April 26, 2021

COSC 5931/4931: Machine Learning Principles

Introduction to Causal Inference

Instructor: Dr. Sheikh Iqbal Ahamed
Presenter: Riddhiman Adib

Overview

- ➔ A Machine Learning Problem (or a causal problem ??)
- ➔ Before Causal Inference
  - ◆ Confounding Variables
  - ◆ Causal Models (vs. Machine Learning Models)
  - ◆ Intervention (vs. association)
- ➔ Intro to Causal Inference
  - ◆ do-Intervention
  - ◆ Biases through Causal Models
  - ◆ Causal Structure Learning Algorithms
  - ◆ Ladder of Causation
- ➔ Open Research Problems
- ➔ My Research Area

Correlation vs Causation

- ➔ Data show that as the number of fires increase, so does the number of firefighters. Therefore, to cut down on fires, we should reduce the number of firefighters.

- ➔ Data show that people who hurry tend to be late to their meetings. So don’t hurry, or you’ll be late.

- ➔ Humans tend to associate correlation with causation.
- ➔ Correlation does not imply Causation. (Wiki)

- ➔ How can we get causes of an event from a dataset?

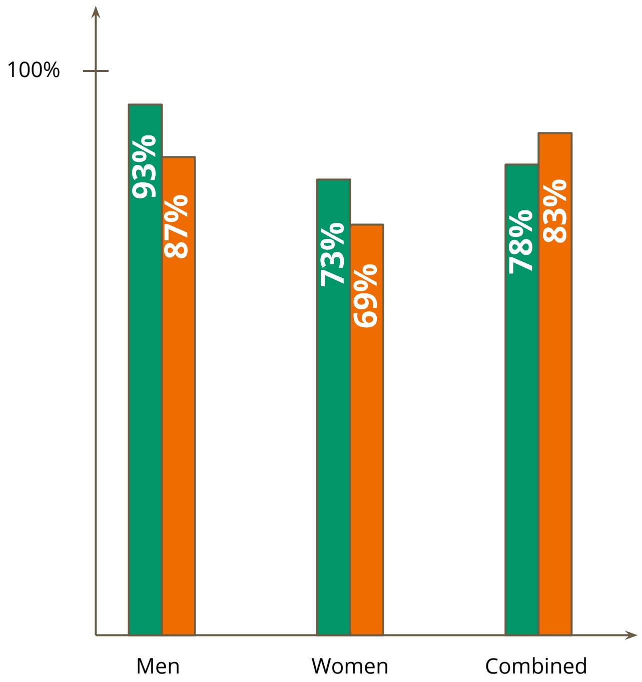
Question for all

- New drug study
  - Trial on Men and Women → Experimental Dataset

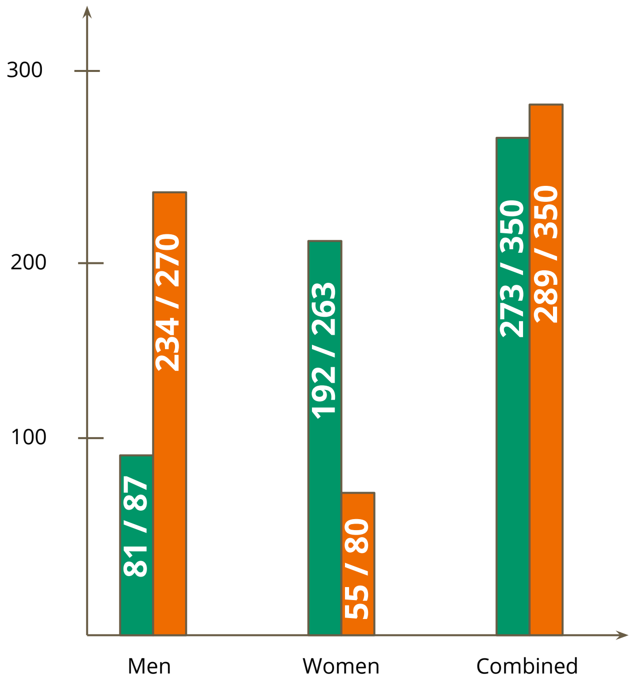
	Drug	No drug
Men	81 out of 87 (93%)	234 out of 270 (87%)
Women	192 out of 263 (73%)	55 out of 80 (69%)
Combined data	273 out of 350 (78%)	289 out of 350 (83%)

- Is the drug good (effective) or bad?
  - Given we DO NOT know the gender of the person,Should we prescribe the drug?
  - Given we know the gender of the person, should we prescribe the drug?

