Causal Variance Decompositions for Measuring Health Inequalities

Lin Yu[1], Zhihui (Amy) Liu[2], Olli Saarela[1]
[1] Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto
[2] Department of Biostatistics, University Health Network

Introduction

Hospital profiling:

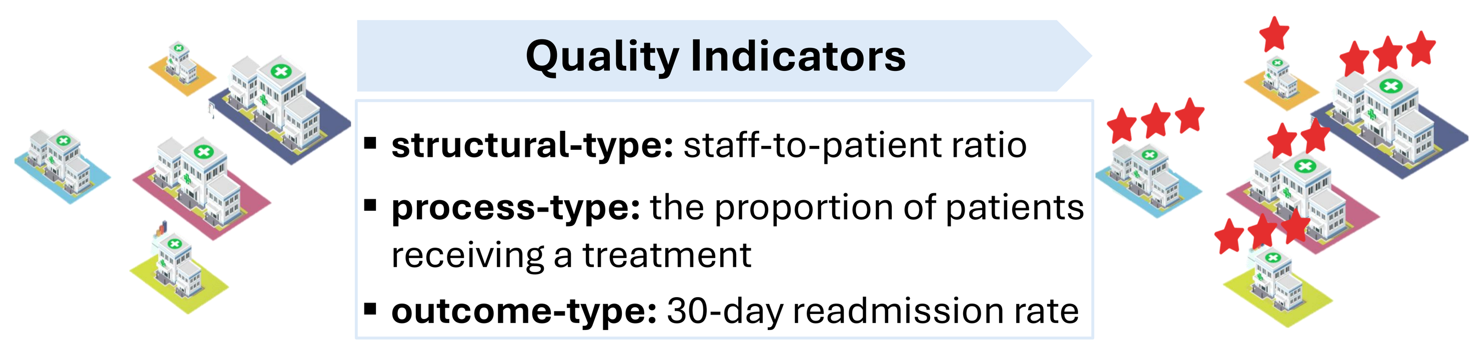

Figure 1: Hospital performance profiling based on quality indicators

Health inequalities: differences in health status across sociodemographic groups

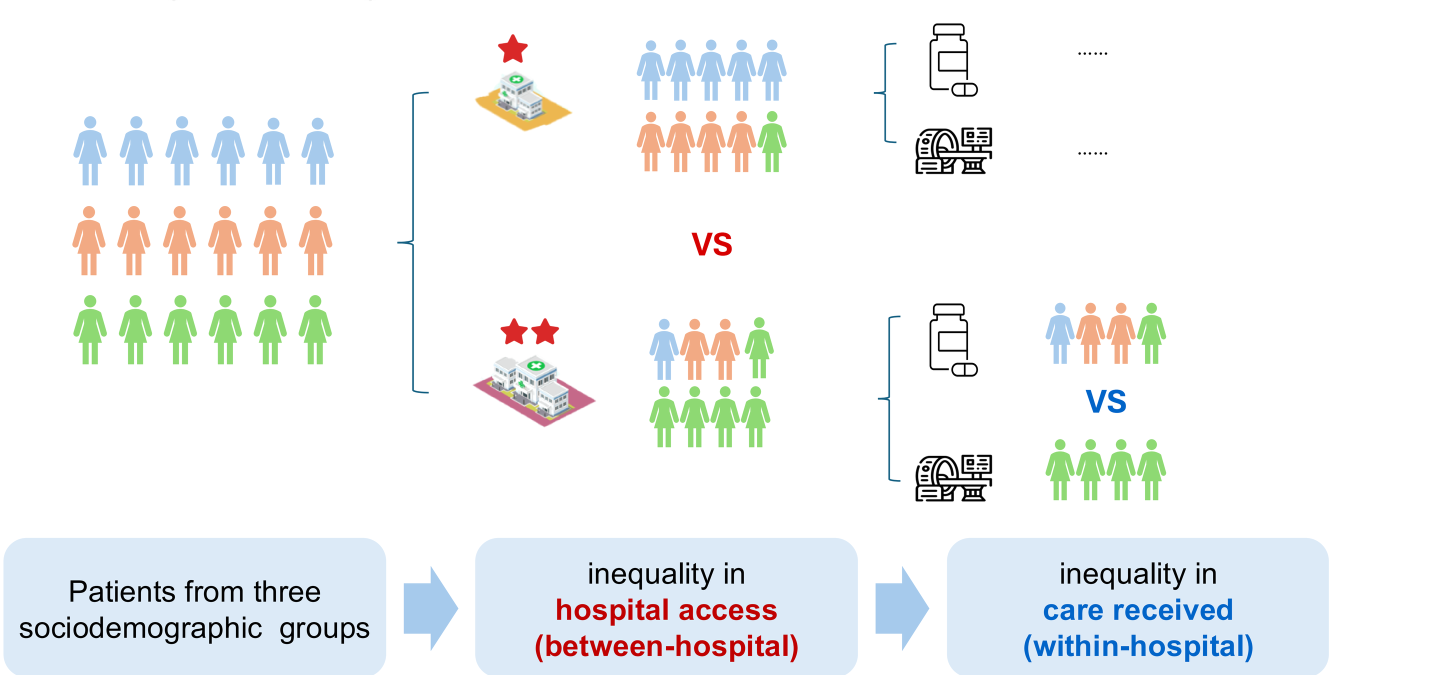

Figure 2: A hypothetical example showing two potential sources of health inequities

