionality reduction assumptions. G: Gaussian; I: independent; nG: non-Gaussian; $\perp$: orthogonal; $\rightarrow$: generate; ANN: additive normal noise; DAG: directed acyclic graph.

Method	Mapping	$p(\mathbf{z})$	$p(\mathbf{x})$
PCA	Linear	IG	IG
ICA	Linear	InG	InG+G
t-SNE	Nonlinear	Local continuty	Local continuty
βVAE	Nonlinear	IG with β	\
NGCA	Linear	G$\perp$nG	ANN
LinGAM	Linear	G$\rightarrow$nG	ANN with DAG

Related works: Toolbox

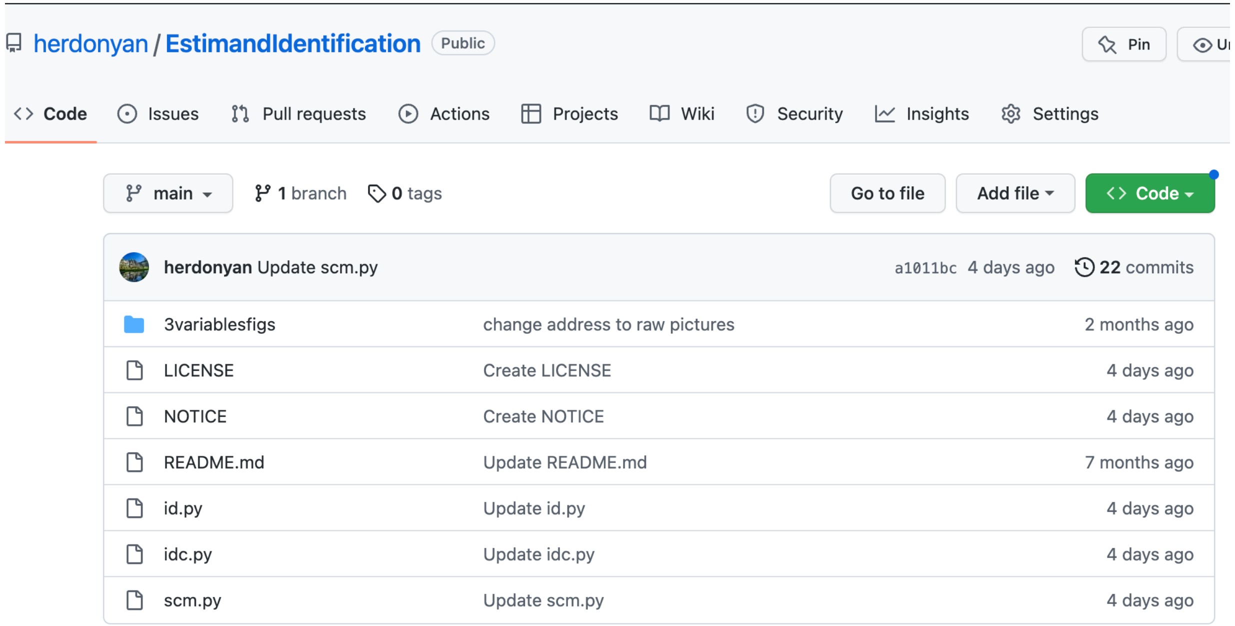
Table 4: Causal Packages. Tetrad: Ramsey et al. (2018); CausalDiscoveryToolbox: Kalainathan & Goudet (2019); Ananke: Nabi et al. (2020), Lee & Shpitser (2020), Bhattacharya et al. (2020); EconML: Keith Battocchi (2019); dowhy: Sharma et al. (2019); causalml: Chen et al. (2020); Causal-Curve: Kobrosly (2020); grf: Athey et al. (2019); dosearch: Tikka et al. (2021); causaleffect: Tikka & Karvanen (2017); dagitty: Textor et al. (2016).

