Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

American Journal of
Preventive Medicine

RESEARCH BRIEF

Race Versus Social Determinants of Health in COVID-19 Hospitalization Prediction

Carrie R. Howell, PhD,[1] Li Zhang, MS,[2] Nengjun Yi, PhD,[2] Tapan Mehta, PhD,[3] W. Timothy Garvey, MD,[4] Andrea L. Cherrington, MD[1]

Introduction: Including race as a biological construct in risk prediction models may guide clinical decisions in ways that cause harm and widen racial disparities. This study reports on using race versus social determinants of health (SDoH) in predicting the associations between cardiometabolic disease severity (assessed using cardiometabolic disease staging) and COVID-19 hospitalization.

Methods: Electronic medical record data on patients with a positive COVID-19 polymerase chain reaction test in 2020 and a previous encounter in the electronic medical record where cardiometabolic disease staging clinical data (BMI, blood glucose, blood pressure, high-density lipoprotein cholesterol, and triglycerides) were available from 2017 to 2020, were analyzed in 2021. Associations between cardiometabolic disease staging and COVID-19 hospitalization adding race and SDoH (individual and neighborhood level [e.g., Social Vulnerability Index]) in different models were examined. Area under the curve was used to assess predictive performance.

Results: A total of 2,745 patients were included (mean age of 58 years, 59% female, 47% Black). In the cardiometabolic disease staging model, area under the curve was 0.767 vs 0.777 when race was included. Adding SDoH to the cardiometabolic model improved the area under the curve to 0.809 ($p<0.001$), whereas the addition of SDoH and race increased the area under the curve to 0.811. In race-stratified models, the area under the curve for non-Hispanic Blacks was 0.781, whereas the model for non-Hispanic Whites performed better with an area under the curve of 0.821.

Conclusions: Cardiometabolic disease staging was predictive of hospitalization after a positive COVID-19 test. Adding race did not markedly increase the predictive ability; however, adding SDoH to the model improved the area under the curve to ≥0.80. Future research should include SDoH with biological variables in prediction modeling to capture social experience of race.
Am J Prev Med 2022;63(1S1):S103–S108.

INTRODUCTION

Including self-identified race in risk prediction may guide clinical decisions in ways that cause harm and widen racial disparities.[1] Including race as a predictor has become standard, with many prediction models developed without explanation of causal biological mechanisms of racial differences[1–3] and others with reasons that are outdated or fraught with racial bias.[3–5]

From the [1]Division of Preventive Medicine, Department of Medicine, Heersink School of Medicine, The University of Alabama at Birmingham, Birmingham, Alabama; [2]Department of Biostatistics, School of Public Health, The University of Alabama at Birmingham, Birmingham, Alabama; [3]Department of Health Services Administration, School of Health Professions, The University of Alabama at Birmingham, Birmingham, Alabama; and [4]Department of Nutrition Sciences, School of Health Professions, The University of Alabama at Birmingham, Birmingham, Alabama

Address correspondence to: Carrie R. Howell, PhD, Division of Preventive Medicine, Department of Medicine, Heersink School of Medicine, The University of Alabama at Birmingham, Medical Towers 638, 1717 11th Avenue South, Birmingham AL 35205. E-mail: chowell@uabmc.edu.

This article is part of a supplement entitled Obesity-Related Health Disparities: Addressing the Complex Contributors, sponsored by the National Institute on Minority Health and Health Disparities (NIMHD), part of the National Institutes of Health (NIH), an agency of the U.S. Department of Health and Human Services (HHS).

0749-3797/$36.00

https://doi.org/10.1016/j.amepre.2022.01.034

Furthermore, electronic medical record (EMR) data, often the entry point for recording race data in health care,[6] are notoriously characterized by high levels of missingness or unknown race[7] and inaccuracies.[8,9] A growing body of literature points to race/ethnicity as a social construct,[3] conceptualized as a measure that captures a life-long social experience,[6] wherein the causal effect of race on outcomes is assumed through racism and social factors as opposed to biological differences.[2] For these reasons, examining the inclusion of race and social determinants of health (SDoH) in models has become increasingly important, further highlighted by the coronavirus disease 2019 (COVID-19) pandemic.[10]

This holds particularly true when examining obesity-related risk factors and outcomes where racial and low-income disparities have been well described.[11,12] Obesity can exacerbate insulin resistance, progressing to metabolic syndrome and prediabetes, culminating in manifestations of Type 2 diabetes and cardiovascular disease. This one disease process, cardiometabolic disease (CMD), could explain the associations between poor COVID-19 outcomes and obesity, hypertension, diabetes, and cardiovascular disease.[13] This study reports the lessons learned using race versus SDoH measures to account for disparity when predicting the associations between CMD severity, using cardiometabolic disease staging (CMDS),[14–16] and hospitalization after a positive COVID-19 polymerase chain reaction (PCR) test among patients captured in the EMR at an academic medical center in Alabama.

