

Causal Variance Decompositions for Measuring Health Inequalities

Lin Yu¹ Zihui (Amy) Liu² Olli Saarela¹

¹Dalla Lana School of Public Health, University of Toronto

²University Health Network



UNIVERSITY OF TORONTO
DALLA LANA SCHOOL OF PUBLIC HEALTH

May 27, 2025

Hospital Performance Comparison



Hospital 1

VS



Hospital 2

Hospital Performance Comparison



Hospital 1

VS



Hospital 2

- Hospital profiling: hospital comparison using **quality indicators**

Hospital Performance Comparison



Hospital 1

VS



Hospital 2

- Hospital profiling: hospital comparison using **quality indicators**
 - Staff-to-patient ratio

Hospital Performance Comparison



VS



Hospital 1

Hospital 2

- Hospital profiling: hospital comparison using **quality indicators**
 - Staff-to-patient ratio
 - Proportion of patients receiving recommended treatments

Hospital Performance Comparison



VS



Hospital 1

Hospital 2

- Hospital profiling: hospital comparison using **quality indicators**
 - Staff-to-patient ratio
 - Proportion of patients receiving recommended treatments
- Perform better on average \xrightarrow{x} **subgroups** experience care similarly

Hospital Performance Comparison



VS



Hospital 1

Hospital 2

- Hospital profiling: hospital comparison using **quality indicators**
 - Staff-to-patient ratio
 - Proportion of patients receiving recommended treatments
- Perform better on average \xrightarrow{x} **subgroups** experience care similarly
- **Health inequality**: the differences in care received/health outcome in different **sociodemographic groups** (Hartman and Dahlerus, 2024)

Sources of Health Inequalities in Healthcare Delivery

- First source: **Hospital access/selection to hospitals**

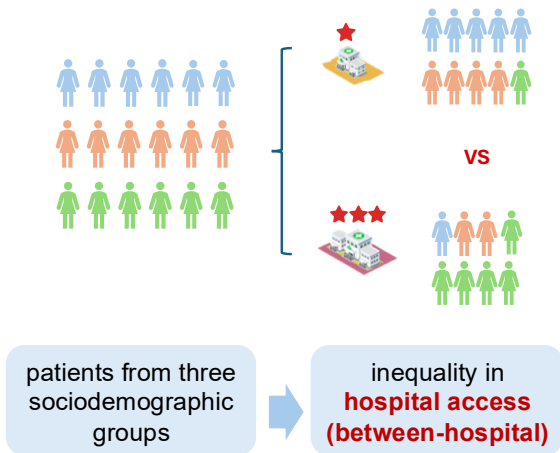


Figure 1: health inequalities in hospital access

Sources of Health Inequalities in Healthcare Delivery

- Second source: **quality of care received/effect modification**

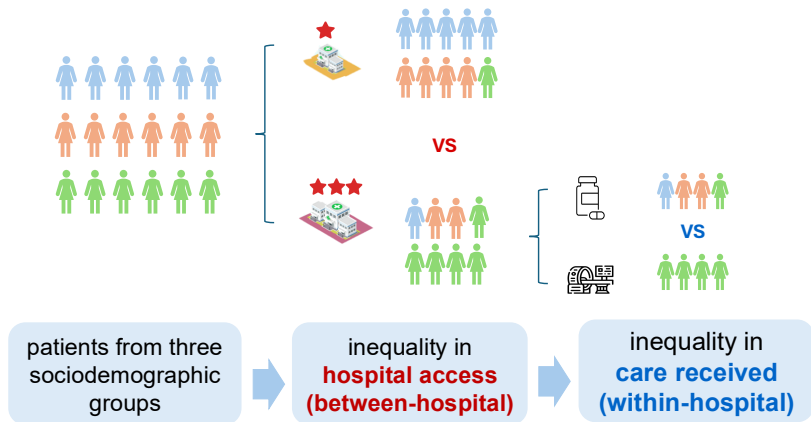


Figure 2: health inequalities in quality of care received

Research Question

- How can we quantify health inequalities in healthcare delivery?

