

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

Mackenzie Edmondson, Chongliang Luo, Nazmul Islam, David Asch, Jiang Bian, Yong Chen

Background

Throughout the COVID-19 pandemic, several studies have found that black patients are more likely than white patients to test positive for or be hospitalized with COVID-19 [1-5]. Many of these studies also found that there was no difference in in-hospital mortality for black and white patients after adjusting for patient-level sociodemographic and clinical characteristics [1-2, 4-5]. However, controlling for patient-level characteristics that are largely driven by racial inequities can risk obscuring racial differences in health outcomes rather than identifying them.

A factor thought to contribute to racial disparities in health outcomes is site of care [6]. To investigate the potential association between attending hospital and racial disparity in mortality for COVID-19 patients, a counterfactual modeling approach can be used [7]. A recent study by Asch et al. (2020) explored the potential connection between attending hospital and racial disparities in COVID-19 using counterfactual modeling, fitting a generalized linear mixed model (GLMM) to model log odds of mortality while adjusting for both common patient-level fixed effects as well as hospital-specific fixed and random effects [8]. Estimation of hospital-specific effects is the key component for counterfactual modeling, allowing for estimating patient-specific mortality risk as if the patient (counterfactually) attended a hospital different from the one they truly attended. While effective for this particular study, which featured a centralized data repository allowing direct access to all patient data, GLMM is not able to be used if data are not centralized.

We propose a novel application of a recently developed algorithm for performing GLMM estimation, namely the distributed penalized quasi-likelihood (dPQL) algorithm [9], to study hospital-associated racial disparity in COVID-19 mortality via counterfactual modeling. The penalized quasi-likelihood (PQL) algorithm is an iterative procedure for GLMM estimation [10-11]. The dPQL algorithm is a distributed version of PQL, where in each iteration it requires participating hospitals to share aggregate, summary-level data rather than patient-level data. The result of the dPQL algorithm is lossless, as fixed-effect and random-effect estimates are identical to those produced by traditional PQL as if one had access to centralized patient-level data. After estimating the fixed and random effects via the distributed algorithm, individual hospitals within a multi-site study can further estimate counterfactual mortality risk for each of their own patients, quantifying the effect of their patients receiving care at another hospital in the study. The proposed distributed analysis framework creates the potential for large-scale, multi-site studies for assessing hospital-associated racial disparities when sharing patient-level data is not possible.

Methods

Given the distributively estimated common fixed effects and random effects for hospital h , for patient i who truly attended hospital h , their estimated mortality risk can be calculated using dPQL. We refer to this quantity as the factual mortality risk. Our primary interest for this analysis is to investigate the association between attending hospital and in-hospital mortality differences for black and white patients. We use counterfactual modeling to study whether observed mortality rate for black patients differs from their simulated mortality rate given they (hypothetically) attended hospitals in the same distribution as white patients [7]. Counterfactual mortality risk estimates differ from factual mortality risk estimates in that they are purely hypothetical, with estimates calculated using an estimated random intercept for a hospital that a given patient did not attend. The difference between factual and counterfactual mortality risk estimates and how they are calculated are detailed in Figure 1.