METHODS

This retrospective study included patients in the EMR with a positive PCR test (collected in 2020), as well as a previous encounter where clinical data were available (2017–2020), and were aged ≥35 years (Appendix 1, available online). Data were extracted as part of a larger study (reviewed and approved as exempt by the institution's IRB) examining the associations between retrospective cardiometabolic markers, SDoH, and poor COVID-19 outcomes (hospitalization, intensive care unit stay, mortality). For brevity, only results for the *hospitalization outcome* defined as new hospital admission and a positive COVID-19 PCR within 14 days of admission are reported. *Cardiometabolic severity* was defined using risk factors from the CMDS,[14–16] which included BMI, glucose, blood pressure, high-density lipoprotein cholesterol, and triglycerides. Owing to small group numbers of Hispanics and Asian Americans in the sample, the analysis was limited to non-Hispanic Black and non-Hispanic White. Guided by the National Institute on Minority Health and Health Disparities SDoH Framework,[17] marital and insurance statuses were included as individual SDoH; neighborhood-level SDoH included census tract level Social Vulnerability Index,[18] rurality,[19] and Health Professional Shortage Area[20] designation (Appendix 2, available online). Bayesian logistic regression with Cauchy priors[21] was used to model COVID-19 hospitalization (reported as OR and 95% CI) using CMDS, age, and sex adding race and SDoH in different models. Model 1 included CMDS only, with Model 2 adding race; Model 3 included CMDS and SDoH with Model 4 adding both SDoH and race to CMDS. To evaluate the predictive performance of fitted models, tenfold cross-validation with several measures was used, including area under the curve (AUC) (a measure that reflects the accuracy of the model to correctly classify those with and without the outcome; higher values indicated better performance, whereas a value of 1 means a perfect prediction), mean squared error, and misclassification. AUCs were compared using DeLong's test. Statistical analysis was performed in 2021 using R software (version 4.0.3).

RESULTS

A total of 2,745 patients had complete data for analysis. Overall, patients had a mean age of 58 years (SD=13.2) and were mostly female (59%), and 47% were Black. Table 1 shows the study population characteristics by race category. Black patients had higher proportions of suboptimal SDoH, had higher BMI and clinical parameter values, and were more likely to be hospitalized. Table 2 shows the predictive power and validation for models. In the CMDS only model, AUC was 0.767 vs AUC of 0.777 when race was included (p=0.002). Adding SDoH to the CMDS model improved AUC to 0.809 (p<0.0001), whereas the addition of race to the CMDS and SDoH model increased AUC to 0.811 (p<0.0001). Black race was significant (footnote in Table 2) when added to the CMDS (OR=2.10, 95% CI=1.72, 2.57) but was attenuated and marginally nonsignificant when SDoH were added (OR=1.28, 95% CI=0.99, 1.65). Figure 1 shows the ORs, AUC, mean squared error, and misclassification statistics for models stratified by race. Male sex, higher glucose levels, insurance status, and higher Social Vulnerability Index were associated with increased odds of hospitalization for both Blacks and Whites at mostly comparable magnitudes. AUC for Blacks was 0.781, whereas the model for Whites performed better with an AUC of 0.821 (p=0.02). Although misclassification was higher in the model for Black patients, the false-negative rate (e.g., underestimation) decreased when SDoH were added (Appendix 3, available online).

DISCUSSION

Using CMDS, CMD severity was predictive of hospitalization after a positive COVID-19 test in Black and White patients in the sample. Although adding race to the CMDS model did slightly improve predictive ability, race became nonsignificant when SDoH were accounted for. In addition, adding SDoH to the model improved AUC to >0.80, indicating excellent predictive ability.