Drug Study Data Visualized



Percentage



Count

Simpson's Paradox (Wiki)

- ➔ If it is good for individual groups, it is good for all.

	Drug	No drug
Men	81 out of 87 (93%)	234 out of 270 (87%)
Women	192 out of 263 (73%)	55 out of 80 (69%)
Combined data	273 out of 350 (78%)	289 out of 350 (83%)

- ➔ UC Berkeley Gender Bias Case in 1973 (Wiki)
- ➔ Kidney Stone Treatment in 1986

Confounding (Lurking) Variable

- A confounding variable (also lurking variable) is a variable that
  - **influences** both the dependent variable and independent variable
  - causing a “fake” association.
- In Kidney Stone example, the confounding variable is the severity of the case
  - represented by the doctors' treatment decision trend of favoring B for less severe cases
- This was not previously known to be important until its effects were included.

Causal Models

- ➔ Causal models represent the mechanism by which data were generated.
- ➔ In Healthcare data, the features (bp, heart rate, temperature) are not mutually independent, rather dependent.
- ➔ We need Causal Model to understand their effects.
- ➔ Causal model in needed for changing the expected outcome.

Causal Models through Graphs

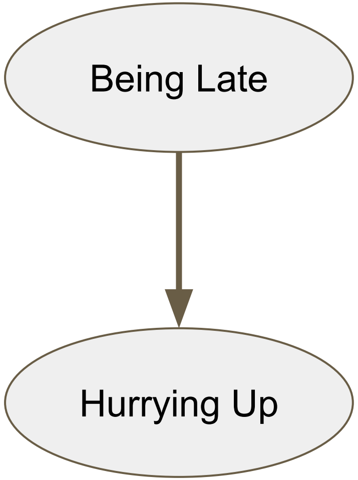


Fig: 1 (a)

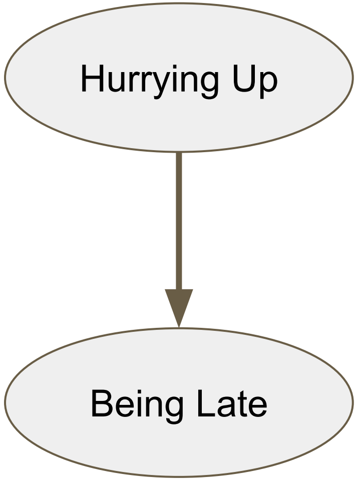


Fig: 1 (b)