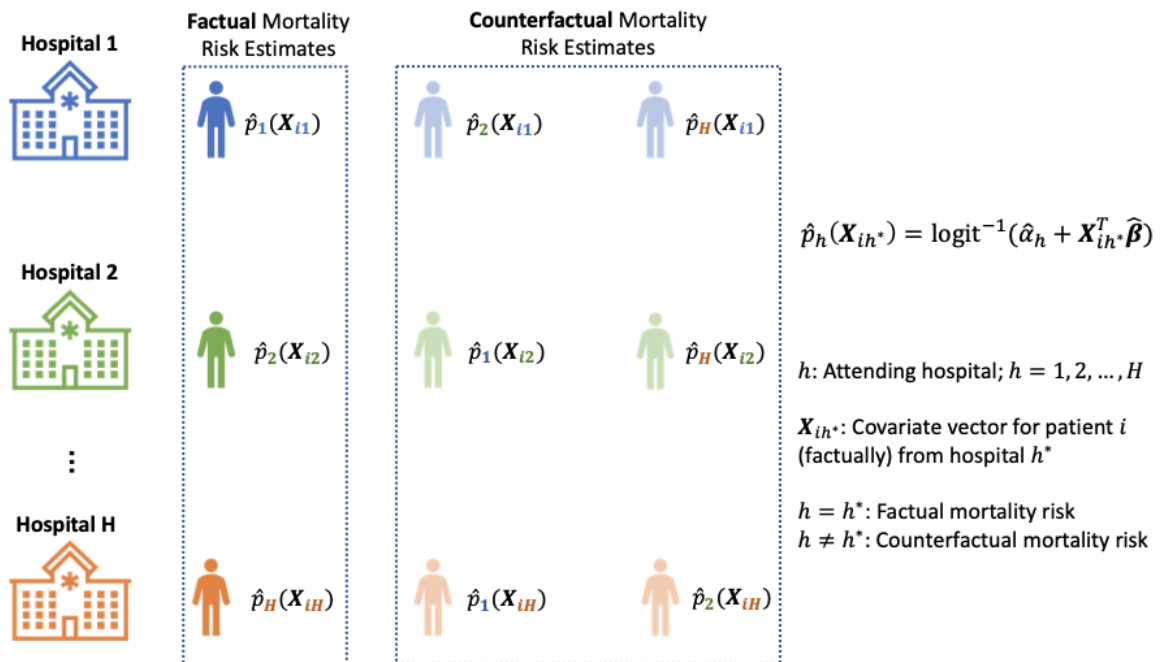


Figure 1. Schematic diagram demonstrating how to calculate both factual and counterfactual mortality risk for patients as if they (hypothetically) attended each hospital. Each row depicts how patient data within a given hospital can be used to estimate mortality risks as if their patients had attended any hospital in the study. The first column features factual mortality risk estimates, calculated using common fixed effect estimates, the random intercept estimate for that given hospital, and the patient data for that hospital. The last two columns provide examples for computable counterfactual mortality risks, which use random intercept estimates for hospitals different from those that patients truly attended.

We quantify racial disparity by conducting a simulation to investigate the counterfactual mortality rate of black patients had they been admitted to hospitals in the same distribution as white patients while retaining their sociodemographic and clinical characteristics. Our simulation procedure is depicted at a high level in Figure 2.