Table 1. Comparison of the Characteristics of the Study Population Between Black and White Patients

Characteristics	Non-Hispanic Black, *n*=1,285	Non-Hispanic White, *n*=1,460	*p*-value[a]
Age, mean (SD)	56.8 (12.8)	59.7 (13.4)	**<0.0001**
Sex, *n* (%)			**<0.0001**
Male	434 (33.8)	687 (47.0)	
Female	851 (66.2)	773 (53.0)	
Cardiometabolic parameters			
BMI, mean (SD), kg/m^2	34.0 (7.7)	30.9 (6.7)	**<0.0001**
Plasma glucose, mean (SD), mg/dl	128.5 (48.0)	116.7 (36.2)	**<0.0001**
Systolic blood pressure, mean (SD), mmHg	134.5 (12.1)	129.9 (11.7)	**<0.0001**
Diastolic blood pressure, mean (SD), mmHg	81.4 (6.8)	78.7 (6.5)	**<0.0001**
HDL cholesterol, mean (SD), mg/dl	49.0 (12.6)	48.9 (13.2)	0.86
Triglycerides, mean (SD), mg/dl	128.0 (69.4)	149.8 (80.6)	**<0.0001**
Individual SDoH			
Marital status, *n* (%)			**<0.0001**
Married	510 (39.7)	1,012 (69.3)	
Single	485 (37.7)	174 (11.9)	
Divorced/widowed	290 (22.6)	274 (18.8)	
Insurance status, *n* (%)			**<0.0001**
Private	664 (51.7)	903 (61.9)	
Public	558 (43.4)	531 (36.4)	
None	36 (2.8)	17 (1.2)	
Other	27 (2.1)	9 (0.6)	
Neighborhood SDoH			
Urbanicity, *n* (%)			**<0.0001**
Metropolitan	1,236 (96.2)	1,337 (91.6)	
Micropolitan	32 (2.5)	78 (5.3)	
Rural	6 (0.5)	20 (1.4)	
Small town	11 (0.8)	25 (1.7)	
Social Vulnerability Index, *n* (%)			**<0.0001**
Low	220 (17.1)	841 (57.6)	
Moderate	309 (24.1)	406 (27.8)	
High	756 (58.8)	213 (14.6)	
Healthcare access, *n* (%)			**<0.0001**
Not designated HPSA	546 (42.5)	1,348 (92.3)	
Designated HPSA	739 (57.5)	112 (7.7)	
COVID-19 hospitalization, *n* (%)			**<0.0001**
Yes	494 (38.4)	406 (27.8)	
No	791 (61.6)	1,054 (72.2)	

Note: Boldface indicates statistical significance ($p<0.01$).
HDL, high-density lipoprotein; HPSA, health professional shortage area; SDoH, social determinants of health.
[a]Comparison of characteristics using chi-square tests for categorical variables and *t* tests for continuous variables.

In models stratified by Black and White race, similar SDoH risk factors and magnitudes of risk for hospitalization were found. Although stratified models revealed that cardiometabolic markers and SDoH had better predictive ability in White patients than in Black patients, these results indicate the need to interrogate additional social determinants specific to health inequalities among Black patients. Importantly, the false-negative rate (e.g., underestimation) decreased for Black patients, indicating that SDoH aided in identifying those truly at risk.

These findings are similar to those of a recent study[22] where census tract income but not race was associated with COVID-19 intensive care unit stay after controlling for BMI, age, sex, and comorbidities. Conversely, another study[23] found that both Black race and insurance status were associated with hospitalization early in the pandemic (e.g., March 2020). However, the authors found that the associations did not persist when examining ventilation or mortality outcomes (e.g., worse survival). Instead, they found a greater incidence of COVID-19 among Blacks, likely

Table 2. Predictive Power and Validation for Models With and Without Race Variable Added

Model	AUC	MSE	Misclassification rate
CMDS only	0.767[a]	0.179	0.260
CMDS+race[b]	0.777[c]	0.175	0.258
CMDS+SDoH	0.809[d]	0.162	0.245
CMDS+SDoH+race[e]	0.811[d]	0.162	0.243

Note: CMDS includes BMI, glucose, blood pressure, HDL and triglycerides, age, and sex; SDoH individual level includes marital and insurance status; and SDoH neighborhood level includes census tract Social Vulnerability Index, urbanicity, and Health Professional Shortage Area designation.

AUC, area under the curve; CMDS, cardiometabolic disease staging; HDL, high-density lipoprotein; MSE, mean squared error; ROC, receiver operating characteristic; SDoH, social determinant of health.

[a]Groups with the different superscript letters are significantly different from each other, DeLong's test for 2 correlated ROC curves.

[b]Black race was significant in the model (OR=2.10; 95% CI=1.72, 2.57).

