Causal Variance Decompositions for Measuring Health Inequalities

Lin Yu, Zhihui (Amy) Liu, Olli Saarela

2025-02-27

Background

Hospital performance is typically measured using quality indicators (QIs) that assess structural, process, or outcome elements related to a specific condition(Lazar, Fleischut, and Regan 2013).

Examples of QIs:

- structural-type: staff-to-patient ratio
- process-type: the proportion of patients receiving a treatment
- outcome-type: 30-day readmission rate

Comparison of hospital performance could reveal **health inequalities** in care delivery between sociodemographic groups.

- Health inequality is the differences in access to healthcare, treatment, and health outcomes in different sociodemographic groups(Gakidou, Murray, and Frenk 2000)

Sources of Health Inequalities in Healthcare Delivery

Setting: Consider a group of patients from different racial backgrounds but with similar clinical profiles, and we compare a process-type QI , such as receiving medication or undergoing surgery.

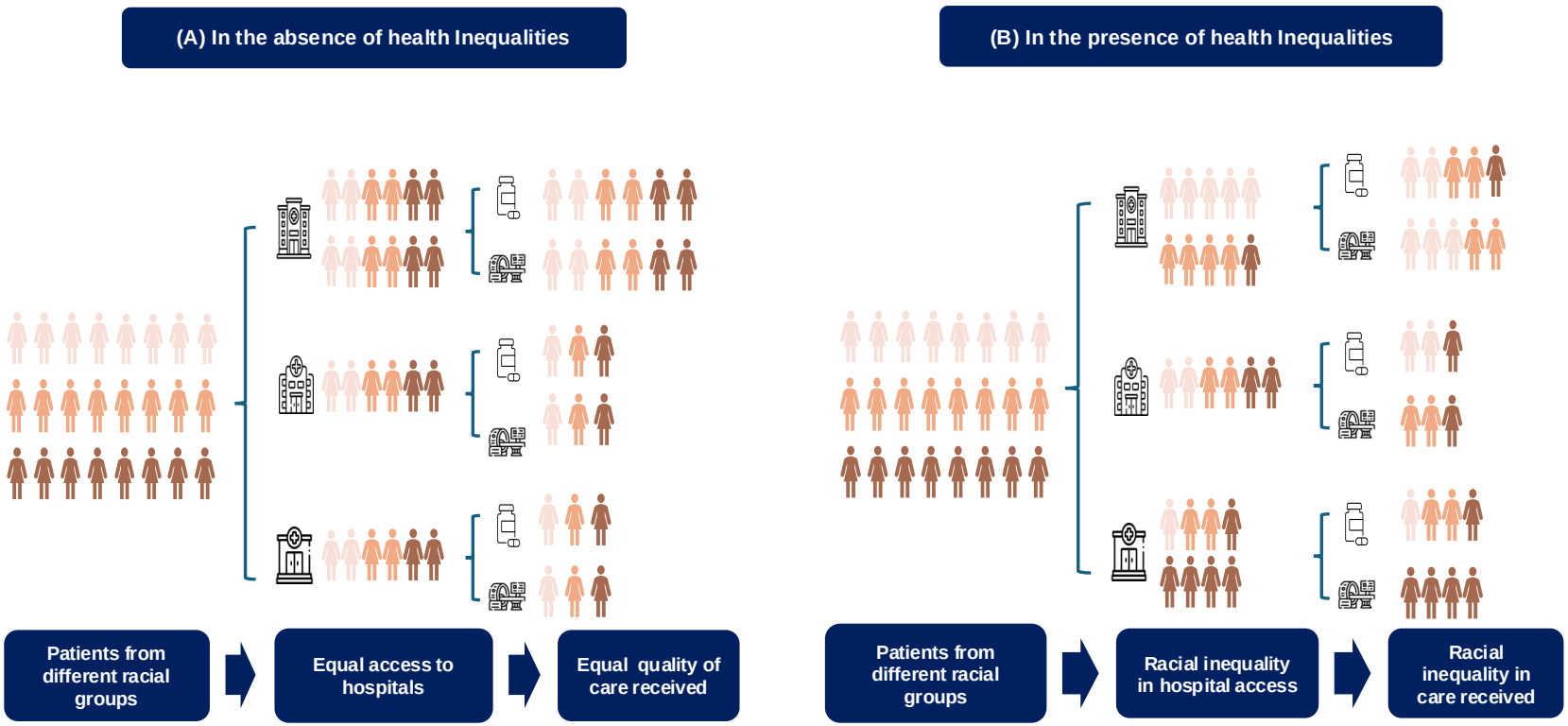

Figure 1: Potential sources of health inequalities

Two sources: Access to healthcare + Quality of care received

Research Question

How can we quantify health inequalities in healthcare delivery?