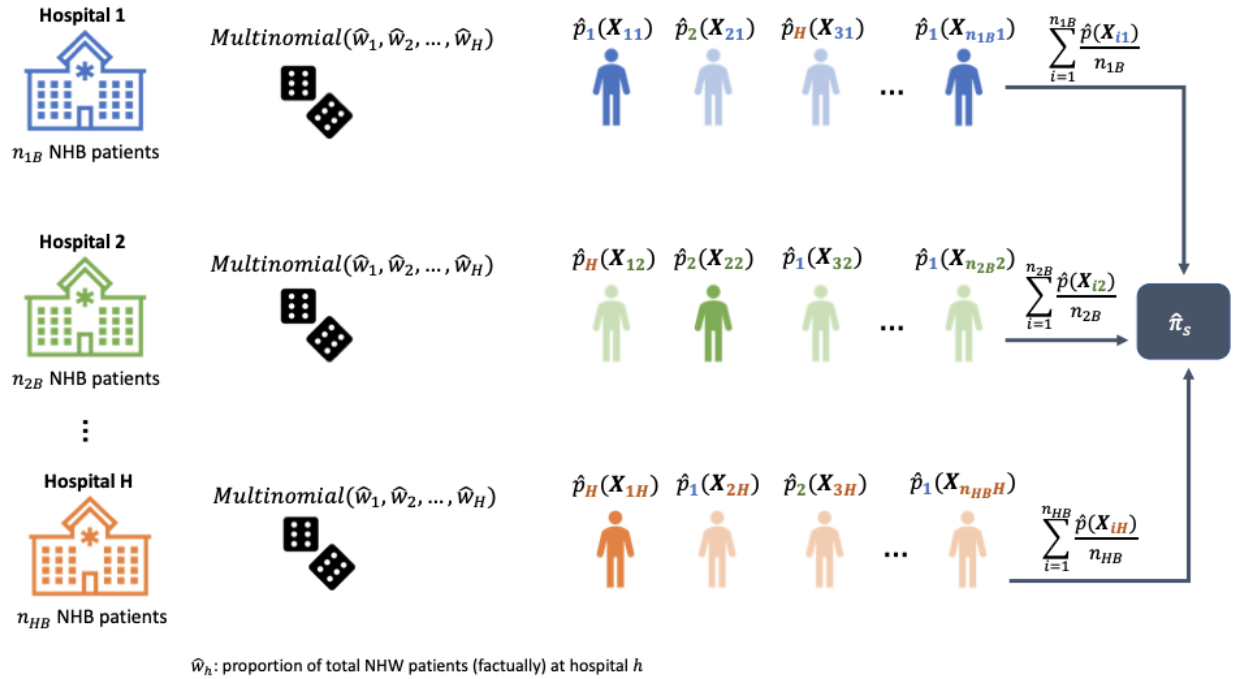


Figure 2. Schematic overview of our simulation procedure for calculating simulated mortality risk estimates using counterfactual modeling. In each iteration of the simulation, within each hospital, a multinomial distribution with probabilities equal to the proportion of total white patients at each hospital (\hat{w}_h) is used to assign each black patient to a hospital. Hospital assignments can be factual (denoted by darker colors) or counterfactual (denoted by lighter colors). Hospital assignments are then used to calculate mortality risk estimates for each black patient as if they attended the assigned hospital. Refer to Figure 1 to see how mortality risk estimates are calculated. Each hospital then averages their patient-level mortality risk estimates and communicates them to the coordinating center, where overall simulated mortality risk estimate $\hat{\pi}_s$ is obtained for each simulation replicate. Multinomial hospital assignment and resultant calculation of (counter)factual mortality risk estimates is performed in each simulation replicate.

As a proof-of-concept analysis, we conducted a counterfactual modeling simulation using patient data from four hospitals in the OneFlorida Clinical Research Consortium. Using the dPQL algorithm, we distributively modeled the log odds of 30-day COVID-19 in-hospital mortality as a function of various patient characteristics, including age, gender, a collection of nine comorbidities, and index quarter, defined as one of four three-month intervals when a given patient was admitted. Our primary analysis concerned calculating the difference between observed mortality rates and average simulated mortality risk estimates for all non-Hispanic black (NHB) patients across the four hospitals. We also performed a series of secondary analyses stratified by index quarter; four sub-analyses were conducted, with each only including patients with an index date for COVID-19 hospitalization in a particular quarter. We conducted 500 iterations of the simulation for each of the five analyses, reporting the average difference in observed mortality rate and mean simulated mortality risk across iterations along with a corresponding 95% empirical confidence interval.

Results

Boxplots summarizing results from each of our simulation studies are presented in Figure 3. The

results presented here are intended to be the product of a proof-of-concept analysis to demonstrate the utility of this simulation method for performing counterfactual modeling. These results therefore should not be interpreted as having clinical significance.

