



Research Paper

Race, ethnicity, community-level socioeconomic factors, and risk of COVID-19 in the United States and the United Kingdom

Chun-Han Lo^{a,b,c,†}, Long H. Nguyen^{a,b,d,†}, David A. Drew^{a,b,†}, Erica T. Warner^{a,e,†}, Amit D. Joshi^{a,b}, Mark S. Graham^f, Adjoa Anyane-Yeboa^b, Fatma M. Shebl^g, Christina M. Astley^{h,i}, Jane C. Figueiredo^j, Chuan-Guo Guo^{a,b,k}, Wenjie Ma^{a,b}, Raaj S. Mehta^{a,b,d}, Sohee Kwon^{a,b}, Mingyang Song^{a,b,c,l}, Richard Davies^m, Joan Capdevila^m, Carole H. Sudre^f, Jonathan Wolf^m, Yvette C. Cozierⁿ, Lynn Rosenbergⁿ, Lynne R. Wilkens^o, Christopher A. Haiman^{p,q}, Loïc Le Marchand^o, Julie R. Palmerⁿ, Tim D. Spector^r, Sebastien Ourselin^f, Claire J. Steves^{r,s}, Andrew T. Chan^{a,b,i,t,*}, on behalf of the COPE Consortium

^a Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, 100 Cambridge Street, 15th Floor, Boston, MA 02114, USA

^b Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

^c Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

^d Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

^e Harvard/MGH Center on Genomics, Vulnerable Populations, And Health Disparities, Massachusetts General Hospital, Boston, MA, USA

^f School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK

^g Medical Practice Evaluation Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

^h Computational Epidemiology Lab and Division of Endocrinology, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA

ⁱ Broad Institute of Harvard and MIT, Cambridge, MA, USA

^j Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles California, USA

^k Department of Medicine, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China

^l Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA

^m Zoe Limited, London, UK

ⁿ Slone Epidemiology Center, Boston University, Boston, MA, USA

^o Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA

^p Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA, USA

^q Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

^r Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

^s Department of Ageing and Health, Guy's and St Thomas's NHS Foundation Trust, London, UK

^t Department of Immunology and Infectious Disease, Harvard T.H. Chan School of Public Health, Boston, MA, USA

ARTICLE INFO

Article History:

Received 5 April 2021

Revised 29 June 2021

Accepted 29 June 2021

Available online xxx

Keywords:

COVID-19

Race

Ethnicity

Socioeconomic factor

Epidemiology

Inequity

ABSTRACT

Background: There is limited prior investigation of the combined influence of personal and community-level socioeconomic factors on racial/ethnic disparities in individual risk of coronavirus disease 2019 (COVID-19).

Methods: We performed a cross-sectional analysis nested within a prospective cohort of 2,102,364 participants from March 29, 2020 in the United States (US) and March 24, 2020 in the United Kingdom (UK) through December 02, 2020 via the COVID Symptom Study smartphone application. We examined the contribution of community-level deprivation using the Neighborhood Deprivation Index (NDI) and the Index of Multiple Deprivation (IMD) to observe racial/ethnic disparities in COVID-19 incidence. ClinicalTrials.gov registration: NCT04331509.

Findings: Compared with non-Hispanic White participants, the risk for a positive COVID-19 test was increased in the US for non-Hispanic Black (multivariable-adjusted odds ratio [OR], 1.32; 95% confidence interval [CI], 1.18–1.47) and Hispanic participants (OR, 1.42; 95% CI, 1.33–1.52) and in the UK for Black (OR, 1.17; 95% CI, 1.02–1.34), South Asian (OR, 1.39; 95% CI, 1.30–1.49), and Middle Eastern participants (OR, 1.38; 95% CI, 1.18–1.61). This elevated risk was associated with living in more deprived communities according to the NDI/IMD. After accounting for downstream mediators of COVID-19 risk, community-level

* Corresponding author at: Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, 100 Cambridge Street, 15th Floor, Boston, MA 02114, USA.

E-mail address: achan@mgh.harvard.edu (A.T. Chan).

† These authors contributed equally to this work.

deprivation still mediated 16.6% and 7.7% of the excess risk in Black compared to White participants in the US and the UK, respectively.

Interpretation: Our results illustrate the critical role of social determinants of health in the disproportionate COVID-19 risk experienced by racial and ethnic minorities.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Research in context

Evidence before this study

Racial and ethnic minorities have disproportionately high rates of SARS-CoV-2 infection, hospitalization, and death related to coronavirus disease 2019 (COVID-19). Prior studies, largely in the US, have examined the association between race/ethnicity and COVID-19 infection. However, these studies have not been able to jointly investigate personal risk factors in the context of community-level socioeconomic risk factors across countries.

