Causal Exploration on Large Healthcare Datasets through Emulation of Randomized Controlled Trials: An Application on Delirium Patients in the ICU

Riddhiman Adib [1], Mohammad Adibuzzaman [2], **Advisor:** Sheikh Iqbal Ahamed [1]

[1] Ubicomp Lab, Department of Computer Science, Marquette University

[2] Regenstrief Center for Healthcare Engineering, Purdue University

Background

- Availability and access to large healthcare datasets, collected over a long period, holding valuable information
- Commonly explored through traditional Machine Learning algorithms (Regressions, Neural Networks, Curve fitting algorithms)
- Perform great in finding patterns out of datasets
  - lacks extensive interpretability to be used in the healthcare sector
- Without exploring underlying causal relationships, the algorithms fail to explain their reasoning
- Causal Inference, a relatively newer branch of Artificial Intelligence, deals with the issue of interpretability
  - works towards an explanation of causality in data through graphical models (Pearl et al. 2016)

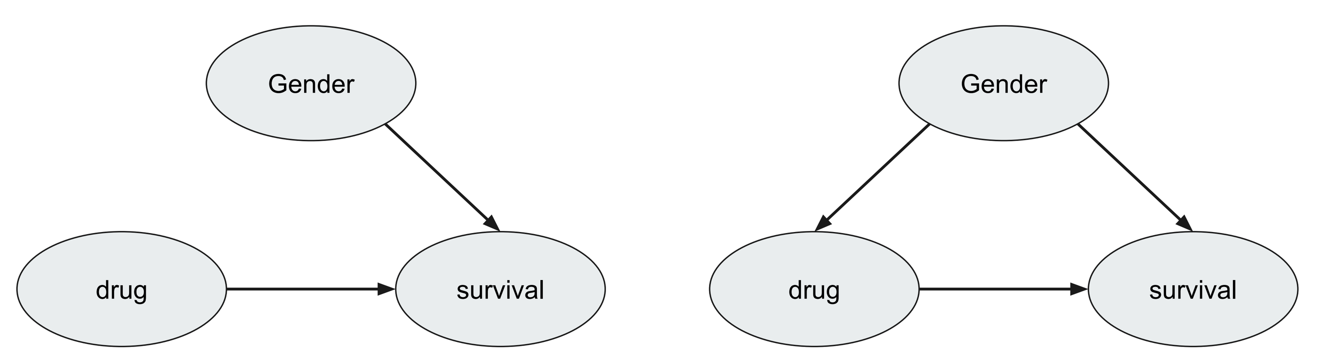

Figure 1: Causal Structure for experimental studies through RCTs versus observational studies (presence of confounding variable as backdoor)

Highlights of Research Idea

- Our research goal is to use Causal Inference to build an applied framework that lets researchers leverage observational datasets in understanding causal relationships between different variables (events, treatments, outcomes, demographic features)
- To achieve that, our objectives are:
  - **Emulation**: to emulate a Randomized Controlled Trial (RCT) from observational dataset with consideration on randomization and minimal biases
  - **Theories**: to generate theories that explain causal relationships through standard statistical analyses (e.g. Odds Ratio, Survival Analysis) using Causal Inference framework
  - **Application**: to apply it to find the efficacy of antipsychotic drugs prescribed in the treatment of Delirium in the ICU.

Approach

- **Data Mining:** extraction of relevant data from large dataset, exploratory data analysis
- **Statistical Analyses:** emulating target trial, relevant statistical analyses on emulated trial endpoint efficacy measurement along with survival analysis
- **Causal Structure Model Generation:** structure learning algorithms, background knowledge and existing literature
- **Causal Inference:** Fundamentally work through theories of Causal Inference by studying existing publications and books, estimation of effects using causal structure, generation of theories relating Survival Analysis and Causal Inference

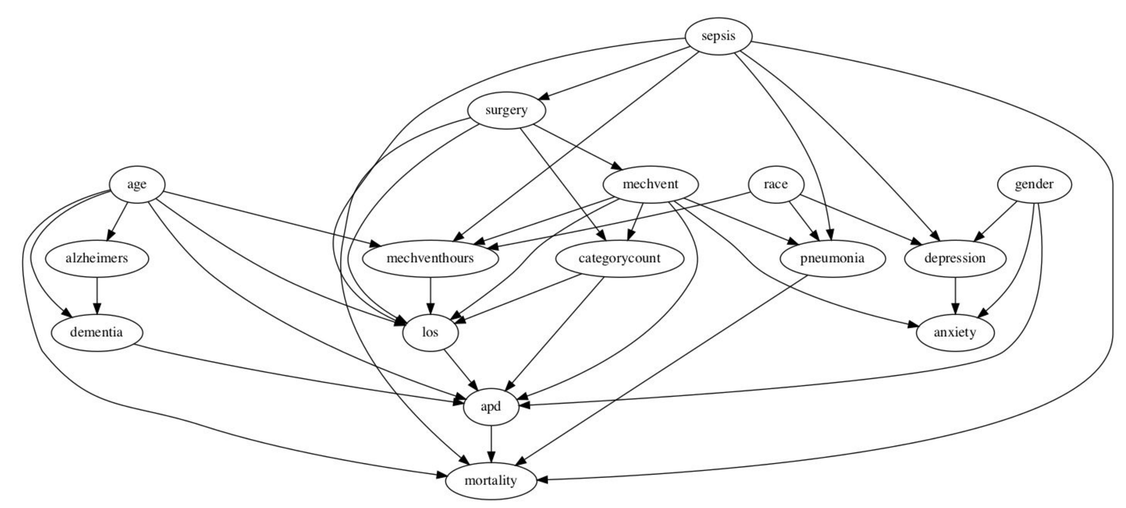

Figure 2: A work-in-progress showcasing causal structure for delirium patients in ICU