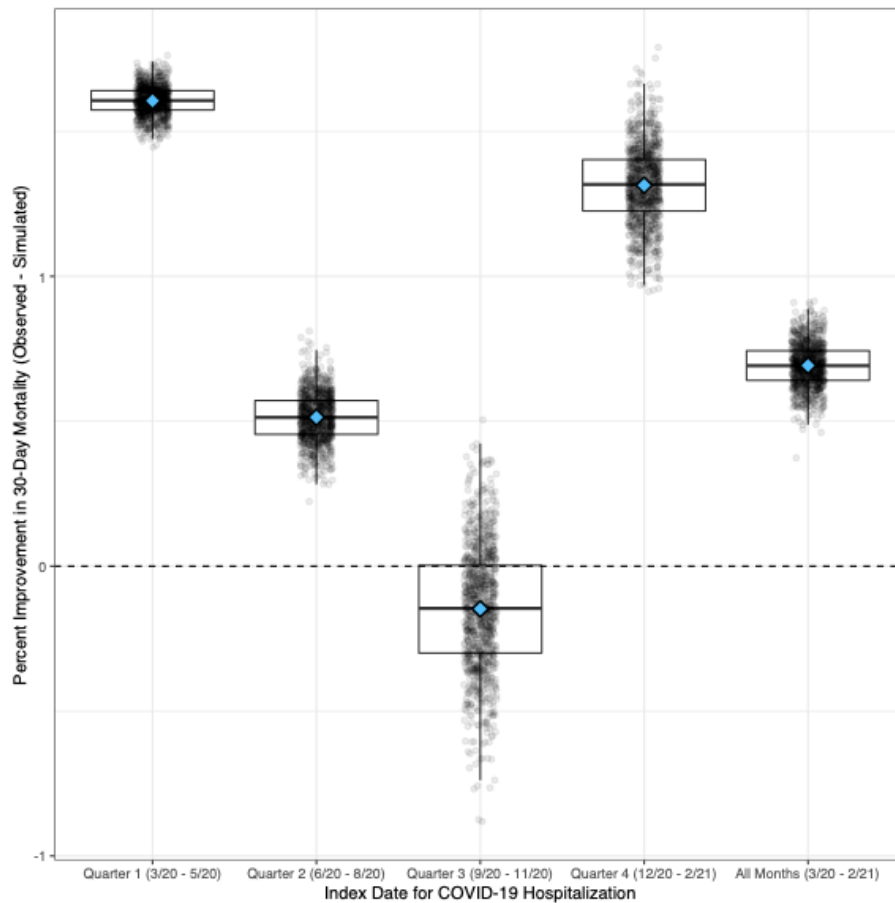


Figure 3. Boxplots depicting results from our five simulation studies, with index date for COVID-19 hospitalization on the x-axis (by index quarter and for all months) and the percent reduction in 30-day in-hospital mortality (observed mortality rate – simulated mortality risk) on the y-axis. For each boxplot, each circle represents the result from one replicate of the simulation. Blue diamonds in the center of each boxplot denote the mean difference across 500 replicates of the simulation.

Conclusion

In this work, we presented a novel application of a method for performing distributed generalized linear mixed modeling, the dPQL algorithm, to study the association between attending hospital and racial disparities in COVID-19 mortality. By not requiring patient-level data sharing, our proposed application could be used to study hospital-associated racial disparities using larger, more heterogeneous collections of patient data, allowing for more generalizable and clinically impactful conclusions. Our counterfactual modeling approach via simulation is also generalizable, able to be used in a variety of applied contexts beyond the application studied here.

References/Citations

1. Price-Haywood, E.G., Burton, J., Fort, D. and Seoane, L., 2020. Hospitalization and mortality among black patients and white patients with Covid-19. *New England Journal of Medicine*, 382(26), pp.2534-2543.
2. Yehia, B.R., Winegar, A., Fogel, R., Fakhri, M., Ottenbacher, A., Jessor, C., Bufalino, A., Huang, R.H. and Cacchione, J., 2020. Association of race with mortality among patients hospitalized with coronavirus disease 2019 (COVID-19) at 92 US hospitals. *JAMA network open*, 3(8), pp.e2018039-e2018039.
3. Muñoz-Price, L.S., Nattinger, A.B., Rivera, F., Hanson, R., Gmehlin, C.G., Perez, A., Singh, S., Buchan, B.W., Ledebore, N.A. and Pezzin, L.E., 2020. Racial disparities in incidence and outcomes among patients with COVID-19. *JAMA network open*, 3(9), pp.e2021892-e2021892.
4. Kabarriti, R., Brodin, N.P., Maron, M.I., Guha, C., Kalnicki, S., Garg, M.K. and Racine, A.D., 2020. Association of race and ethnicity with comorbidities and survival among patients with COVID-19 at an urban medical center in New York. *JAMA network open*, 3(9), pp.e2019795-e2019795.
5. Ogedegbe, G., Ravenell, J., Adhikari, S., Butler, M., Cook, T., Francois, F., Iturrate, E., Jean-Louis, G., Jones, S.A., Onakomaiya, D. and Petrilli, C.M., 2020. Assessment of racial/ethnic disparities in hospitalization and mortality in patients with COVID-19 in New York city. *JAMA network open*, 3(12), pp.e2026881-e2026881.
6. Hasnain-Wynia, R., Baker, D.W., Nerenz, D., Feinglass, J., Beal, A.C., Landrum, M.B., Behal, R. and Weissman, J.S., 2007. Disparities in health care are driven by where minority patients seek care: examination of the hospital quality alliance measures. *Archives of internal medicine*, 167(12), pp.1233-1239.
7. George, E.I., Ročková, V., Rosenbaum, P.R., Satopää, V.A. and Silber, J.H., 2017. Mortality rate estimation and standardization for public reporting: Medicare's hospital compare. *Journal of the American Statistical Association*, 112(519), pp.933-947.
8. Asch DA, Islam MN, Sheils NE, et al. Patient and Hospital Factors Associated With Differences in Mortality Rates Among Black and White US Medicare Beneficiaries Hospitalized With COVID-19 Infection. *JAMA Netw Open*. 2021;4(6):e2112842. doi:10.1001/jamanetworkopen.2021.12842
9. Luo, C., Islam, M.N., Sheils, N.E., Buresh, J. and Chen, Y., 2021. dPQL: a lossless distributed algorithm for generalized linear mixed model with application to privacy-preserving hospital profiling. *medRxiv*.
10. Breslow, N.E. and Clayton, D.G., 1993. Approximate inference in generalized linear mixed

models. *Journal of the American Statistical Association*, 88(421), pp.9-25.

11. Breslow, N.,-- 2004. Whither PQL?. In *Proceedings of the Second Seattle Symposium in Biostatistics* (pp. 1-22). Springer, New York, NY.