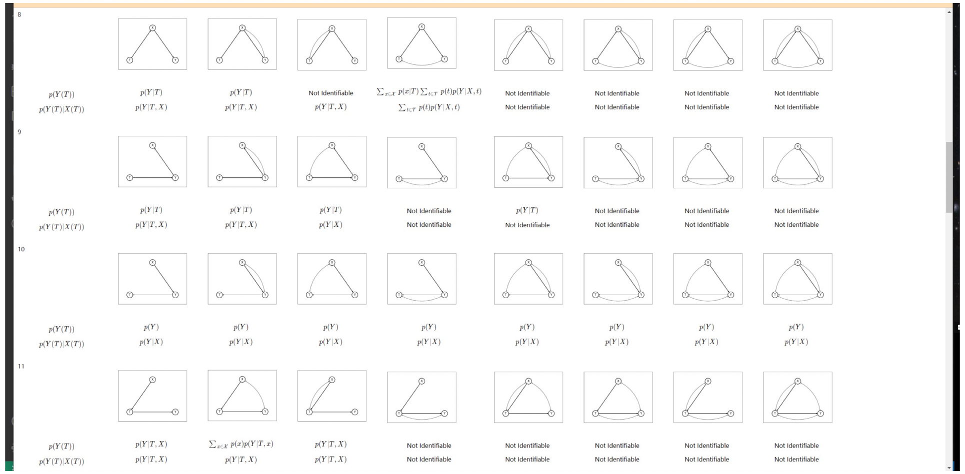
Motivation	Toolbox	Support Team	Introduction
Causal Learning	causal-learn	CMU, DMIR, Gong Mingming team, Shouhei Shimizu team	python version of Tetrad
	Tetrad	CMU	Java
	CausalDiscoveryToolbox	FenTechSolutions	python, DAG/Pair, dataset, independence, structure learning, metrics
	gCastle	Huawei Noah	python, data generation and process, causal structure learning, metrics
	tigramite	Jakob Runge	python, learning from time-series data
Causal Reasoning	Ananke	Ilya Shpitser team	python, support do-calculus
	EconML	Microsoft	python, Econometrics
	dowhy	Microsoft	python
	causalml	Uber	python, campaign target optimization, personalized engagement
	CausalImpact	Google	R, time-series, adertisement and click
	WhyNot	John Miller	python, simulator and environment
	Causal-Curve	Kobrosly, R.W.	python, continuous variable such as price, time and income
	grf	grf-lab of Standford	R
	dosearch	Santtu Tikka	R
	causaleffect	Santtu Tikka	R
	dagitty	\	R, support adjustment formula
End-to-End	causalnex	QuantumBlack	python, 0.11.0, structure learning, domain knowledge, estimation

Our Preliminary works

- Open Package: Identification and Structural Causal Model
- A Rejected Paper (UAI 2022): OOD Robustness

Our Preliminary works: Open Package





- Characteristics
  - ✓Automatic Identification
  - ✓Sampling data from given SCM with parameters

Our Preliminary works: Robustness

- Novelty: introduce auto identification into causal effect estimation.

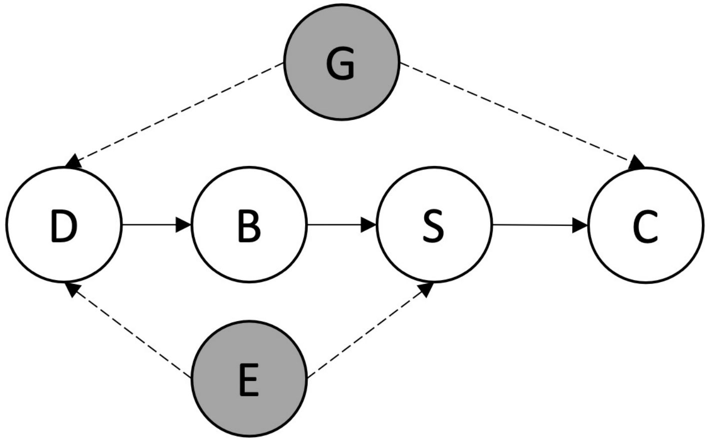
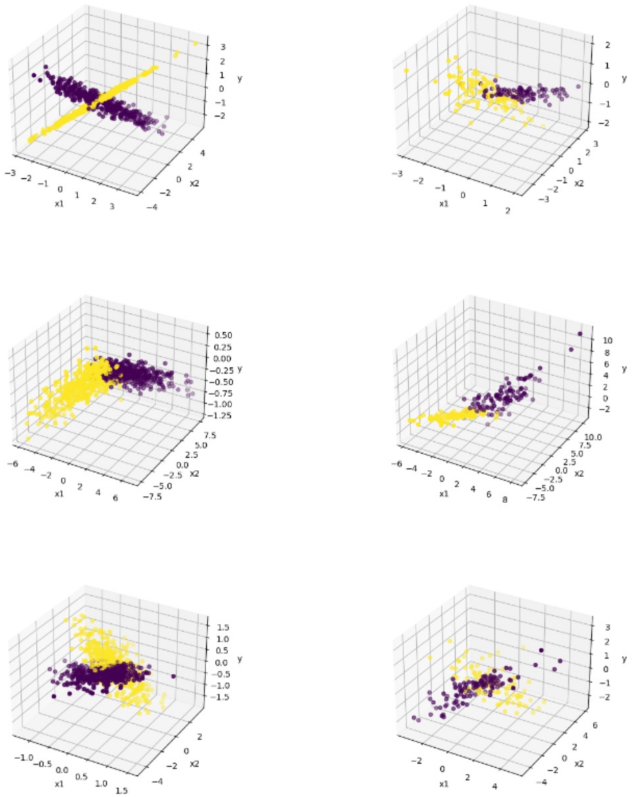


