Distributed Counterfactual Modeling Approach for Investigating Hospital-Associated Racial Disparities in COVID-19 Mortality

Mackenzie Edmondson[a], Chongliang Luo[a], Nazmul Islam[b], David Asch[c,d], Jiang Bian[e,f], Yong Chen[a]

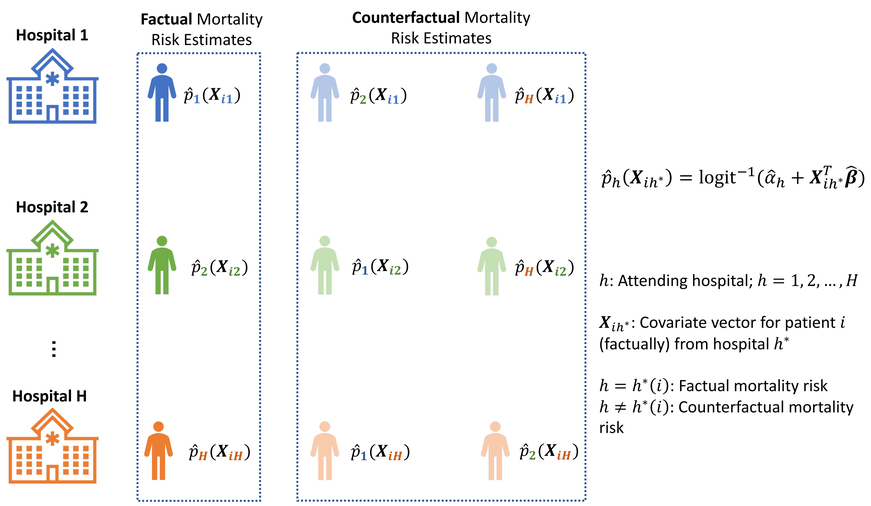
a. Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
b. OptumLabs at UnitedHealth Group, Minnetonka, MN
c. Division of General Internal Medicine, University of Pennsylvania, Philadelphia, PA
d. Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA
e. Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL
f. Cancer Informatics Shared Resource, University of Florida Health Cancer Center, Gainesville, FL

Background

- Several studies have found that black patients are more likely than white patients to test positive for or be hospitalized with COVID-19 but have found no difference in in-hospital mortality.
- Previous studies may have underestimated racial differences due to reliance on data from single hospital system.
- Adequate control of patient-level characteristics requires aggregation of highly granular data from several institutions.
- Want to investigate role of care site in health outcome disparity by race. Patients of different races tend to live in different areas, and sources of care and referral patterns tend to differ.
- **Goal**: Use counterfactual modeling to study potential association between admitting hospital and racial disparity in mortality for COVID-19 patients **without requiring patient-level data sharing**.

Methods

- **Idea**: Fit generalized linear mixed model (GLMM) to model log odds of mortality while adjusting for common patient-level fixed effects as well as hospital-specific random effects.
- **Counterfactual modeling:** Through estimating hospital-specific effects, can estimate patient-specific mortality risk as if patient (counterfactually) attended hospital different from the one truly attended.