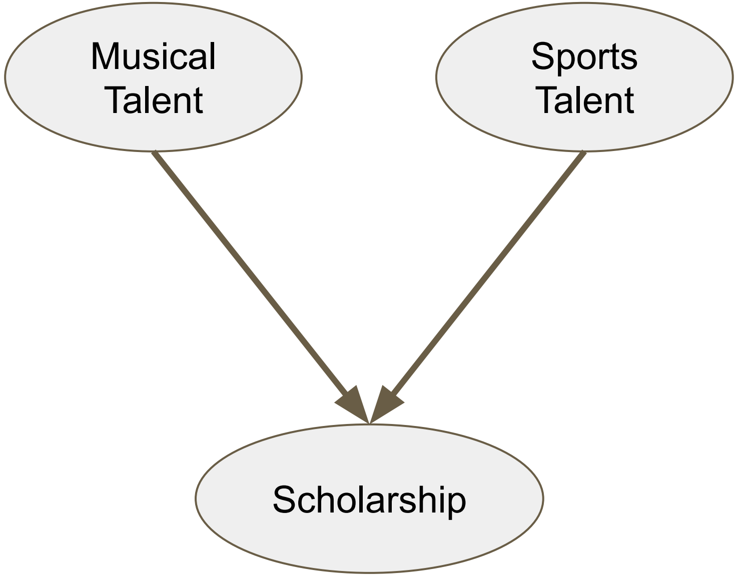


Fig: 2

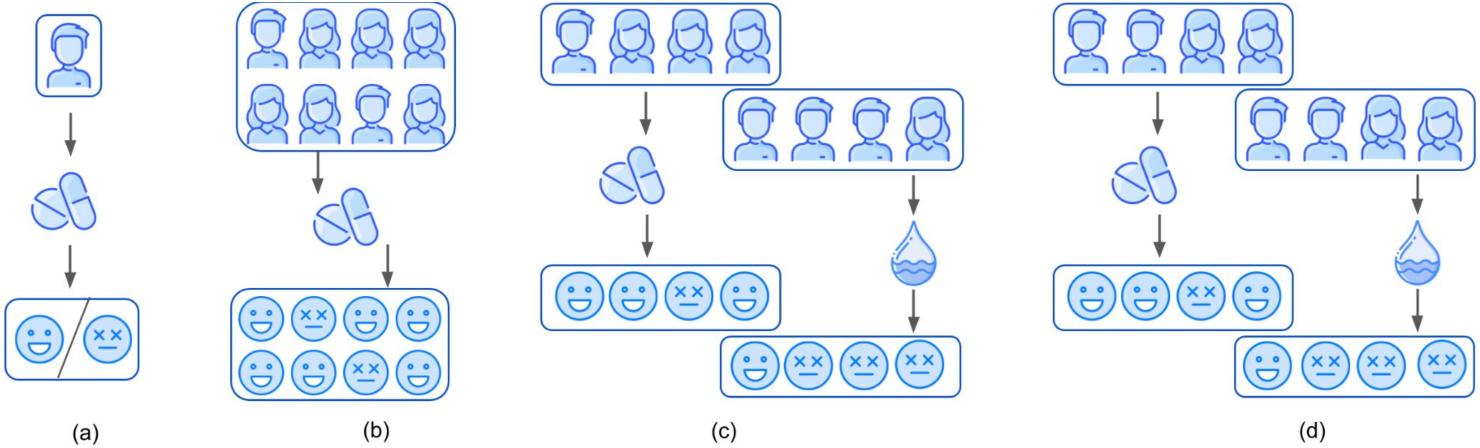
Causal Models vs Machine Learning Models

- ➔ Prediction based on correlation vs. intervention based on causality
  - Search for causality is the goal of science
- ➔ Data alone cannot hold causality
  - need additional assumptions and knowledge
- ➔ No true causal model
  - Granular models vs broad models
- ➔ Association → intervention → Counterfactuals
  - "A happens <u>when</u> B happens" → "A happens <u>because</u> B happens" → "<u>What would happen</u> to A if we alter B"
- ➔ Needs its own mathematical formulations

Intervention

- ➔ The easiest way to find causality is an intervention.
- ➔ Proper randomization
- ➔ We randomly force A to have different values and we measure B.
- ➔ If we can do that, you're done, but you can't always do that.
- ➔ It may be unethical to give people ineffective treatments to deadly diseases, or they may be have some say in their treatment, e.g., they may choose the less harsh (treatment B) when their kidney stones are small and less painful.

Randomized Controlled Trials

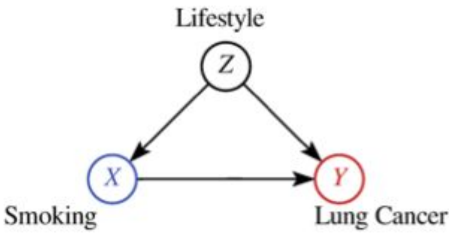


- ➔ Clinical experiments → explores causal effect
- ➔ Treatment-Control groups → compares causal effect of treatment
- ➔ Randomization → ensures similar population groups

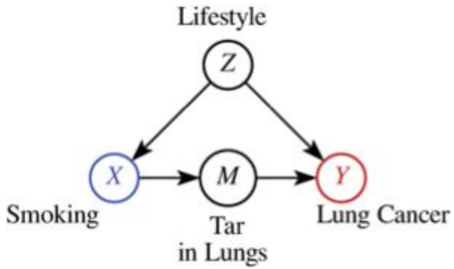
Structural Causal Models

- Structural Causal Model **M**
  - 4-tuple **< U, V, f, P(u) >**
  - **U** → background (exogenous) variables
  - **V** → observable (endogenous) variables
  - **F** → functions, mapping from **U** to **V**
  - **P(u)** → probability distribution over exogenous variables **U**