Added value of this study

Our analysis concurrently examines personal risk factors, including personal exposure to COVID-19, occupation, and comorbid conditions, and community-level socioeconomic factors on a population-wide scale in the US and the UK. We demonstrate that comorbid conditions do not account for a substantial proportion of the increased likelihood of COVID-19 infection among minority populations. Racial and ethnic minority groups were more likely to live in more deprived communities. Community-level socioeconomic factors mediated a significant portion of the COVID-19 risk in Black participants even after accounting for personal risk factors.

Implications of all the available evidence

The disproportionate impact of COVID-19 experienced by racial and ethnic minority populations highlights the importance of efforts to achieve health equity. Systemic factors are persistent in mediating excess risk among racial and ethnic minorities. Our data support the need for targeted allocation of resources to specific communities to build a more robust and equitable public health infrastructure.

1. Introduction

Racial and ethnic minorities have been disproportionately affected by coronavirus disease 2019 (COVID-19), resulting in higher rates of SARS-CoV-2 infection, hospitalization, and death [1]. These stark disparities underscore fundamental differences in medical, social, economic, and environmental circumstances that predate the current pandemic. One fundamental cause is residential segregation, which reinforces physical isolation of groups from each other and has been deemed a significant contributor to racial and ethnic differences in socioeconomic status, perhaps more so in the United States (US) than in the United Kingdom (UK) [2]. This is especially important in the context of COVID-19, as person-to-person transmission within a community is the primary mode of infection. Specifically, neighborhoods with concentrated poverty and relatively crowded housing are at a higher risk of COVID-19 infection [3,4]. Moreover, minorities with lower levels of educational attainment are more likely to be essential workers unable to fully engage in social isolation [5–7].

There is a pressing need to examine how such social determinants of health jointly contribute to disparities in COVID-19 incidence.

Previous studies investigating racial and ethnic disparities have largely focused on individual socioeconomic risk factors in relation to COVID-19 incidence without considering socioeconomic status as a composite of multiple domains [8,9]. On the other hand, reports utilizing ecologic data on COVID-19 incidence (county- or ZIP Code-level) [8,10,11] have been largely unable to consider individual-level risk factors or include only a single institution or limited geographic regions [12,13]. Data collected from a large sample of participants that jointly examine the contribution of personal and community-level risk factors are needed.

Thus, we conducted a population-scale investigation in the US and the UK to examine (a) the association between race/ethnicity and COVID-19 risk adjusting for personal risk factors; (b) community-level socioeconomic factors; and (c) the degree of COVID-19 risk mediated by such socioeconomic factors. By concurrently studying two countries with a common data collection tool, we were able to examine factors contributing to racial and ethnic disparities in COVID-19 risk in the context of country-specific social determinants of health.

2. Methods

2.1. Real-time assessment of COVID-19 using smartphone technology

We conducted a cross-sectional analysis nested within a prospective cohort of 2,102,364 participants who reported race and ethnicity through the COVID Symptom Study smartphone application since March 29, 2020 in the US and March 24, 2020 in the UK through December 02, 2020 [14]. The app was developed by Zoe Ltd. IT (London, UK) in collaboration with Massachusetts General Hospital (Boston, Massachusetts, USA) and King's College London (London, UK). It offered users a guided interface to report baseline demographic information and comorbidities. Users were prompted to use the application daily to allow for longitudinal, prospective collection of concomitant symptoms, health care visits, and COVID-19 test results. Study participants were recruited through general media and social media outreach, as well as direct invitations from the investigators of long-running prospective cohorts [15]. At enrollment, participants provided electronic informed consent to the use of aggregated information for research purposes and agreed to applicable privacy policies and terms of use. For the current study, participants were followed up for the duration for which they used the app (median: 17 days). This research study was approved by the Partners Human Research Committee (Protocol 2020P000909) and King's College London Ethics Committee (REMAS ID 18210, LRS-19/20-18210). The COVID Symptom Study app is registered with ClinicalTrials.gov, NCT04331509. C-HL, LHN, DAD, and ATC had full access to all the data in the study.

2.2. Assessment of risk factors, symptoms, and testing for COVID-19

Information collected through the application has previously been described [14]. Briefly, at enrollment, participants were asked to provide information on demographic factors and suspected risk factors for COVID-19 (Table 1 and 2). On first use and daily, participants were asked if they felt physically normal, and if not, what symptoms they were experiencing. They were also asked daily if they had been tested for COVID-19 and the results (none, negative, pending,

Table 1
Characteristics of study participants according to race and ethnicity in the United States.