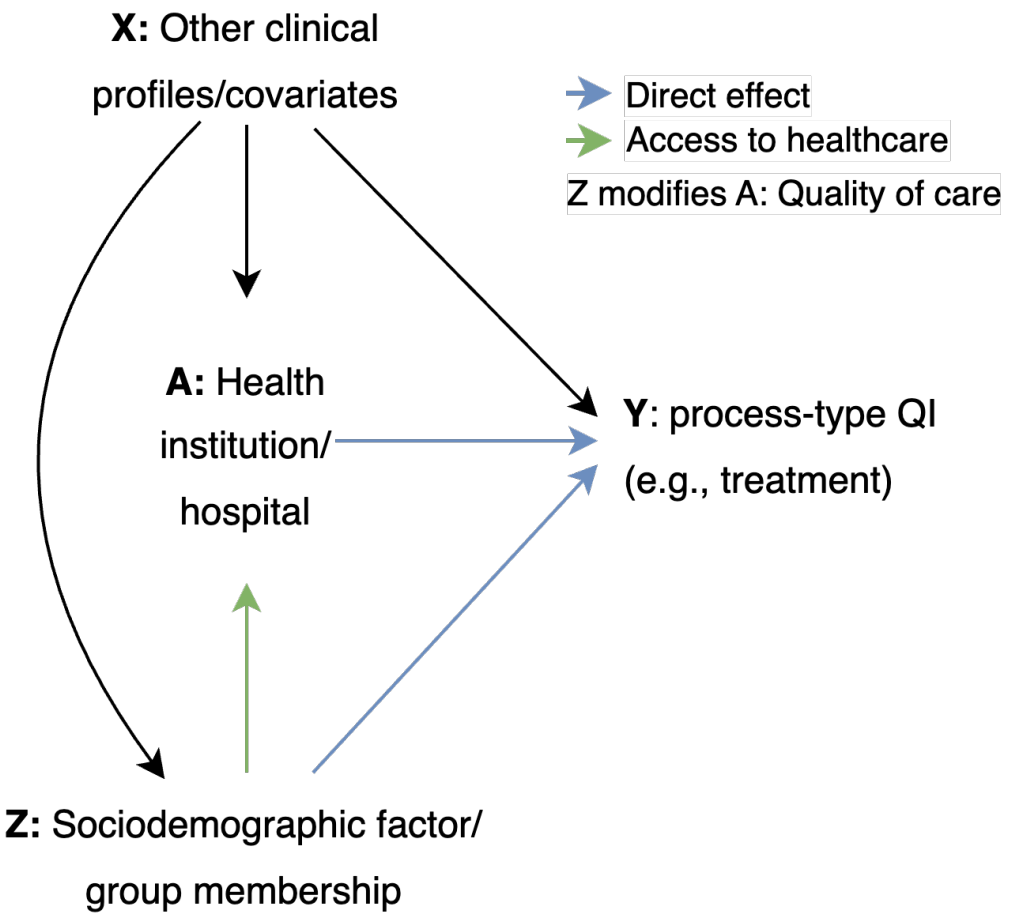

Figure 2: Causal DAG illustrating our research question

Related Literature

Health inequalities are typically assessed by decomposing the observed inequalities in outcome into contributions from different explanatory factors.

Two types of causal decomposition methods:

1) Effect Decomposition(Yu and Elwert 2023):

$E_{Z=z_1}(Y) - E_{Z=z_2}(Y) =$
baseline + hospital + selection + effect modification

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

$V(Y) =$ baseline covariates + hospital quality + residual

Related Literature

Limitations of effect decomposition methods:

- Focus on pairwise comparisons between two groups
- In hospital profiling/comparison, outcomes should capture variation in quality of care across hospitals

This motivates using variance as the key causal quantity for decomposition:

- Handle multi-categorical cases
- Identify quality indicator that capture variation in care

However, to the best of our knowledge, existing variance decomposition method **does not explicitly account for effect modification (within-hospital inequalities) or selection effect (between-hospital inequalities)**. As a result, it cannot be directly used to evaluate health inequalities in care delivery.

Objectives

- Develop a causal variance decomposition framework to quantify effect modification and selection effect in multi-categorical settings
- Construct corresponding estimators for these quantities
- Assess the performance of these estimators through simulation studies

Proposed estimands

The three-way variance decomposition from (Chen et al. 2019) separates the observed variance in care into individual-level case-mix factors, hospital-level, and residual components:

$$V[Y(a) \mid X = x] = V_{Z|x}[E(Y(a) \mid Z, x)] + E_{Z|x}[V(Y(a) \mid Z, x)]$$

where:

$$V(Y(a) \mid Z, x) = V_{A|Z,x}\left\{E[Y(a) \mid A, Z, x]\right\} + E_{A|Z,x}\left\{V[Y(a) \mid A, Z, x]\right\}$$