Treatment and Outcome



With Confounder

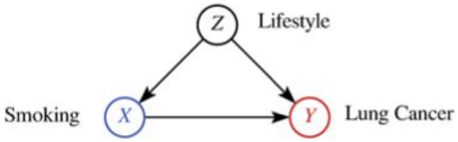


With Mediator

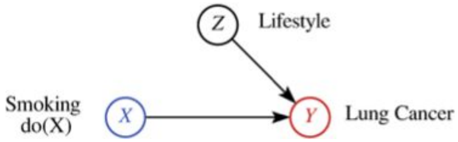
do-Intervention: Intervention by do-Calculus

- Conditioning → Observing data without action → Conditional dependence
- Intervention → Taking an action → Manipulated data
- do-Intervention → **P(Y|do(X))**

Conditioning
P(Y|X), P(Y|X,Z), P(Z|X,Y)

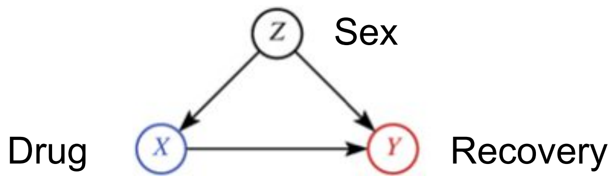


Manipulation
P(Y|do(X)), P(Y|do(X),Z)



Backdoor Adjustment on Simpson's Paradox

- Backdoor Adjustment Formula
  - Based on do-Intervention
  - **$P(y|do(x)) = \sum P(y|x,z) P(z)$**

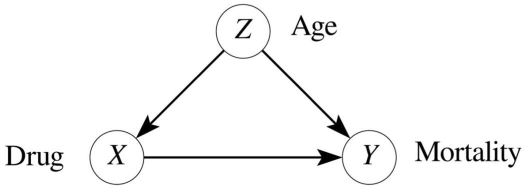


- Effect:
  $P(Y=1|X=1) - P(Y=1|X=0)$
  $= 78\% - 83\% = -5\% =$ **-0.05**
- Causal effect:
  $P(Y=1|do(X=1)) - P(Y=1|do(X=0))$
  $= \sum P(Y=1|X=1,Z) P(Z) - \sum P(Y=1|X=0,Z) P(Z)$
  $= [\ P(Y=1|X=1,Z=1) P(Z=1) + P(Y=1|X=1,Z=1) P(Z=1)\ ]$
  $-\ [\ P(Y=1|X=1,Z=1) P(Z=1) + P(Y=1|X=1,Z=1) P(Z=1)\ ]$
  $= [\ 93\% + 73\%\ ] - [\ 87\% + 69\%\ ] = 10\% =$ **0.1**

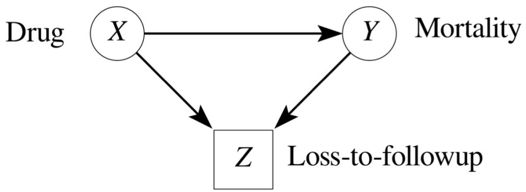
	Drug	No drug
Men	81 out of 87 (93%)	234 out of 270 (87%)
Women	192 out of 263 (73%)	55 out of 80 (69%)
Combined data	273 out of 350 (78%)	289 out of 350 (83%)

Biases

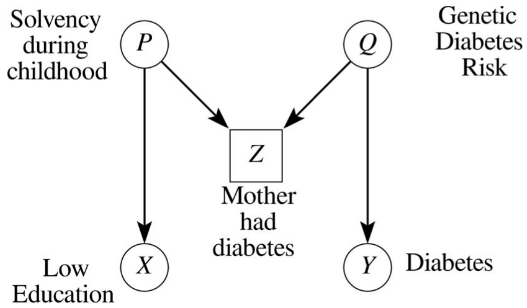
Confounding Bias: Common cause for treatment and outcome