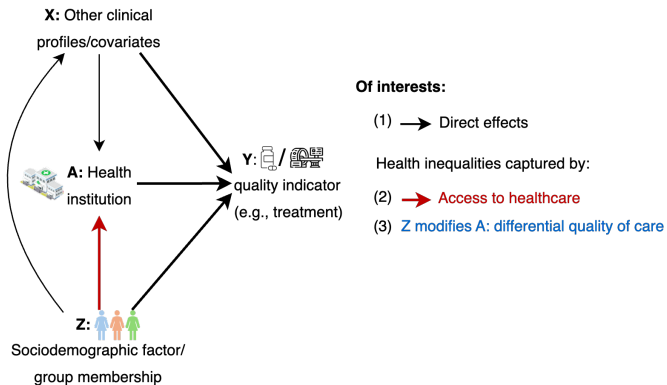


Figure 3: Causal DAG illustrating our research question

Related Literature


1) Effect Decomposition:

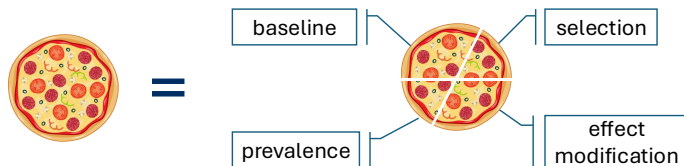
- Observed health inequality = $E_{z_1}(Y) - E_{z_2}(Y) =$



Related Literature


1) Effect Decomposition:

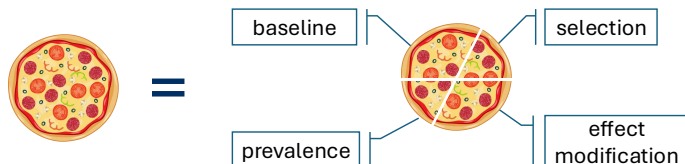
- Observed health inequality = $E_{z_1}(Y) - E_{z_2}(Y) =$ 
- Decompose observed health inequality (Yu and Elwert, 2025)



Related Literature

1) Effect Decomposition:

- Observed health inequality = $E_{z_1}(Y) - E_{z_2}(Y) =$ 
- Decompose observed health inequality (Yu and Elwert, 2025)

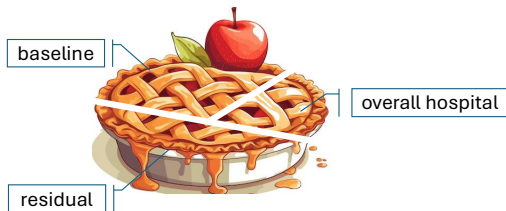


- Limitation: pairwise comparisons**
 - does not fit the hospital profiling context well (**multiple** hospitals and sociodemographic groups)

Related Literature

2) Variance Decomposition (Chen et al., 2019):

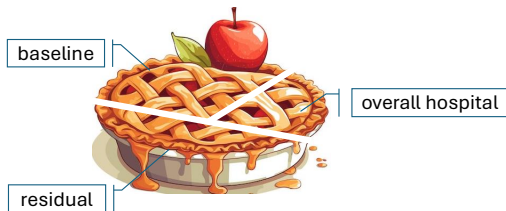
- Observed variation = $V(Y)$



Related Literature

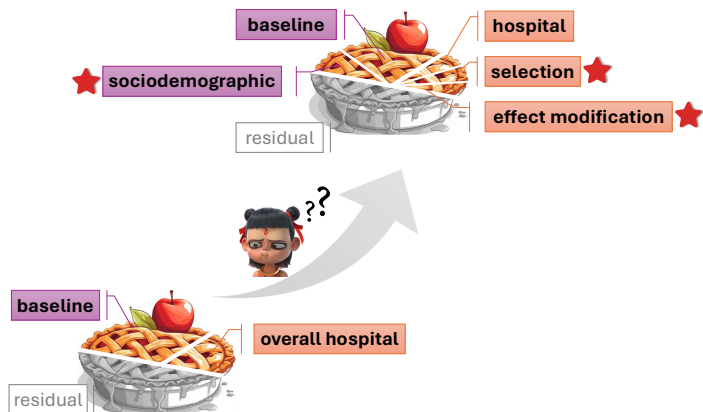
2) Variance Decomposition (Chen et al., 2019):

- Observed variation = $V(Y)$



- Limitation: does not directly account for effect modification and selection

Objectives



1) Variance Decomposition Derivation → 2) Estimator Development → 3) Performance Evaluation