Variables ^b	Race/ethnicity ^a				
	White	Black	Hispanic	Asian	More than one/other race
No. participants	210961	8140	19277	11596	8814
Age, years, median [IQR]	55.0 [37.0, 67.0]	52.0 [34.0, 63.0]	45.0 [32.0, 62.0]	41.0 [25.0, 61.0]	39.0 [23.0, 57.0]
< 25, %	10.4	13.8	26.1	23.1	27.0
25-34, %	11.0	11.4	20.4	16.8	16.2
35-44, %	14.0	13.2	18.6	15.1	16.3
45-54, %	14.5	17.4	13.9	13.4	12.5
55-64, %	18.9	21.2	10.2	9.2	10.8
≥ 65, %	31.1	23.0	10.8	22.4	17.1
Male sex, %	35.9	32.3	39.7	43.4	36.9
Body mass index, kg/m ² , median [IQR]	25.8 [22.7, 30.0]	28.6 [24.5, 33.7]	26.5 [22.8, 31.3]	23.6 [21.1, 26.6]	25.7 [21.9, 30.5]
17-18.4, %	4.1	3.5	5.6	7.1	8.6
18.5-24.9, %	39.7	23.9	34.3	56.4	37.2
25-29.9, %	31.1	30.4	29.6	26.7	27.2
≥ 30, %	25.1	42.2	30.6	9.8	27.0
Comorbidities, %					
Diabetes	4.4	8.8	3.9	5.0	4.6
Heart disease	6.1	5.3	3.4	4.4	5.0
Lung disease or asthma	9.5	10.2	7.3	5.7	10.8
Kidney disease	1.5	2.1	1.0	1.3	1.7
Cancer (active or in the past)	2.1	1.6	1.0	1.3	1.4
Pregnant (females only), %	0.4	0.2	0.4	0.5	0.3
Medication usage, %					
Immunosuppressants	3.8	4.4	2.8	2.3	4.0
Chemo/Immunotherapy	0.5	0.5	0.3	0.4	0.4
ACE inhibitor	9.8	10.6	5.8	5.0	6.5
Aspirin/NSAIDs	17.4	13.1	13.9	5.6	15.3
Current smoker, %	5.8	10.5	8.1	3.8	9.8
Frontline healthcare worker, %	8.7	10.5	9.6	9.9	8.6
Contact with COVID-19 cases in community, %					
Suspected	6.4	6.5	9.2	5.4	9.0
Confirmed	7.1	11.9	15.1	7.6	10.1

Abbreviations: ACE, angiotensin converting enzyme; IQR, interquartile range; NSAIDs, nonsteroidal anti-inflammatory drugs.

^a "Hispanic" was defined as any race of Hispanic or Latino ancestry. Other racial categories were defined as each respective race not of Hispanic or Latino ancestry.

^b Proportions are presented for categorical variables and were calculated based on the total number of participants with available data. Median (IQR) are presented for continuous variables.

positive). To validate our case ascertainment, a subset of individuals who had reported symptoms in the COVID Symptom Study application were invited to provide a copy of the test results. Among 235 participants, we found that self-reported COVID-19 testing yielded a positive predictive value of 77% and a negative predictive value of 97% for confirmed medical record results. Although this was not examined by individual strata, given the simple in-app guidance and the uniform access to the app by our study participants, we did not expect case ascertainment by self-report to vary greatly by race and ethnicity. On first use of the app, participants also reported whether they had ever been exposed to someone with presumed or confirmed COVID-19 infection (such as co-workers, family members, or others). This variable served as an indicator of individual-level exposure to SARS-CoV-2 and had subtle inferences about the socioeconomic environment in which an individual resided.

2.3. Assessment of race and ethnicity

Individuals were asked to report with which race and/or ethnicity they self-identified. Questions were based on standard categories from the National Institutes of Health (race: White, Black, Asian, Native Hawaiian or other Pacific Islander, American Indian or Alaska Native, other, prefer not to say; ethnicity: Hispanic, non-Hispanic) [16] and the Office for National Statistics (White, Black, South Asian, Middle Eastern, Chinese, White and Black, mixed race, other, prefer not to say) [17] (Supplementary Table 1). Individuals who identified their race or ethnicity as "other" were provided an option to enter a free-text description. Those who identified as "Mixed Race" or selected more than one race were described as "more than one race"

and grouped with "other race". In the US, "Hispanic" was defined as any race of Hispanic or Latino ancestry. Other racial categories were defined as each respective race not of Hispanic or Latino ancestry (e.g., non-Hispanic White, non-Hispanic Black). Due to limited sample sizes, "Native Hawaiian or other Pacific Islander" and "American Indian or Alaskan Native" were grouped into "More than one/other race". In the UK, individuals were asked to identify as "Chinese" or "South Asian", where provided examples included "Indian, Pakistani, Bangladeshi, other", but were not specifically asked about other racial identities from Asia. East/Southeast Asian countries such as Japan, the Philippines, and Thailand were the most common "other" free-text responses and were included as "East/Southeast Asian" if they had not previously identified as an existing category.