[c]Groups with the different superscript letters are significantly different from each other, DeLong's test for 2 correlated ROC curves.

[d]Groups with the different superscript letters are significantly different from each other, DeLong's test for 2 correlated ROC curves.

[e]Race became nonsignificant in the model once SDoHs were added (OR=1.28; 95% CI=0.99, 1.65).

attributable to social vulnerabilities, which may explain the higher likelihood of hospitalization.[23] Both studies, along with these findings and others,[24] find that social factors rather than race are stronger risks for poor COVID-19 outcomes.

The impact of Black race on hospitalization was attenuated when SDoH were accounted for. This suggests that racial differences in outcome were driven more by social vulnerabilities and social experiences rooted in structural racism as opposed to biological race differences.[2] Of interest, the model limited to Black patients did not perform as well, even when accounting for SDoH. This could be due to unmeasured social factors, specific to Black patients, not captured in this study's models such as perceived discrimination and racism.[25] Future research should focus on including, examining, and reporting on such factors to inform risk prediction modeling.

In exploring the use of race versus SDoHs in COVID-19 outcome prediction models, several key lessons were learned (limitations of the study noted in Appendix 4, available online). First, it is imperative that social factors such as socioeconomic disparities and social experience be

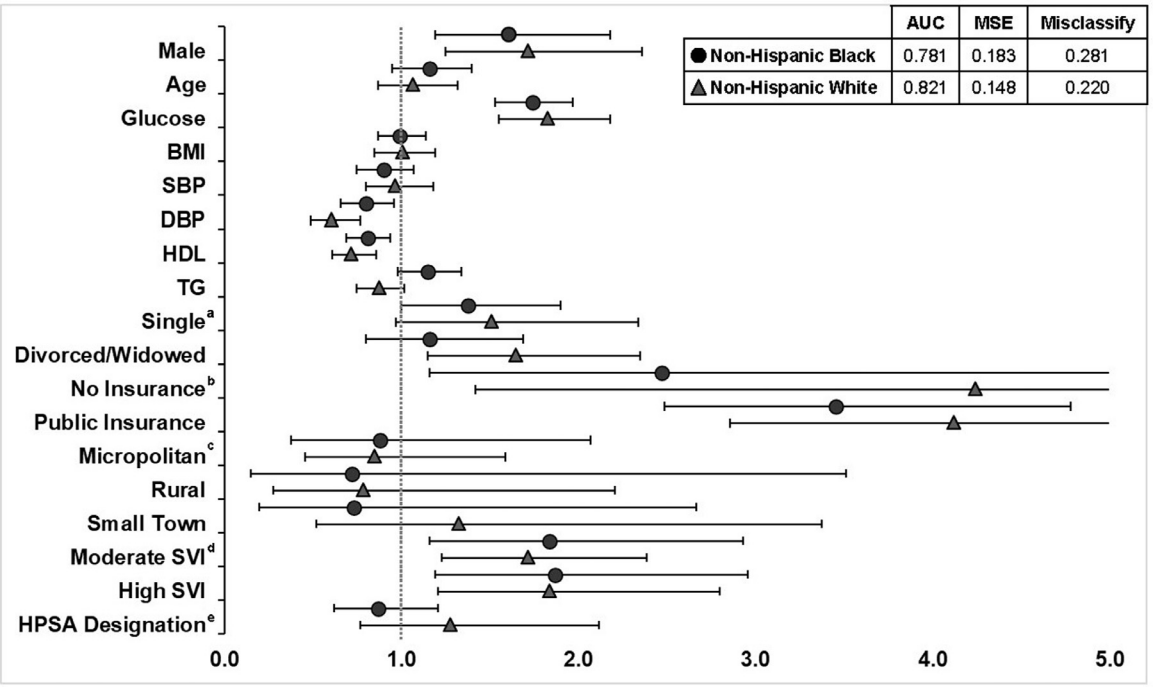

Figure 1. OR plots predicting hospitalization outcome, stratified by Black and White patients.

Note: The points and lines present the estimated values and 95% CIs, respectively. The points and lines present the estimated values and 95% CIs.

[a]Marital status reference category is married.

[b]Insurance status reference category is private insurance.

[c]Urbanicity reference category is metropolitan.

[d]SVI reference category is low SVI.

[e]HPSA reference category is no HPSA designation.