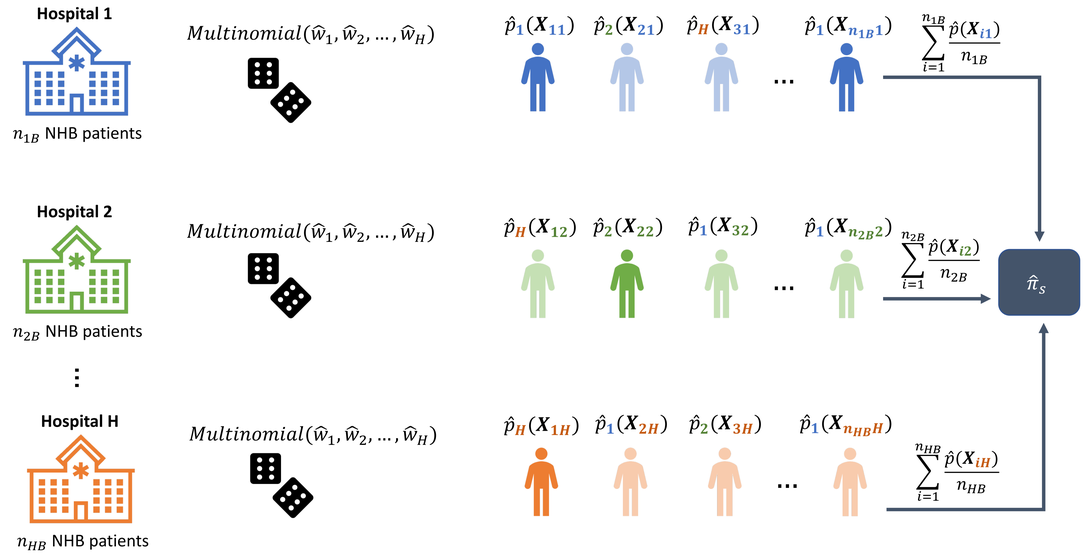
$$\hat{p}_h(\boldsymbol{X}_{ih^*}) = \text{logit}^{-1}(\hat{\alpha}_h + \boldsymbol{X}_{ih^*}^T\hat{\boldsymbol{\beta}})$$

h: Attending hospital; $h = 1, 2, \ldots, H$

$\boldsymbol{X}_{ih^*}$: Covariate vector for patient i (factually) from hospital h^*

$h = h^*(i)$: Factual mortality risk
$h \neq h^*(i)$: Counterfactual mortality risk

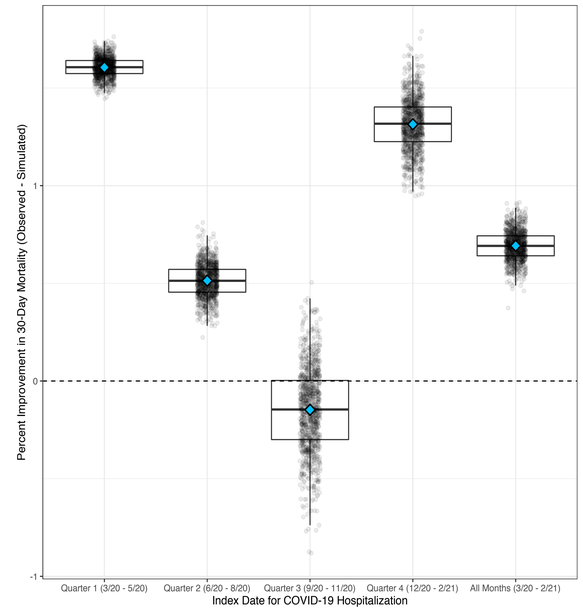
- **Distributed penalized quasi-likelihood (dPQL, Luo et al.)** algorithm distributively fits GLMM using data stored separately at different hospital systems.
- Aggregate, summary-level information shared rather than patient-level data.
- **Simulation** used to estimate racial disparity: produce counterfactual mortality rate estimate for black patients had they attended hospitals in the same distribution as white patients (while retaining sociodemographic/clinical characteristics (see schematic overview of simulation procedure on right).



$\hat{w}_h$: proportion of total NHW patients (factually) at hospital h

Results

- **Proof of concept:** Counterfactual modeling simulation using data from OneFlorida Clinical Research Consortium.
- Modeled in-hospital mortality as function of patient characteristics and estimated hospital-specific random effects (4 hospitals).
- Boxplots display difference between observed (factual) mortality rates and average simulated mortality risk estimates for all Non-Hispanic Black patients.
- Performed analysis stratified by quarter and across all quarters.
- **Results not meant to be clinically interpreted;** intent is to demonstrate utility of this method for performing counterfactual modeling.



Conclusions

- Presented novel application of method for performing distributed generalized linear mixed modeling to study association between admitting hospital and racial disparities in COVID-19 mortality.
- Privacy preserving: requires participating institutions to share only aggregate, summary-level data to perform counterfactual modeling.
- Potential for this method to be used in OHDSI study to explore racial differences in COVID-19 mortality using data from OHDSI network (as well as in other applications of interest), allowing for more generalizable and clinically impactful conclusions.

Contact: mackenzie.edmondson@merck.com