We excluded participants who did not provide information on racial or ethnic identity or selected "prefer not to say". Country-specific percentages of racial groups in the study are provided in Supplementary Table 1 along with those in the general population as estimated by the US Census Bureau Population Estimates Program (2019) [18] and the England and Wales 2011 Census by the Office for National Statistics [19].

2.4. Assessment of community-level sociodemographic characteristics in the US

Socioeconomic variables at the ZIP Code Tabulation Area (ZCTA) level were collected from the US Census Bureau's 2014-2018 American Community Survey [20]. Percentage of essential workers was calculated as the proportion of individuals employed in 14 major occupational categories, including "healthcare practitioners and

Table 2
Characteristics of study participants according to race and ethnicity in the United Kingdom.

Variables ^a	Race/ethnicity						
	White	Black	South Asian	Middle Eastern	Chinese	East/Southeast Asian	More than one/other race
No. participants	1736547	10949	37638	6828	5762	1713	44139
Age, years, median [IQR]	48.0 [31.0, 61.0]	44.0 [31.0, 55.0]	41.0 [30.0, 52.0]	41.0 [30.0, 55.0]	40.0 [28.8, 51.0]	44.0 [35.0, 53.0]	31.0 [15.0, 48.0]
< 25, %	17.6	16.5	18.6	16.2	18.3	10.0	40.5
25–34, %	11.4	14.5	15.7	16.9	18.3	14.1	14.3
35–44, %	14.8	19.6	25.2	24.2	24.6	27.2	15.3
45–54, %	18.7	23.7	19.6	17.6	18.8	27.7	13.3
55–64, %	18.6	19.1	11.4	13.3	10.8	12.3	9.9
≥ 65, %	18.8	6.6	9.5	11.7	9.3	8.6	6.7
Male sex, %	44.0	49.4	47.9	54.0	38.8	28.4	44.7
Body mass index, kg/m ² , median [IQR]	25.3 [22.1, 29.3]	26.9 [23.1, 31.3]	24.6 [21.5, 28.0]	25.3 [22.2, 29.1]	22.5 [20.2, 25.4]	23.1 [20.5, 26.2]	23.5 [19.7, 27.9]
17–18.4, %	8.5	8.1	9.8	7.1	12.1	9.1	19.3
18.5–24.9, %	39.6	29.4	44.6	41.2	60.6	58.3	40.8
25–29.9, %	30.1	31.2	29.4	31.0	19.6	23.6	21.9
≥ 30, %	21.8	31.3	16.1	20.8	7.7	8.9	18.0
Comorbidities, %							
Diabetes	3.0	6.0	6.2	4.3	2.4	3.1	2.2
Heart disease	3.3	2.4	3.5	3.8	1.8	1.9	1.9
Lung disease or asthma	8.6	9.1	7.9	6.6	5.7	6.2	8.7
Kidney disease	0.8	1.3	1.0	1.1	0.7	0.2	0.7
Cancer (active or in the past)	1.4	1.2	0.8	1.1	0.7	1.2	0.8
Pregnant (females only), %	0.6	0.7	1.0	0.8	0.8	0.8	0.7
Medication usage, %							
Immunosuppressants	3.6	4.1	3.5	3.0	2.3	2.5	3.1
Chemo/Immunotherapy	0.3	0.3	0.2	0.2	0.2	0.4	0.2
ACE inhibitor	7.3	6.1	5.5	5.0	2.9	3.5	3.5
Aspirin/NSAIDs	6.7	7.1	4.4	6.0	2.8	4.0	5.5
Current smoker, %	3.0	3.9	3.4	5.8	2.4	2.8	4.8
Frontline healthcare worker, %	5.6	13.4	10.1	7.1	6.4	14.1	5.1
Contact with COVID-19 cases in community, %							
Suspected	9.5	11.4	9.6	10.5	8.2	11.7	12.7
Confirmed	4.7	8.6	7.9	7.0	6.7	10.4	6.1

Abbreviations: ACE, angiotensin converting enzyme; IQR, interquartile range; NSAIDs, nonsteroidal anti-inflammatory drugs.

^a Proportions are presented for categorical variables and were calculated based on the total number of participants with available data. Median (IQR) are presented for continuous variables.

technical occupations”, “farming, fishing, and forestry occupation”, and “transportation occupation”, among others, consistent with a report from the American Civil Liberties Union of Massachusetts [37].