Challenges and Objectives

Challenges: existing methods are limited to pairwise comparisons or cannot account for health inequalities

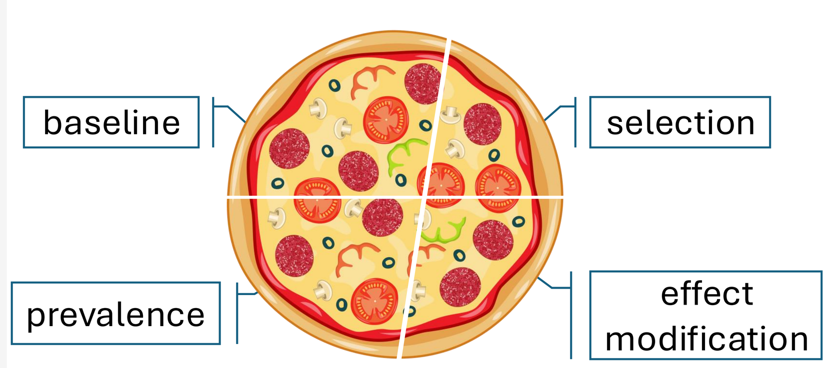

Figure 3: Effect Decomposition

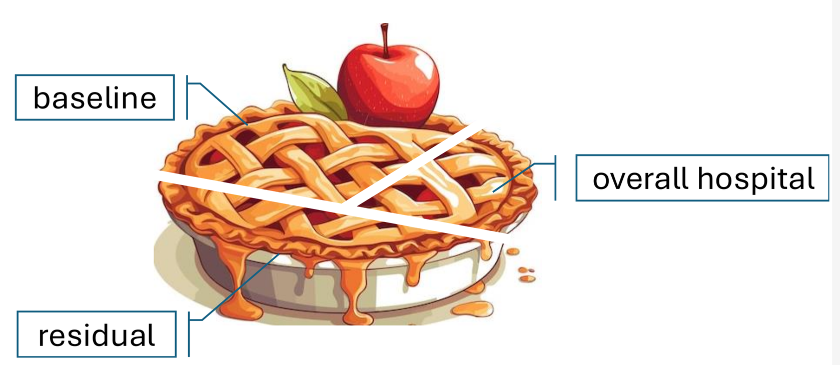

Figure 4: Variance Decomposition

Objectives:

- Develop a new causal variance decomposition to quantify inequalities in healthcare access and differential care delivery, especially for **multi-categorical** exposure and effect modifier
- Construct corresponding estimators and assess the performance of these estimators in simulation