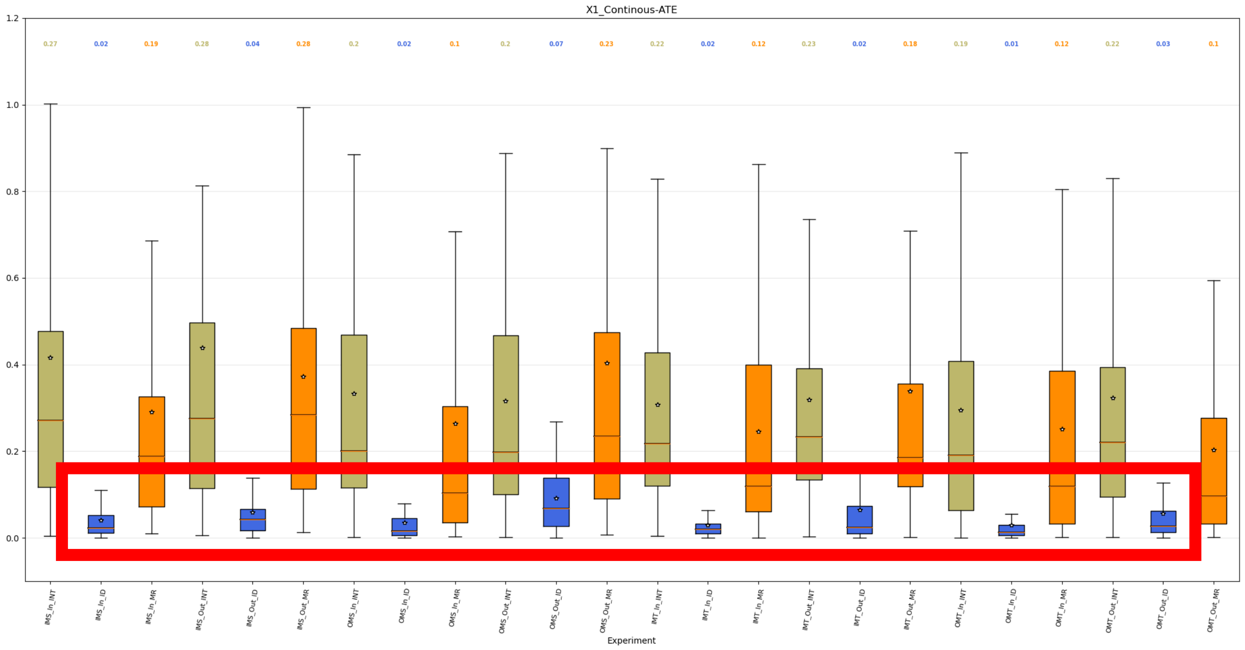
Figure 1: Example of four variables. D means dopamine; B means senior brain activity (frontal lobe); G means unobserved gene/physique; E means social environment not easy to measure. S means smoking behaviour, and C means cancer. For example, $E \to D$ may represent some life pressures, and $E \to S$ may be unconscious mimic nature.