2.5. Creation of the Neighborhood Deprivation Index in the US

Details about the Neighborhood Deprivation Index (NDI) have been described previously [23]. Briefly, based on a review of literature, we identified 25 census variables that have been used consistently to approximate neighborhood-level environments for possible inclusion in the deprivation index. We used principal component analysis (PCA) for census data reduction and retained the first principal component. Variables were then assessed for inclusion based on two *a priori* criteria: First, variables that had a loading above 0.25 were included in the index. Second, we stipulated that the lower 95% confidence limit of the variable loading could not be below 0.68, which was chosen because it was the lower 95% confidence limit for the median variable loading. Of the 25 variables included in the PCA, seven variables were retained for the index (Supplementary Table 2). Ordination was then repeated using only these census variables to obtain the final loadings. The NDI was standardized to have a mean of 0 and a standard deviation of 1 by dividing the index by the square of the eigenvalue.

2.6. Collection of Index of Multiple Deprivation in the UK

For the UK, we collected the Index of Multiple Deprivation (IMD) for each Lower Layer Super Output Area (LSOA) through the Office for National Statistics (England) [24], the Welsh Government (Wales) [25], the Scottish Government (Scotland) [26], and the Northern

Ireland Statistics and Research Agency (Northern Ireland) [27]. We combined scores from each of the UK's four constituent countries into a unified scale by assuming the same distribution of deprivation scores in each country. Individual components of the IMD include income, education, employment, barriers to housing and services, living environment, health, and crime (Supplementary Table 2). To provide consistency in the presentation of the directionality of deprivation between the NDI and the IMD, the order of the IMD scale was reversed such that the highest quintile in both the NDI and the IMD reflected the “most deprived” category.

2.7. Statistical analysis

We performed logistic regression analysis conditioned on age, sex, and date of study entry to estimate the odds ratios (ORs) and 95% confidence intervals (CIs) for testing positive for COVID-19 throughout follow-up according to race/ethnicity. Additional covariates included in the multivariable models were selected *a priori* as putative risk factors for COVID-19 and grouped into comorbidities, occupation as frontline healthcare worker (HCW), occupational risk factors, and personal contact with COVID-19 in the community. We applied inverse probability weighting (IPW). Each participant was weighted according to the inverse probability of receiving a test for COVID-19 calculated as a function of age, sex, date of study entry, race/ethnicity, symptoms, and occupation as frontline HCW. Because we previously found a higher risk of testing positive for COVID-19 in the healthcare setting [28], we also examined the race/ethnicity-COVID-19 relationship within frontline HCWs. Given the dynamic nature of the pandemic, we further performed an analysis stratified by time period (first wave [March 29–June 15, 2020], second wave

Table 3
Risk of COVID-19 according to race and ethnicity in the United States

	Race/ethnicity ^a				
	White	Black	Hispanic	Asian	More than one/other race
Overall participants^b					
No. of cases	6972	413	1352	258	362
No. of participants	210961	8140	19277	11596	8814
Age-adj OR (95% CI) ^c	1.0 (reference)	1.52 (1.36-1.71)	1.73 (1.62-1.85)	0.80 (0.69-0.92)	1.05 (0.93-1.19)
Comorbidity-adj OR (95% CI) ^d	1.0 (reference)	1.49 (1.33-1.67)	1.70 (1.59-1.82)	0.82 (0.71-0.94)	1.06 (0.94-1.19)
Comorbidity + occupation-adj OR (95% CI) ^e	1.0 (reference)	1.47 (1.31-1.65)	1.70 (1.58-1.81)	0.80 (0.70-0.92)	1.06 (0.94-1.19)
Comorbidity + occupation + personal contact with COVID-19-adj OR (95% CI) ^f	1.0 (reference)	1.32 (1.18-1.47)	1.42 (1.33-1.52)	0.88 (0.76-1.01)	1.00 (0.89-1.13)
Healthcare workers^b					
No. of cases	1381	79	201	62	76
No. of participants	18343	856	1849	1146	759
Age-adj OR (95% CI) ^c	1.0 (reference)	1.51 (1.14-2.00)	1.26 (1.04-1.51)	0.84 (0.62-1.12)	1.16 (0.88-1.52)
Comorbidity-adj OR (95% CI) ^d	1.0 (reference)	1.44 (1.09-1.90)	1.24 (1.03-1.49)	0.87 (0.65-1.16)	1.16 (0.88-1.53)
Comorbidity + occupational risk factors-adj OR (95% CI) ^e	1.0 (reference)	1.47 (1.12-1.93)	1.27 (1.05-1.53)	0.85 (0.63-1.14)	1.16 (0.88-1.52)
Comorbidity + occupational risk factors + personal contact with COVID-19-adj OR (95% CI) ^f	1.0 (reference)	1.38 (1.05-1.82)	1.23 (1.02-1.49)	0.85 (0.63-1.14)	1.13 (0.85-1.49)

Abbreviation: adj, adjusted; CI, confidence interval; OR, odds ratio.