AUC, area under the curve; DBP, diastolic blood pressure; HDL, high-density lipoprotein; HPSA, Health Professional Shortage Area; MSE, mean squared error; SBP, systolic blood pressure; SVI, Social Vulnerability Index; TG, triglyceride.

included in health prediction modeling, instead of solely conceptualizing race through a biological framework.[25] Second, there is a pressing need to explore the risk factors specific to minority and low-income populations. For the EMR, this can be accomplished by assimilating data warehouses that contain rich data on social and behavioral factors in these populations that can be harnessed for prediction models. Finally, researchers and clinicians should question and appreciate the social context that may lead to a patient's poor health outcome as opposed to myopically focusing on biological racial differences.

CONCLUSIONS

Considering SDoH in the context of race in prediction modeling should become standard practice in health research, particularly in the age of precision public health,[26] where healthcare models are moving toward tailoring interventions to a population's specific risk and circumstances.

ACKNOWLEDGMENTS

The authors would like to thank the COVID-19 Collaborative Outcomes Research Enterprise, a collaboration involving the institution's COVID-19 Enterprise Research Initiative, Scientific Community of Outcomes Researchers, and the Center for Clinical and Translational Science, for facilitating the utilization of institutional data to examine population health outcomes.

The National Institute on Minority Health and Health Disparities had no role in study design; collection, analysis, and interpretation of data; writing of the report; or the decision to submit the report for publication.

This work was supported by funding from the National Institute on Minority Health and Health Disparities (U54MD000502).

No financial disclosures were reported by the authors of this paper.

CREDIT AUTHOR STATEMENT

Carrie R. Howell: Conceptualization, Funding acquisition, Methodology, Supervision, Writing - original draft. Li Zhang: Data curation, Formal analysis, Visualization. Nengjun Yi: Formal analysis, Supervision, Writing - review & editing. Tapan Mehta: Formal analysis, Supervision, Writing - review & editing. W. Timothy Garvey: Conceptualization, Supervision, Writing - review & editing. Andrea L. Cherrington: Conceptualization, Methodology, Supervision, Writing - review & editing.

SUPPLEMENTAL MATERIAL

Supplemental materials associated with this article can be found in the online version at https://doi.org/10.1016/j.amepre.2022.01.034.

SUPPLEMENT NOTE

This article is part of a supplement entitled Obesity-Related Health Disparities: Addressing the Complex Contributors, which is sponsored by the National Institute on Minority Health and Health Disparities (NIMHD), National Institutes of Health (NIH), U.S. Department of Health and Human Services (HHS). The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of NIMHD, NIH, or HHS.

REFERENCES

1. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight - reconsidering the use of race correction in clinical algorithms. *N Engl J Med.* 2020;383(9):874–882. https://doi.org/10.1056/NEJMms2004740.
2. Julian BA, Gaston RS, Brown WM, et al. Effect of replacing race with apolipoprotein L1 genotype in calculation of kidney donor risk index. *Am J Transplant.* 2017;17(6):1540–1548. https://doi.org/10.1111/ajt.14113.
3. Kowalsky RH, Rondini AC, Platt SL. The case for removing race from the American Academy of Pediatrics Cnical Practice Guideline for urinary tract infection in infants and young children with fever. *JAMA Pediatr.* 2020;174(3):229–230. https://doi.org/10.1001/jamapediatrics.2019.5242.
4. Eneanya ND, Yang W, Reese PP. Reconsidering the consequences of using race to estimate kidney function. *JAMA.* 2019;322(2):113–114. https://doi.org/10.1001/jama.2019.5774.
5. Vyas DA, Jones DS, Meadows AR, Diouf K, Nour NM, Schantz-Dunn J. Challenging the use of race in the vaginal birth after cesarean section calculator. *Womens Health Issues.* 2019;29(3):201–204. https://doi.org/10.1016/j.whi.2019.04.007.
6. Cooper RS, Nadkarni GN, Ogedegbe G. Race, ancestry, and reporting in medical journals. *JAMA.* 2018;320(15):1531–1532. https://doi.org/10.1001/jama.2018.10960.
7. Lee SJ, Grobe JE, Tiro JA. Assessing race and ethnicity data quality across cancer registries and EMRs in two hospitals. *J Am Med Inform Assoc.* 2016;23(3):627–634. https://doi.org/10.1093/jamia/ocv156.
8. Polubriaginof FCG, Ryan P, Salmasian H, et al. Challenges with quality of race and ethnicity data in observational databases. *J Am Med Inform Assoc.* 2019;26(8–9):730–736. https://doi.org/10.1093/jamia/ocz113.
9. Magaña López M, Bevans M, Wehrlen L, Yang L, Wallen GR. Discrepancies in race and ethnicity documentation: a potential barrier in identifying racial and ethnic disparities. *J Racial Ethn Health Disparities.* 2016;4(5):812–818. https://doi.org/10.1007/s40615-016-0283-3.
10. Patel AGM, Hsiao CJ, Fasanya HO, Stoffel M, Lyon ME. Accuracy in measuring racial disparities during the COVID-19 pandemic needs improvement. *J Appl Lab Med.* 2021;6(1):310–313. https://doi.org/10.1093/jalm/jfaa212.
11. Petersen R, Pan L, Blanck HM. Racial and ethnic disparities in adult obesity in the United States: CDC's tracking to inform state and local action. *Prev Chronic Dis.* 2019;16:E46. https://doi.org/10.5888/pcd16.180579.
12. Thornton PL, Kumanyika SK, Gregg EW, et al. New research directions on disparities in obesity and type 2 diabetes. *Ann N Y Acad Sci.* 2020;1461(1):5–24. https://doi.org/10.1111/nyas.14270.
13. Nadolsky KZ, Hurley DL, Garvey WT. Covid-19 & obesity: beyond Bmi. *Endocr Pract.* 2020;26(8):923–925. https://doi.org/10.4158/EP-2020-0302.
14. Guo F, Garvey WT. Development of a weighted cardiometabolic disease staging (CMDS) system for the prediction of future diabetes. *J Clin Endocrinol Metab.* 2015;100(10):3871–3877. https://doi.org/10.1210/jc.2015-2691.
15. Guo F, Moellering DR, Garvey WT. The progression of cardiometabolic disease: validation of a new cardiometabolic disease staging