Selection Bias: Bias injected by selecting certain group



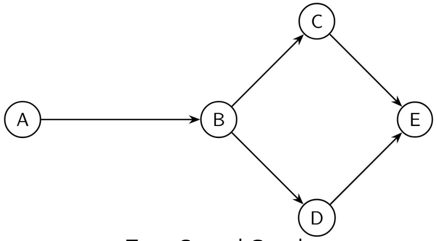
M Bias: Common outcome of causes for treatment and outcome



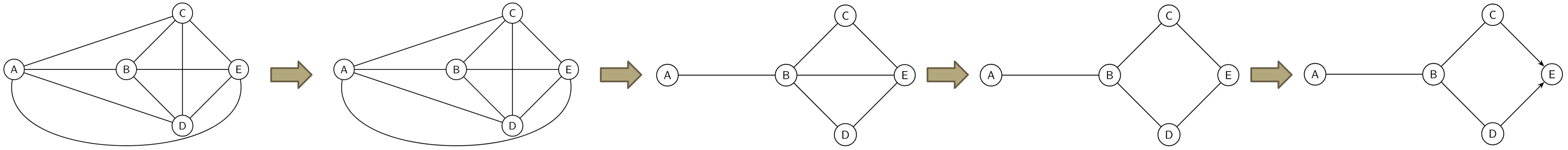
Causal Structure Learning Algorithms

- Learns causal structure from:
  - data → conditional probability distribution
  - Assumptions

- **Constraint-based methods:** Peter-Clark (PC), Fast causal inference (FCI)
- **Score-based methods:** Greedy equivalence search (GES), Greedy interventional equivalence search (GIES)
- **Hybrid methods:** Max-min hill climbing (MMHC)



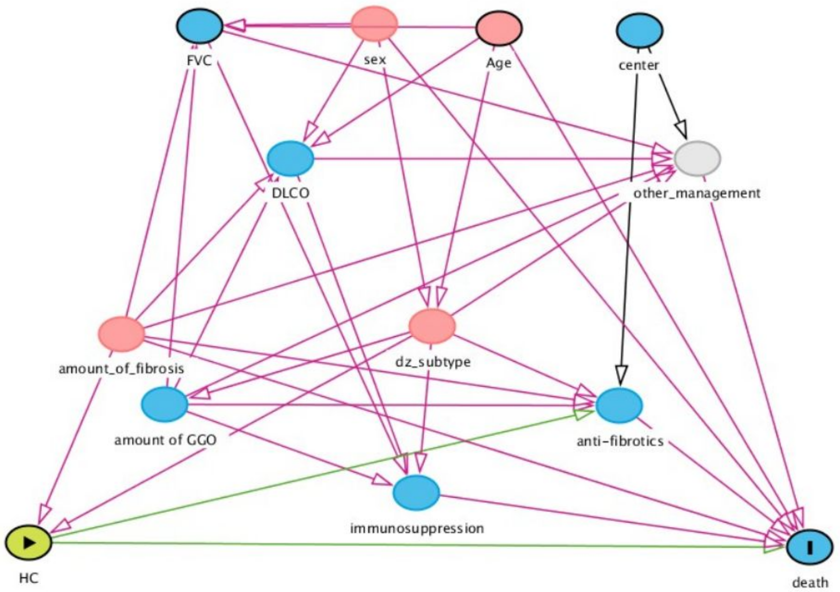
True Causal Graph



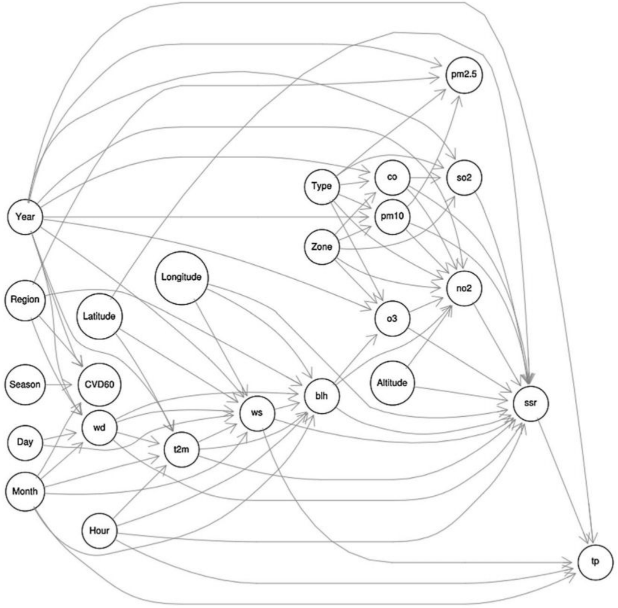
Causal Graph Learning Process in PC Algorithm