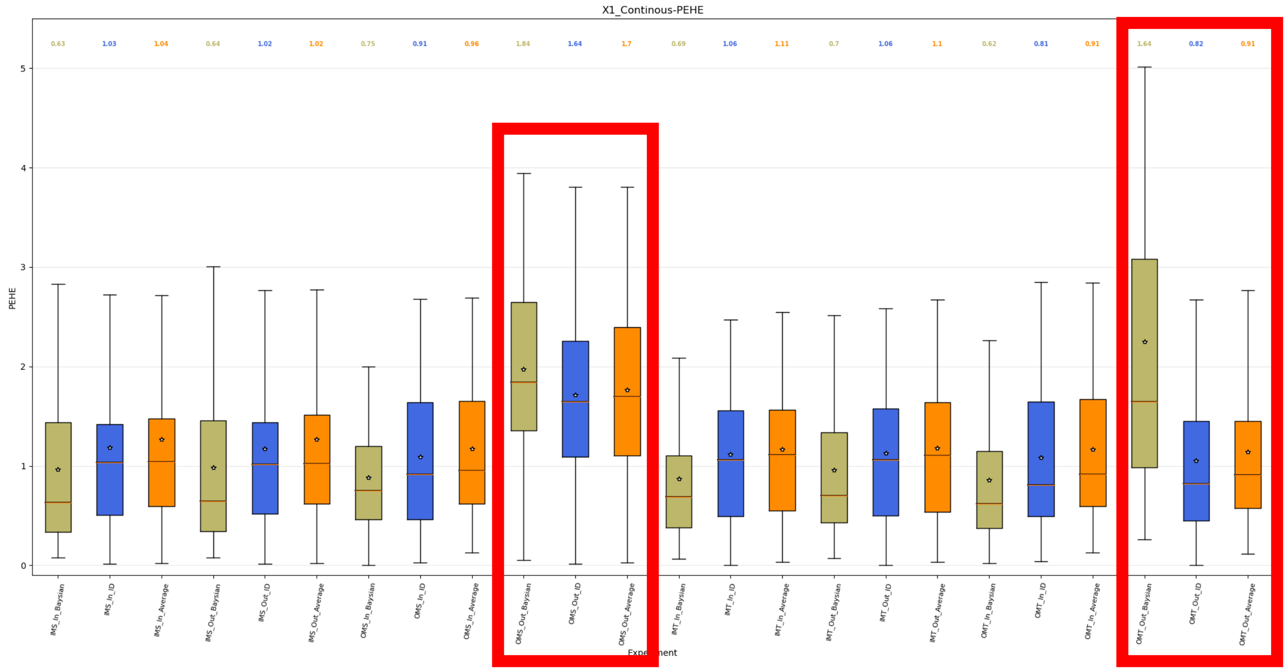


- The left column is train data, and the right column is test data. Yellow and purple indicate smoking or not.
- X_1 and X_2 are variable D and B. Y is variable C.

- In our simulation, we want to calculate the causal effect of smoking on cancer.
- We use $p(c|do(s)) = \frac{\sum_D p(D)p(s,c|D,b)}{\sum_D p(D)p(s|D,b)}$ and maximum likelihood to estimate $E(c|do(s))$ for all individuals.

Our Preliminary works: Robustness

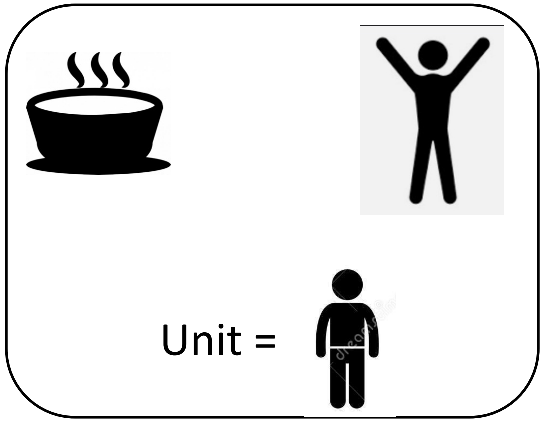


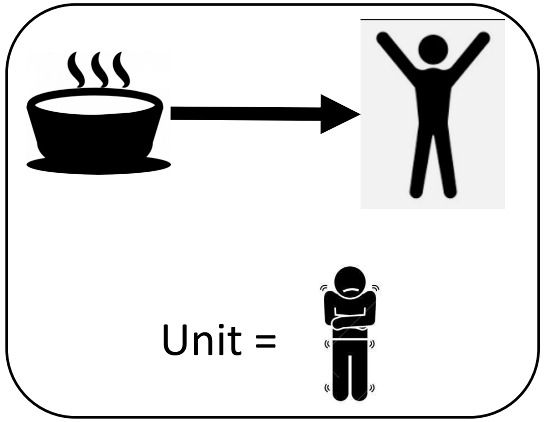


- In unbiasedness testing, estimations after identification are more unbiased than MR Freedman (2008) and INT Lin (2013) from ATE estimation results in both discrete and continuous cases. Considering estimation variance, it got better performance when outer mechanisms (dashed line) are changed.