^a In the United States, "Hispanic" was defined as any race of Hispanic or Latino ancestry. Other racial categories were defined as each respective race not of Hispanic or Latino ancestry.

^b All models were weighted according to the inverse probability of testing for COVID-19 calculated as a function of age, sex, date of study entry, race/ethnicity, symptoms (fatigue, headache, sore throat, chest pain, shortness of breath, persistent cough, diarrhea, abdominal pain, skipped meals/anorexia, hoarse voice, myalgias, delirium, loss of smell/taste, fever), and occupation as frontline healthcare worker (among overall participants).

^c Logistic regression model conditioned on age, sex, and date of study entry.

^d Additionally adjusted for body mass index (17-18.4, 18.5-24.9, 25-29.9, and ≥ 30 kg/m²), history of diabetes (no, yes), heart disease (no, yes), lung disease or asthma (no, yes), kidney disease (no, yes), cancer (active or in the past; no, yes), and smoking status (never/former smokers, current smokers).

^e Additionally adjusted for occupation as frontline healthcare worker (no, yes; among overall participants). For healthcare workers, the model was additionally adjusted for access to personal protective equipment (reuse or inadequate, adequate) and practice setting (inpatient, nursing homes, outpatient hospital clinics, home health sites, ambulatory clinics, other).

^f Additionally adjusted for personal contact with COVID-19 (no, suspected COVID-19, confirmed COVID-19).

[June 16-September 15, 2020], third wave [September 16-December 02, 2020] in the US; first wave [March 24-August 31, 2020], second wave [September 01-December 02, 2020] in the UK).

We then examined the association between race/ethnicity, community-level socioeconomic factors, and risk of COVID-19 using univariate logistic regression analyses conditioned on age, sex, and date of study entry. Community-level socioeconomic factors were assessed through the NDI in the US and the IMD in the UK as well individual components of the NDI and the IMD.

Lastly, structural equation models were implemented to conduct a mediation analysis using the "lavaan" package in R. Potential mediators, from upstream to downstream, included community-level socioeconomic factors, personal contact with COVID-19, occupation, and comorbidities (Fig. 4). We computed the proportion of total effect that was explained by indirect effects of each of the mediators to estimate the relative contributions to the association between race/ethnicity and COVID-19.

We conducted all analyses using R 3.6.1 (Vienna, Austria). All statistical tests were two-sided with a *P* value less than 0.05 indicating statistical significance.

2.8. Role of funding source

The sponsors had no role in study design, analysis, and interpretation of data, report writing, and the decision to submit for publication.

3. Results

3.1. Study population

In the US, 210,961 participants identified as non-Hispanic White (herein described as White), 8,140 as non-Hispanic Black (Black), 19,277 as Hispanic/Latinx (Hispanic), 11,596 as non-Hispanic Asian (Asian), and 8,814 as more than one/other race (Table 1). In the UK,

1,736,547 participants identified as White, 10,949 as Black, 37,638 as South Asian, 6,828 as Middle Eastern, 5,762 as Chinese, 1,713 as East/Southeast Asian, and 44,139 as more than one/other race (Table 2). The median age was 52 years (interquartile range [IQR] 35-66) in the US and 48 years (IQR 31-61) in the UK. In both countries, Black participants were more likely to have obesity, diabetes, and kidney disease compared to other participants. Racial minorities were also more likely to work as frontline HCWs and have personal contact with COVID-19 compared to White participants.

3.2. Risk of COVID-19 among US racial and ethnic minorities

In the US, compared to White participants, Black and Hispanic participants had an increased risk of reporting a positive COVID-19 test (age-adjusted OR [95% CI], 1.52 [1.36-1.71] for Black participants and 1.73 [1.62-1.85] for Hispanic participants) (Table 3). These ORs were not materially altered after additional adjustment for comorbidities and occupation as frontline HCW. In analyses restricted to HCWs, we observed similar increases in risk among Black and Hispanic HCWs compared to White HCWs. For the overall population, additional adjustment for contact in the community with a person with suspected or confirmed COVID-19 attenuated the risk estimates. Notably, in a stratified analysis, the increased risk of COVID-19 experienced by Black and Hispanic participants was more pronounced in the first wave, followed by the second wave in the US (Supplementary Table 3).