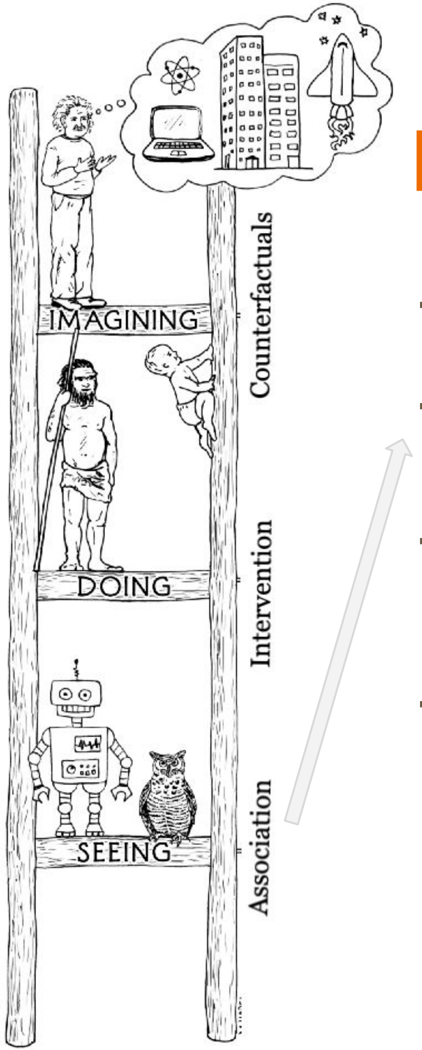
Applications of Causal Structure Learning Algorithms

Computed Tomography Honeycombing Identifies a Progressive Fibrotic Phenotype With Increased Mortality Across Diverse Interstitial Lung Diseases



Modeling air pollution, climate, and health data using Bayesian Networks: A case study of the English regions





Ladder of Causation

- ➔ Availability of big data → Machine Learning → Causal Inference

- ➔ Association
  - One object is associated with another if observing one changes the probability of observing the other.
  - Example: shoppers who buy toothpaste are more likely to also buy dental floss.
  - Mathematically: **P (floss | toothpaste)**
- ➔ Intervention
  - Specific causal relationships between events, assessed by experimentally performing some action that affects one of the events.
  - Example: if we doubled the price of toothpaste, what would be the new probability of purchasing?
  - Mathematically: **P (floss | *do*(toothpaste))**
- ➔ Counterfactuals
  - consideration of an alternate version of a past event, or what would happen under different circumstances for the same experimental unit.
  - Example, what is the probability that, if a store had doubled the price of floss, the toothpaste-purchasing shopper would still have bought it?
  - Mathematically: **P* (floss | toothpaste, *do*(price = x2))**

Research Problems

- ➔ New structure learning algorithms
- ➔ Ensembling of outputs of learning algorithms
- ➔ New / altered assumptions
- ➔ Extraction of proper data from observational data
- ➔ Application in specific branches: Epidemiology, Sociology, Finance, Economics

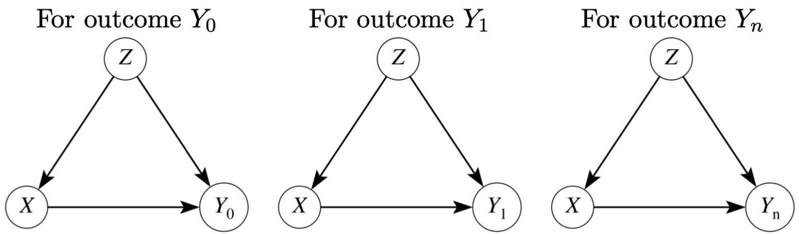
My Research Works

- Theoretical Works:
  - Survival Analysis: Hazard Ratio Estimation for Observational Data through Causal Inference
  - Causal Knowledge Hierarchy: Causal Model Estimation from Data and Background Knowledge, through Causal Knowledge Hierarchy