Proposed Six-way Variance Decomposition

$$V(Y) = \text{[Pie Image]} \left\{ \begin{array}{l} \text{sociodemographic } E_{\mathbf{X}} \left\{ V_{Z|\mathbf{X}} [\mu(Z, \mathbf{X})] \right\} \\ \text{baseline } V_{\mathbf{X}} \left\{ E_{Z|\mathbf{X}} [\mu(Z, \mathbf{X})] \right\} \\ \text{hospital } E_{\mathbf{X}} \left\{ \sum_a \left([E_{Z|\mathbf{X}} \{\tau_a(Z, \mathbf{X})\}]^2 P(A = a | \mathbf{X}) \right) \right\} \\ \text{selection } E_{\mathbf{X}} \left\{ \sum_a \text{Cov} \left([\tau_a(Z, \mathbf{X})]^2, P(A = a | Z, \mathbf{X}) \right) \right\} \\ \text{effect modification } E_{\mathbf{X}} \left\{ \sum_a [V_{Z|\mathbf{X}} \{\tau_a(Z, \mathbf{X})\}] P(A = a | \mathbf{X}) \right\} \\ \text{residual } E_{\mathbf{X}} \left\{ E_{Z|\mathbf{X}} \left[\sum_a V(Y(a) | Z, \mathbf{X}) P(A = a | Z, \mathbf{X}) \right] \right\} \end{array} \right\}$$

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

Identification and Estimation

1 Assumptions: Consistency, Exchangeability and Positivity

2 Identification Results:

$$V(Y(A) | \mathbf{X}) = V(Y | \mathbf{X})$$
$$E(Y(a) | Z, \mathbf{X}) = E(Y | A = a, Z, \mathbf{X})$$

3 Model-based Estimators:

- Outcome model: $E(Y | A, Z, \mathbf{X})$
 - Parametric: Generalized linear/mixed-effect model
 - Non-parametric: Random Forests, XGBoost
- Assignment models:
 - Hospital $P(A | Z, \mathbf{X})$: multinomial regression
 - Sociodemographic $P(Z | \mathbf{X})$: multinomial regression

Simulation Setup

■ Data Generating Mechanism

- Covariates: $X_1 \sim \text{Bern}(0.5)$ and $X_2 \sim N(0, 1)$
- Outcome: logistic regression
- Hospital/group membership:
 - $P(A | Z, \mathbf{X})$: multinomial reg.
 - $P(Z | \mathbf{X})$: multinomial reg.

Simulation Setup

■ Data Generating Mechanism

- Covariates: $X_1 \sim \text{Bern}(0.5)$ and $X_2 \sim N(0, 1)$
- Outcome: logistic regression
- Hospital/group membership:
 - $P(A | Z, \mathbf{X})$: multinomial reg.
 - $P(Z | \mathbf{X})$: multinomial reg.

■ Simulation Scenarios

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

Simulation Results

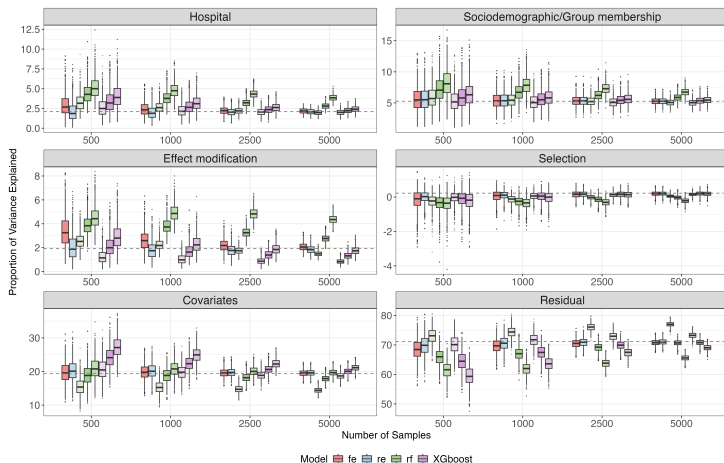


Figure 4: The estimated proportion of variance explained by each component under different sample sizes in 1000 replicates

Conclusions

■ Conclusions:

- Parametric-based estimators capture the true values reasonably well as the sample size increases (provided the models are correctly specified)
- Nonparametric-based estimators are sensitive to hyperparameter tuning

■ Next steps:

- Explore additional scenarios, particularly with a larger number of hospitals
- Extend the analysis to survival outcome

References

- Chen, B., Lawson, K. A., Finelli, A., and Saarela, O. (2019). Causal variance decompositions for institutional comparisons in healthcare. *Statistical Methods in Medical Research*, 29(7):1972–1986.
- Hartman, N. and Dahlerus, C. (2024). Evaluating medical providers in terms of patient health disparities: a statistical framework. *Health Services and Outcomes Research Methodology*, pages 1–18.
- Yu, A. and Elwert, F. (2025). Nonparametric causal decomposition of group disparities. *The Annals of Applied Statistics*, 19(1):821–845.

Thanks for listening:)