We hypothesized that, since personal contact with COVID-19 appeared to be a strong mediator of the risk of COVID-19 among racial minorities, community-level socioeconomic factors may mediate the association between race/ethnicity and risk of COVID-19. Overall, participants living in communities in the highest quintile of the NDI (most deprived) had an OR of 1.24 (95% CI, 1.20-1.28; $P_{\text{trend}} < 0.001$) for personal contact with COVID-19 and 1.71 (95% CI, 1.60-1.84; $P_{\text{trend}} < 0.001$) for testing positive for COVID-19 compared to those in the lowest quintile (least deprived) (Table 5).

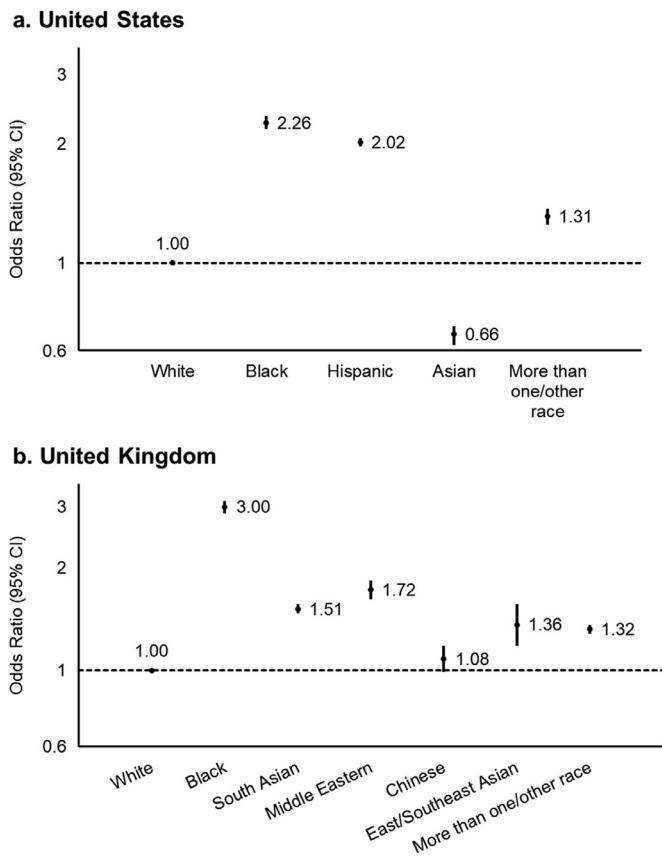


Fig. 1. Risk of living in a community within the highest quintile of community-level deprivation according to race and ethnicity. Data points represent the odds ratios with 95% confidence intervals. In both countries, White participants were used as the reference group. “Hispanic” in the United States was defined as any race of Hispanic or Latino ancestry, while other racial categories were defined as each respective race not of Hispanic or Latino ancestry. Community-level deprivation was represented by the Neighborhood Deprivation Index in the United States and the Index of Multiple Deprivation in the United Kingdom. Higher scores represented more deprived communities.

Similar increases in risk were observed for *a priori* categories of each individual domain of the NDI (**Supplementary Fig. 1**). The OR of living in communities in the highest quintile of the NDI was 2.26 (95% CI, 2.18–2.35) for Black participants, 2.02 (95% CI, 1.97–2.07) for Hispanic participants, 0.66 (95% CI, 0.62–0.69) for Asian participants, and 1.31 (95% CI, 1.25–1.37) for participants of more than one/other race compared to White participants (**Fig. 1**). This broadly corresponded to the findings related to individual NDI domains shown in **Fig. 2**, where Black and Hispanic participants were consistently more likely to live in more deprived communities, while Asian participants were at a lower risk.

Next, in a structural equation model, we observed that in the US, the total proportion of increased COVID-19 risk in Black participants compared to White participants mediated by community-level deprivation was 24% (**Table 6** and **Fig. 4**). When considering potential downstream mediators of community risk, including contact with community members with COVID-19, occupation as frontline HCW, and comorbidities, we found that 16.6% of COVID-19 risk remained mediated by community-level deprivation. The corresponding proportion of additional risk mediated was 9.2% for Hispanic participants and 5.9% after accounting for potential downstream mediators of COVID-19 risk.

3.3. Risk of COVID-19 among UK racial and ethnic minorities

In the UK, after accounting for the likelihood of receiving a test, an increased risk of reporting a positive COVID-19 test was observed

among Black, South Asian, and Middle Eastern participants (age-adjusted ORs ranging from 1.25 to 1.88) (**Table 4**). As with the US participants, we found minimal attenuation of the risk estimates after accounting for comorbidities. Among frontline HCWs, we observed a similar increase in risk among South Asian participants. Additional adjustment for personal contact with COVID-19 attenuated the risk estimates for overall UK participants. Similar to the US, we observed greater racial/ethnic disparities in COVID-19 risk in the first wave compared to the second wave of the pandemic in the UK (**Supplementary Table 4**).