Our Approach to Address Challenges

- Hidden Confounding: Individual Diagram
- Novelty: Will be the **first** to learn **Individual** Structural Causal Model for causal effect estimation.





Diagrams often assume non-parametric dependency among variables for all units.

But different units may have different dependencies and parameters.

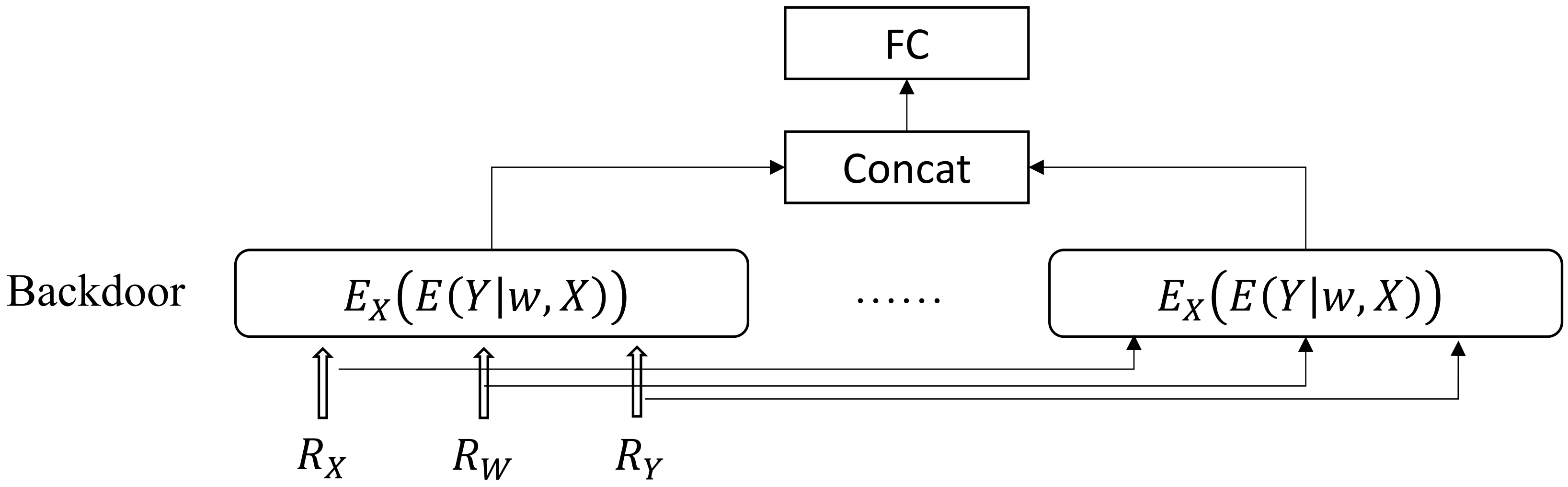
Our Approach to Address Challenges

- High Dimensionality:
  - Balancing Representation
  - Variables Grouping (the dependencies between variables within each group can be learned separately)
  - Multi Diagram Identification
  - Model Learning

- Novelty: The dependency among variables will be simplified by representation learning and variable grouping while preserving causal effect estimation performance.

Our Approach to Address Challenges

- Robustness:
  - Use multi-head techniques in the diagram identification stage to improve robustness for causal effect learning
- Novelty: Will be the first to introduce multi-head identification modules in causal effect learning.



Futural plan

Table 5: Ph.D. Program Timeline with Publication Goals

Year	Activities
1	Coursework, literature review, research proposal
2	Data collection, preliminary analysis, conference paper 1, conference presentation, QE
3	Advanced analysis, paper writing, conference paper 2, journal paper, Candidature
4	Finalize dissertation, defend dissertation, conference paper 3, graduation

Thanks!