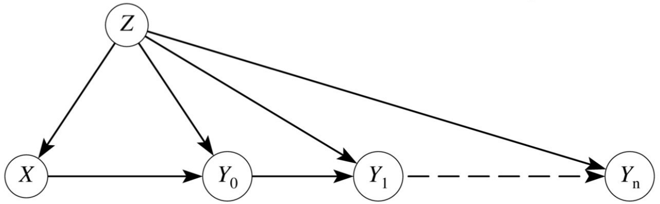
- Applications:
  - Efficacy estimation of Antipsychotic Drugs for Delirium-induced Patients in the ICU

Causally Formulated Hazard Ratio Estimation

- ➔ Survival Analysis → Hazard Ratio
- ➔ *Hypothesis:* SCM can aid in generation of causally interpretable HR.



(a) Converted Causal DAGs with no dependency between Y_is



(b) Converted single Causal DAG with dependencies between Y_is

Algorithm 1 Causally Formulated Hazard Ratio Estimation

Input: Z, X, T, S
Output: $survival\_curve, HR_{drug}$

global $n \leftarrow length(T)$
global $t_{max} \leftarrow max(T)$
$Y_i \leftarrow convert\_single\_to\_multiple\_trials(T, S)$
for $i \leftarrow 0$ to t_{max} **do**
  $adj\_c_i, adj\_p_i \leftarrow adjust\_backdoor(Z, X, Y_i)$
end for
$survival\_curve \leftarrow plot(time, cumulative(adj\_p_i))$
$adj\_X, adj\_T \leftarrow convert\_multiple\_to\_single\_trial(adj\_c_i)$
$model \leftarrow cox\_ph\_model(adj\_X, adj\_T)$
$HR_{drug} \leftarrow exp(model.\beta_{drug})$
return $survival\_curve, HR_{drug}$

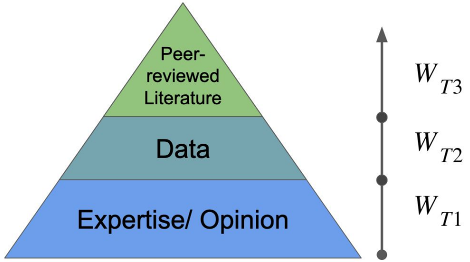
Algorithm 2 Conversion of single trial to multiple trials

Input: T, S
Output: Y_i

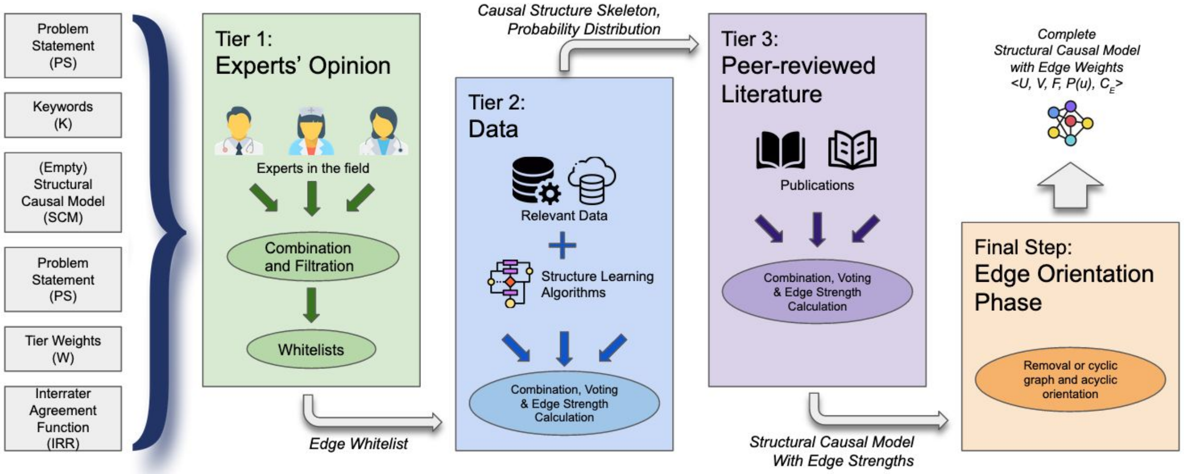
for $i \leftarrow 0$ to t_{max} **do**
  for $j \leftarrow 0$ to n **do**
    $Y_i[j] \leftarrow (T[j] <= i)$ **and** $(S[j] = 1)$? 1 : 0
  end for
end for
return Y_i