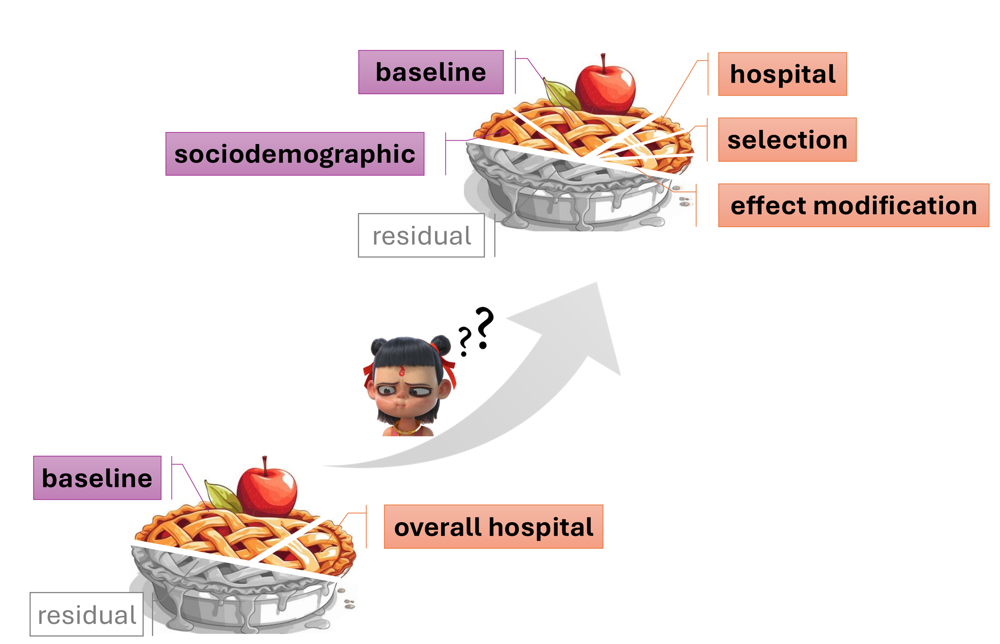

Figure 5: Main research objectives and key contributions

Methods

Causal DAG

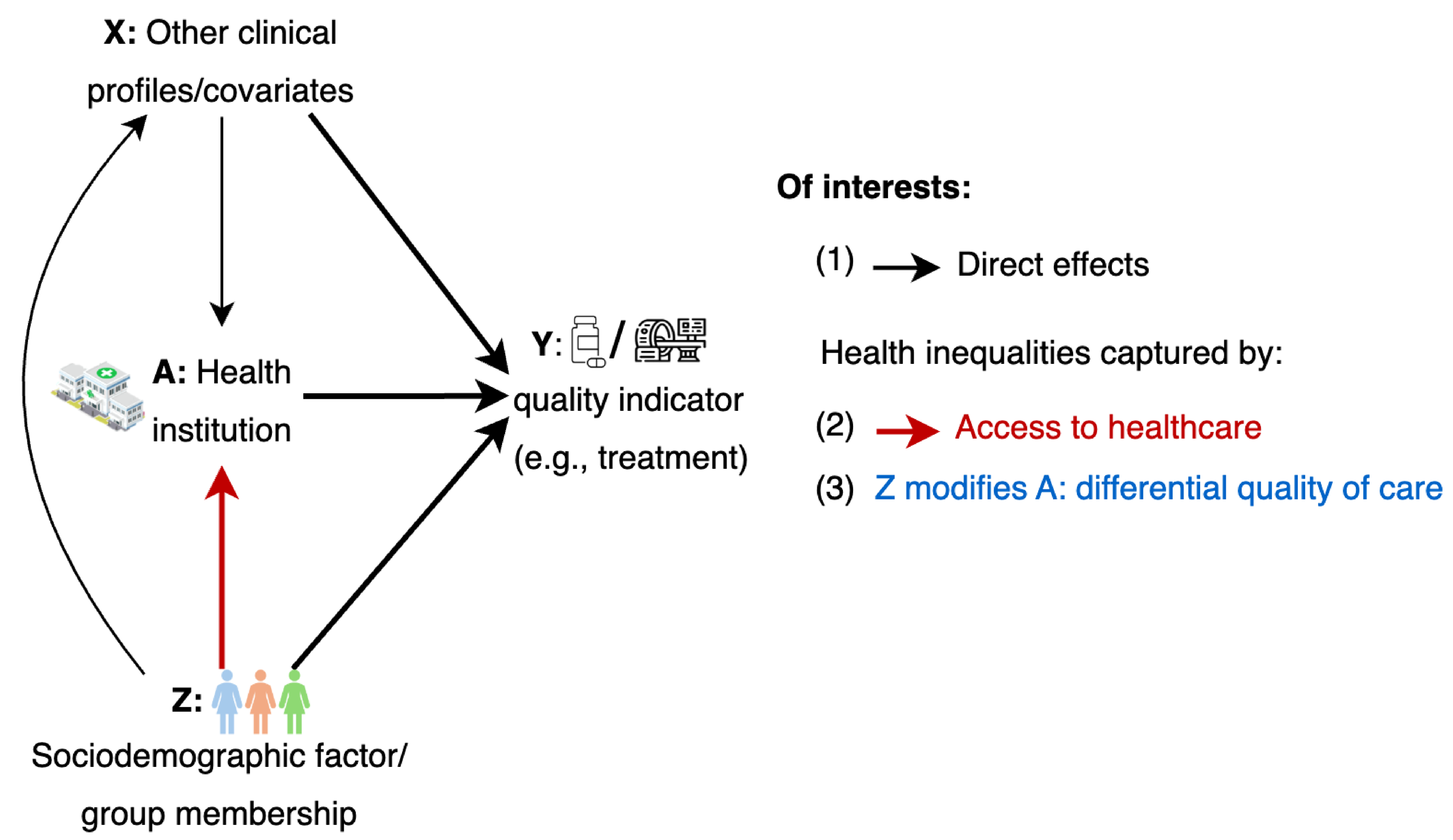

Figure 6: Causal DAG (Directed Acyclic Graph) illustrating our research question

Proposed six-way variance decomposition

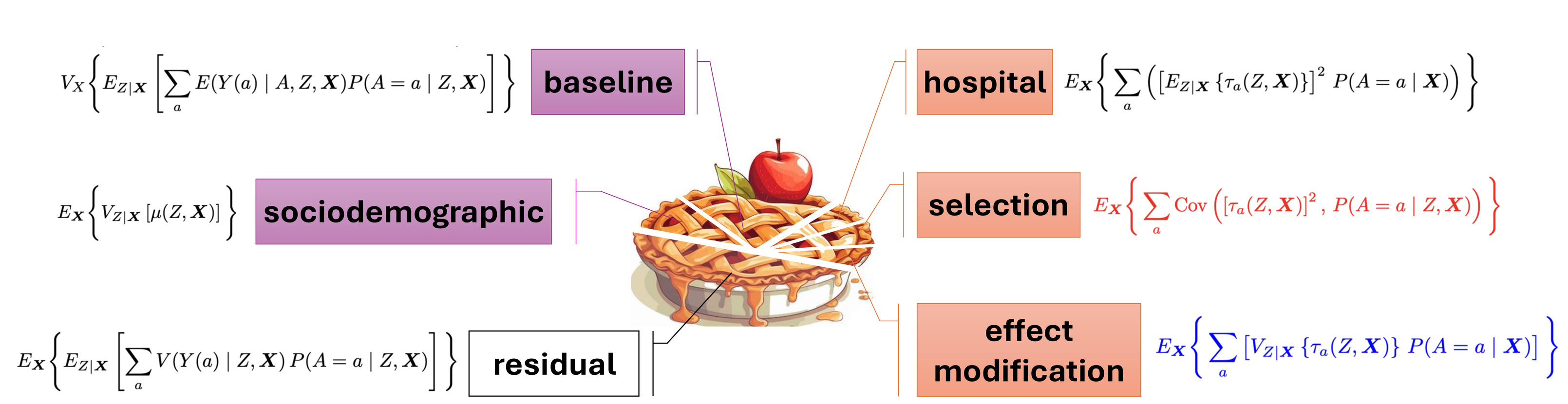

Figure 7: Six-way variance decomposition for measuring health inequalities

Where

- $\mu(Z, \boldsymbol{X}) = \sum_a E(Y(a) \mid Z, \boldsymbol{X}) P(A = a \mid Z, \boldsymbol{X})$
- $\tau_a(Z, \boldsymbol{X}) = E(Y(a) \mid Z, \boldsymbol{X}) - \mu(Z, \boldsymbol{X})$: the causal effect of being treated in hospital a, compared to the "average hospital"

Identification

- Assumptions:
  - A1 (Consistency): $Y = Y(A)$
  - A2 (Exchangeability): $Y(a) \perp A \mid Z, \boldsymbol{X}$
  - A3 (Positivity): $0 < P(A \mid Z, \boldsymbol{X}) < 1$
- Identification Results