The three-way decomposition:

$$V[Y(a) \mid X = x] = \underbrace{V_{Z|x}\left\{E_{A|x}[E(Y(a) \mid A, Z, x)]\right\}}_{\text{Variance due to } Z} + \underbrace{E_{Z|x}\left\{V_{A|Z,x}[E(Y(a) \mid A, Z, x)]\right\}}_{\text{Variance due to hospital quality}} + \underbrace{E_{(A,Z)|X}[V(Y(a) \mid A, Z, x)]}_{\text{Residual variance}}$$

Proposed estimands

We propose further isolating **effect modification and selection** from hospital quality variance:

$$
\begin{aligned}
V[Y(a) \mid X] = & \underbrace{V_{Z|x}\left\{E_{A|x}[E(Y(a) \mid A, Z, x)]\right\}}_{\text{Variance explained by } Z} \\
& + \underbrace{V_{A|x}\{E(Y(a) \mid A, x\}}_{\text{Variance causally explained by A}} \\
& + \underbrace{E_{A|(Z,x)}\{V_{Z|x}\left[\tau_a(Z, x)\right]\}}_{\text{variance causally due to effect modification}} \\
& + \underbrace{\sum_a \operatorname{Cov}\left(\left[\tau_a(Z, x)\right]^2, P(A = a \mid Z, x)\right)}_{\text{Variance explained by selection}} \\
& + \underbrace{E_{(A,Z)|x}\left[V(Y(a) \mid A, Z, x)\right]}_{\text{Residual variance}}
\end{aligned}
$$

$\tau_a(Z, x) = E(Y(a) \mid Z, x) - E(Y(\cdot) \mid Z, x)$: the causal effect of visiting hospital a, compared to the "average hospital"

Interpretation

Conditional on x, each term contributes to the variation in the outcome and can be interpreted as follows:

1) Group membership $V_{Z|x}\left\{E_{A|x}[E(Y(a) \mid A, Z, x)]\right\}$

2) Hospital $V_{A|x}\{E(Y(a) \mid A, x\}$

3) Effect modification $E_{A|(Z,x)}\{V_{Z|x}\left[\tau_a(Z, x)\right]\}$

4) Selection $\sum_a \mathrm{Cov}\left(\left[\tau_a(Z, x)\right]^2, P(A = a \mid Z, x)\right)$

5) Residual $E_{(A,Z)|x}[V(Y(a) \mid A, Z, x)]$

Further, to get the marginal variance, we can calculate the variance due to x as:

6) Covariates $V\{E_{(A,Z)}[E(Y(a) \mid A, Z, x)]\}$

$\tau_a(Z, x) = E(Y(a) \mid Z, x) - E(Y(\cdot) \mid Z, x)$: the causal effect of visiting hospital a, compared to the "average hospital"

Idenfitication

Assumptions:
- $A1$ Consistency: $Y = Y(a)$ for $\forall A$ (if received $A = a$, then the observed outcome should be equal to the potential outcome $Y(a)$)
- $A2$ Conditional exchangeability: $Y(a) \perp A \mid Z, X$
- $A3$ Positivity/Overlap: $0 < P(A \mid Z, X) < 1$

If these assumptions hold, we can identify the causal quantities using observed data as follows:

$$V(Y(a) \mid X) \overset{\text{A1}}{=} V(Y \mid X)$$

$$E(Y(a) \mid Z, X) \overset{\text{A2}}{=} E(Y(a) \mid A, Z, X) \overset{\text{A1,A3}}{=} E(Y \mid A = a, Z, X)$$

Estimators

To obtain the causal estimands, we derive model-based estimators using both parametric and non-parametric approaches:

- The outcome model is estimated using generalized linear models, mixed-effects models, random forests, or XGBoost
- The treatment and sociodemographic factor (e.g., race) assignment models are estimated using multinomial logistic regression

Simulation Settings

Data generating mechanism:

- Covariates: $X_1 \sim \text{Bern}(p = 0.5)$, $X_2 \sim N(0, 1)$
- True models:
  - outcome model: logistic regression conditional on A, Z and X
  - treatment assignment model: multinomial logistic regression with Z and X as covariates
  - group membership assignment model: multinomial logistic regression with X as covariates

Scenarios:

- Sample size: 500, 1000, 2500, 5000
- Z: 3 levels
- A: 5 levels

Simulation Results

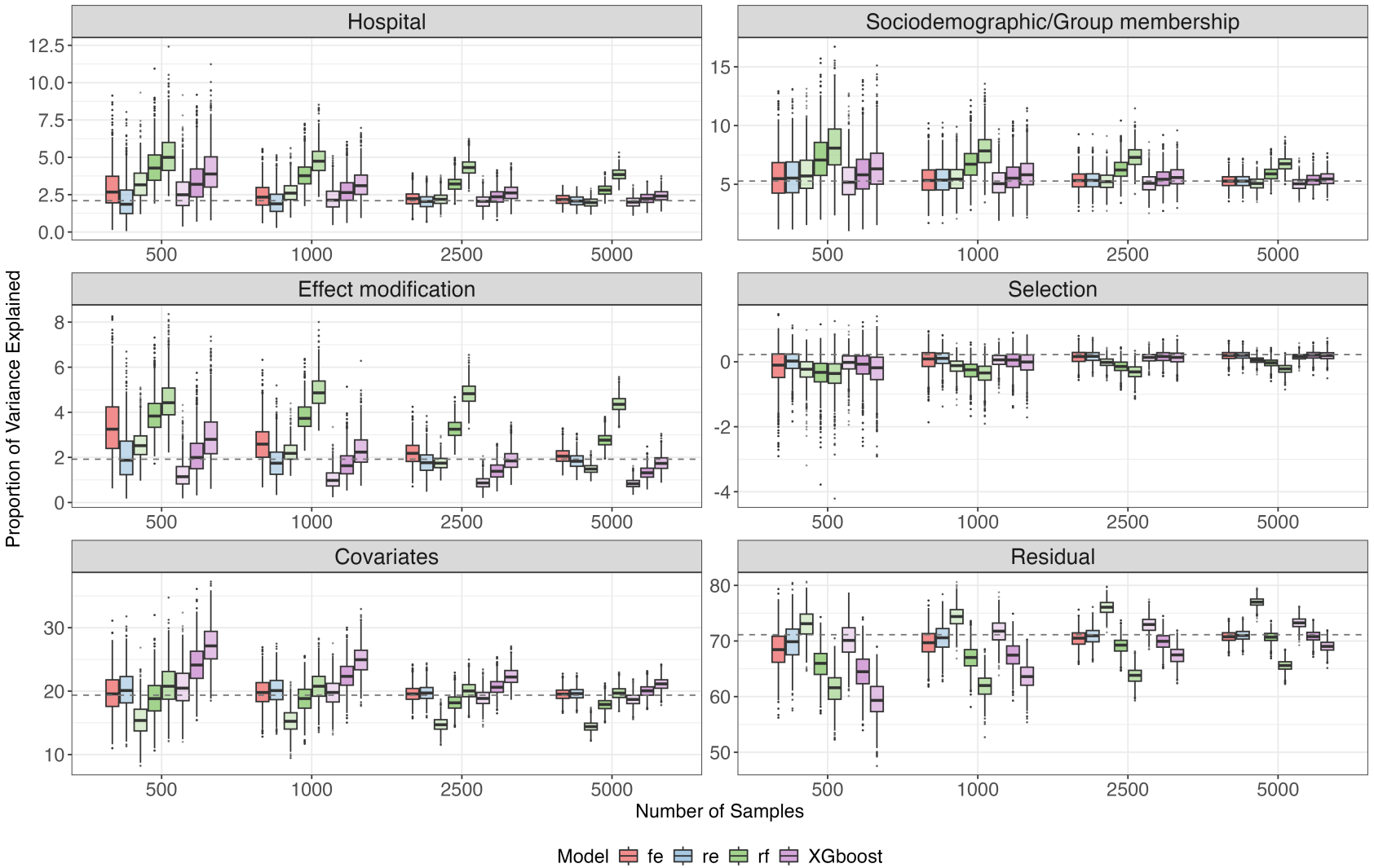

Figure 3: The estimated proportion of variance explained by each component under different sample sizes in 1000 replicates

Conclusions & Next Steps

Conclusions:

- Parametric-based estimators capture the true values reasonably well as the sample size increases (provided the model is 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 continuous outcome

Thanks:)

References

Chen, Bo, Keith A Lawson, Antonio Finelli, and Olli Saarela. 2019. "Causal Variance Decompositions for Institutional Comparisons in Healthcare." *Statistical Methods in Medical Research* 29 (7): 1972–86. https://doi.org/10.1177/0962280219880571.

Gakidou, E. E., C. J. L. Murray, and J. Frenk. 2000. "Defining and Measuring Health Inequality: An Approach Based on the Distribution of Health Expectancy." *Bulletin of the World Health Organization* 78 (1): 42–54. https://scielosp.org/article/bwho/2000.v78n1/42-54/en/.

Lazar, Eliot J., Peter Fleischut, and Brian K. Regan. 2013. "Quality Measurement in Healthcare." *Annual Review of Medicine* 64 (1): 485–96. https://doi.org/10.1146/annurev-med-061511-135544.

Yu, Ang, and Felix Elwert. 2023. "Nonparametric Causal Decomposition of Group Disparities." *arXiv Preprint arXiv:2306.16591*. https://arxiv.org/abs/2306.16591.