Forward Thinking

- The research area of emulating RCTs using an observational dataset is highly expected, however relatively newer, and still contains many assumptions
- Researchers have shown the possibility of using observational datasets in emulating RCTs for antiretroviral therapy (Lodi et al. 2019), and for ARDS (Bikak et al. 2018)
  - Still not perfected, contains biases, lacks trust and validity
  - We plan to build our approach on top of these existing research works and find novel ways to address the shortcomings
- The forward-thinking in our proposition would be:
  - to propose a standard mathematical and statistical framework to build (emulate) RCTs from Observational dataset with minimal bias
  - to propose Causal Inference approach to existing statistical methodologies, like Survival Analysis
  - to showcase new way to explore large datasets and interpret their underlying causal structure

Significance

- Delirium occurs in about 80% cases in the Intensive Care Unit (ICU) and is commonly treated with antipsychotic drugs (APD) (Girard et. al. 2008)
- Controversy over usage of APDs in treating Delirium, since RCTs do not agree in clear evidence of similar efficacy or safety
- Observational data has potential to resolve this issue through Causal Inference
  - MIMIC III database, an extensive EHR dataset with 53,423 distinct hospital admissions (Adibuzzaman et. al. 2016)
- Elimination of controversies, Cost-effective virtual RCTs, Datasets from around the globe
- Causal Inference in observational datasets is not limited to only health sector, it also helps in explaining problems in the field of Social Science, Economics etc.

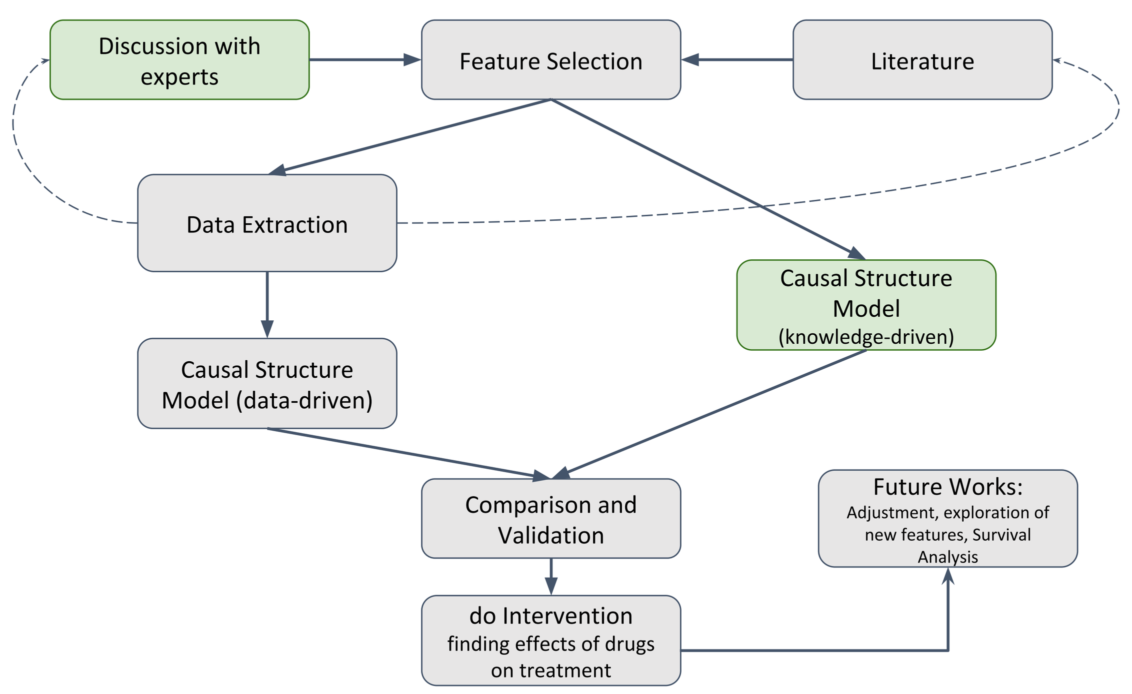

Figure 3: Workflow diagram

References

1. Girard, T. D., Pandharipande, P. P., & Ely, E. W. (2008). Delirium in the intensive care unit. Critical care, 12(S3), S3.
2. Bikak, M., Adibuzzaman, M., Jung, Y., Yih, Y., & Bareinboim, E. (2018). Regenerating Evidence from Landmark Trials in ARDS Using Structural Causal Models on Electronic Health Record. In B104. Critical Care: Big Data in HealthCare - Predictive Analytics, Clinical Decision Support, And Rapid Response (pp. A4290-A4290). American Thoracic Society.
3. Adibuzzaman, M., Musselman, K., Johnson, A., Brown, P., Pitluk, Z., & Grama, A. (2016, September). Closing the data loop: An integrated open access analysis platform for the mimic database. In 2016 Computing in Cardiology Conference (CinC) (pp. 137-140). IEEE.
4. Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.
5. Lodi, S., Phillips, A., Lundgren, J., Logan, R., Sharma, S., Cole, S. R., ... & Horban, A. (2019). Effect estimates in randomized trials and observational studies: comparing apples with apples. American journal of epidemiology, 188(8), 1569-1577.