  - $V(Y(A) \mid \boldsymbol{X}) = V(Y \mid \boldsymbol{X})$ (by A1)
  - $E(Y(a) \mid Z, \boldsymbol{X}) = E(Y \mid A = a, Z, \boldsymbol{X})$ (by A1, A2, A3)

Model-Based Estimators

- Outcome model: $E(Y \mid A, Z, \boldsymbol{X})$
  - Parametric: GLM, GLMM
  - Non-parametric: Random Forests, XGBoost
- Assignment models:
  - Hospital: $P(A \mid Z, \boldsymbol{X})$ via multinomial regression
  - Sociodemographic: $P(Z \mid \boldsymbol{X})$ via multinomial regression

Simulation Study

Data Generating Mechanism

- Covariates: $X_1 \sim \text{Bern}(0.5)$ and $X_2 \sim N(0, 1)$
- Outcome $E(Y \mid A, Z, \boldsymbol{X})$: logistic regression
- Hospital/group membership assignment:
  - $P(A \mid Z, \boldsymbol{X})$: multinomial regression
  - $P(Z \mid \boldsymbol{X})$: multinomial regression

Simulation Scenarios

- Sample size: 500, 1000, 2500, 5000
- Hospitals (A): 5 levels
- Sociodemographic groups (Z): 3 levels
- 1000 replicates