Causal Model Estimation from Data and Background Knowledge

- Data alone cannot express causality
- *Question:* How can we merge Causal Knowledge from different sources
- Causal Knowledge Hierarchy



Causal Knowledge Hierarchy



General Process Flow

Application: Efficacy of APD in Delirium Treatment

- ➔ Delirium occurs in 80% of cases in patients in the ICU
- ➔ Common treatment is prescription of Antipsychotic Drugs
- ➔ Efficacy of APD is not well-established
- ➔ *Question*: Can we find the true causal effect of APD in the ICU from observational data?
- ➔ MIMIC-III dataset, publicly available healthcare data of ~50K patients

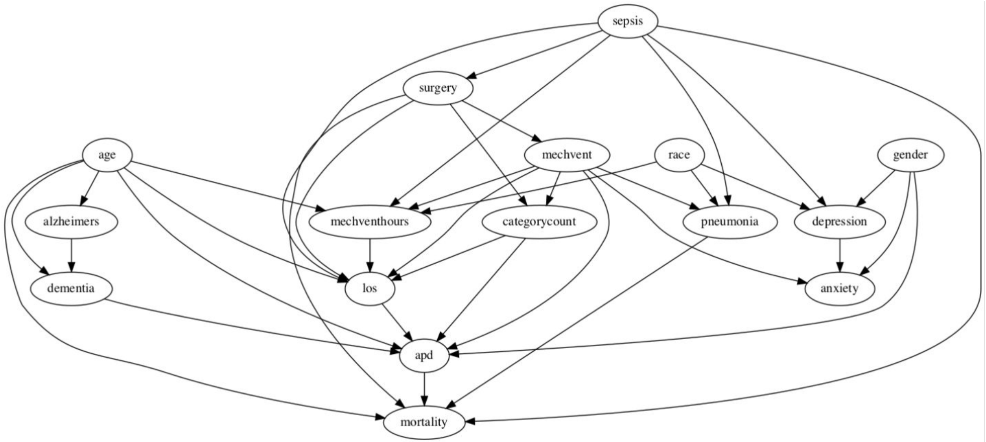


Figure: A simplified Causal DAG for Delirium induction in the ICU with 16 covariates (results of preliminary analysis)

References

- ➔ https://en.wikipedia.org/wiki/Causal_inference
- ➔ https://en.wikipedia.org/wiki/Simpson%27s_paradox
- ➔ Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.
- ➔ Pearl, J. (2009). Causality. Cambridge university press.
- ➔ Bareinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. Proceedings of the National Academy of Sciences, 113(27), 7345-7352.
- ➔ Adib, R., Griffin, P., Ahamed, S. I., & Adibuzzaman, M. (2020, September). A Causally Formulated Hazard Ratio Estimation through Backdoor Adjustment on Structural Causal Model. In Machine Learning for Healthcare Conference (pp. 376-396). PMLR.
- ➔ Bikak, M., Adib, R., Ingram, W., Griffin, P., & Adibuzzaman, M. (2019). Outcomes of Use of Antipsychotic for Delirium in the ICU: A Big Data Approach. In D50. CRITICAL CARE: THE METAMORPHOSIS-PAIN, SEDATION, DELIRIUM, ICU-ACQUIRED WEAKNESS, AND PALLIATIVE CARE (pp. A6672-A6672). American Thoracic Society.

Thank You

- Any questions?
- Please feel free to reach out: *riddhiman.adib@marquette.edu*
- More about my research updates: *adib2149.github.io*