system applicable to obesity. *Obesity (Silver Spring).* 2014;22(1):110–118. https://doi.org/10.1002/oby.20585.
16. Wilkinson L, Yi N, Mehta T, Judd S, Garvey WT. Development and validation of a model for predicting incident type 2 diabetes using quantitative clinical data and a Bayesian logistic model: a nationwide cohort and modeling study. *PLoS Med.* 2020;17(8):e1003232. https://doi.org/10.1371/journal.pmed.1003232.
17. Alvidrez J, Castille D, Laude-Sharp M, Rosario A, Tabor D. The National Institute on Minority Health and Health Disparities research framework. *Am J Public Health.* 2019;109(suppl 1):S16–S20. https://doi.org/10.2105/AJPH.2018.304883.
18. Flanagan BE, Hallisey EJ, Adams E, Lavery A. Measuring community vulnerability to natural and anthropogenic hazards: the Centers for Disease Control and Prevention's Social Vulnerability Index. *J Environ Health.* 2018;80(10):34–36.
19. Rural-urban commuting area codes. U.S. Department of Agriculture Economic Research Service. https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx. Accessed June 5, 2018.
20. HPSA data downloads. Health Resource and Service Administration. https://data.hrsa.gov/data/download. Accessed April 30, 2019.
21. Gelman A, Jakulin A, Pittau MG, Su YS. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat.* 2008;2(4):1360–1383. https://doi.org/10.1214/08-AOAS191.
22. Quan D, Luna Wong L, Shallal A, et al. Impact of race and socioeconomic status on outcomes in patients hospitalized with COVID-19. *J Gen Intern Med.* 2021;36(5):1302–1309. https://doi.org/10.1007/s11606-020-06527-1.
23. Muñoz-Price LS, Nattinger AB, Rivera F, et al. Racial disparities in incidence and outcomes among patients with COVID-19. *JAMA Netw Open.* 2020;3(9):e2021892. https://doi.org/10.1001/jamanetworkopen.2020.21892.
24. Kabarriti R, Brodin NP, Maron MI, et al. Association of race and ethnicity with comorbidities and survival among patients with COVID-19 at an urban medical center in New York. *JAMA Netw Open.* 2020;3(9):e2019795. https://doi.org/10.1001/jamanetworkopen.2020.19795.
25. Khazanchi R, Evans CT, Marcelin JR. Racism, not race, drives inequity across the COVID-19 continuum. *JAMA Netw Open.* 2020;3(9):e2019933. https://doi.org/10.1001/jamanetworkopen.2020.19933.
26. Khoury MJ, Iademarco MF, Riley WT. Precision public health for the era of precision medicine. *Am J Prev Med.* 2016;50(3):398–401. https://doi.org/10.1016/j.amepre.2015.08.031.