Simulation Results

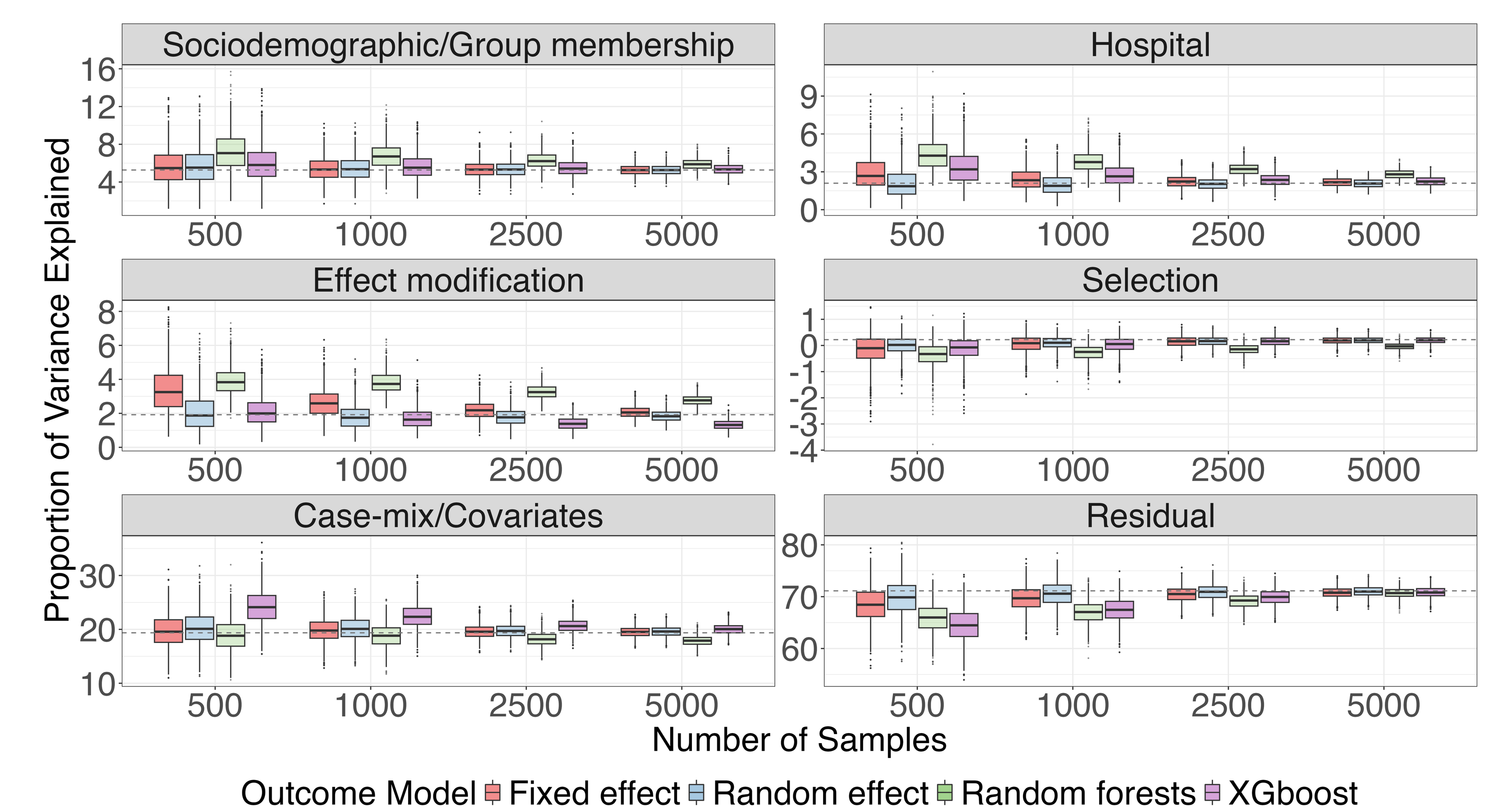

Figure 8: The estimated proportion of variance explained by each component under different sample sizes in 1000 replicates

- Evidence of **consistency** for all six variance component estimators
- Small-sample bias, perform better as the sample size increases

Conclusion

Contributions: Our proposed method reveals **sources and magnitudes** of health inequalities, offering insights for policymakers and healthcare professionals to promote equity in healthcare delivery

Limitation:

- Estimators are sensitive to model misspecification
- Non-parametric methods are sensitive to hyperparamter tuning

Future work:

- Extend the analysis to survival outcome
- Establish a decomposition methodology to incorporate **moderated mediation** into current decomposition