We assigned each UK app user an Index of Multiple Deprivation (IMD) [24]. Individual domains and their distribution across quintiles of the IMD are shown in **Supplementary Table 2**. Participants living in communities in the highest quintile of the IMD (most deprived; see *Methods*) had an OR of 1.11 (95% CI, 1.10–1.12; $P_{\text{trend}} < 0.001$) for contact with someone with suspected or confirmed COVID-19 and 1.26 (95% CI, 1.22–1.29; $P_{\text{trend}} < 0.001$) for testing positive for COVID-19 compared to the lowest quintile (least deprived) (**Table 5**). Each domain comprising the IMD except for housing was associated with an increased risk of personal contact with COVID-19 and testing positive for COVID-19 (**Supplementary Fig. 2**). Black participants, in particular, were more likely to live in socioeconomically deprived communities as defined by the IMD (OR, 3.00; 95% CI, 2.88–3.12) (**Fig. 1**) and individual IMD domains compared to White participants (**Fig. 3**).

In the UK, using a structural equation model, the proportion of increased COVID-19 risk compared to White participants mediated by community-level deprivation was higher in Black participants (18.2%) than in South Asian (4.0%) and Middle Eastern participants (4.3%) despite all three racial groups showing an excess risk of testing positive for COVID-19 (**Table 6** and **Fig. 4**). Additionally, considering potential downstream mediators of community risk resulted in a remaining 7.7% of excess risk mediated by community-level deprivation in Black participants.

4. Discussion

Among over two million individuals, racial and ethnic minorities, particularly Black and Hispanic participants in the US and Black, South Asian, and Middle Eastern participants in the UK, experienced a greater risk of COVID-19 compared to White participants, especially during the first wave of the pandemic. These minority groups experienced a higher risk of living in more socioeconomically deprived communities; personal contact with COVID-19 in the community was a significant mediating factor for the association between race/ethnicity and COVID-19 risk. A substantial proportion of COVID-19 risk was mediated by community-level deprivation, especially for Black participants. Similar findings were observed in the US and the UK, providing evidence that country-specific social determinants of health observed among minority communities are likely significant risk factors for COVID-19.

These findings are consistent with prior studies of racial and ethnic disparities in COVID-19 risk. Prior studies using ecological data found attenuated associations between race/ethnicity and COVID-19 infection after adjusting for community-level factors [8,10,11]. However, these studies did not examine individual risk in the context of community-level impacts [12,13]. A recent study of 18,917 participants in the UK Biobank showed that ethnic minority background and community-level socioeconomic status assessed by the Townsend Deprivation Index are important COVID-19 risk factors among individuals with cancer [29]. Our analysis provides robust results that address this gap by concurrently examining personal exposure to COVID-19, occupation, and comorbid conditions, and community-level socioeconomic factors on a population scale using participant information not commonly available in registry or hospital-based cohorts. Our results demonstrate that comorbid conditions do not explain a substantial proportion of the increased likelihood of

Table 4
Risk of COVID-19 according to race and ethnicity in the United Kingdom.

	Race/ethnicity						
	White	Black	South Asian	Middle Eastern	Chinese	East/Southeast Asian	More than one/other race
Overall participants^a							
No. of cases	39594	323	1247	229	113	51	1071
No. of participants	1736547	10949	37638	6828	5762	1713	44139
Age-adj OR (95% CI) ^b	1.0 (reference)	1.41 (1.23-1.61)	1.56 (1.46-1.67)	1.54 (1.32-1.79)	0.85 (0.68-1.06)	1.21 (0.87-1.67)	1.05 (0.98-1.13)
Comorbidity-adj OR (95% CI) ^c	1.0 (reference)	1.40 (1.23-1.60)	1.57 (1.47-1.68)	1.53 (1.31-1.79)	0.86 (0.69-1.08)	1.23 (0.89-1.70)	1.06 (0.99-1.15)
Comorbidity + occupation-adj OR (95% CI) ^d	1.0 (reference)	1.28 (1.12-1.46)	1.48 (1.38-1.59)	1.50 (1.29-1.75)	0.86 (0.69-1.08)	1.12 (0.81-1.55)	1.06 (0.99-1.15)
Comorbidity + occupation + personal contact with COVID-19-adj OR (95% CI) ^e	1.0 (reference)	1.17 (1.02-1.34)	1.39 (1.30-1.49)	1.38 (1.18-1.61)	0.84 (0.67-1.05)	0.95 (0.69-1.32)	1.00 (0.93-1.08)
Healthcare workers^a							
No. of cases	9029	117	402	50	28	30	222
No. of participants	97671	1470	3805	485	371	241	2272
Age-adj OR (95% CI) ^b	1.0 (reference)	1.07 (0.88-1.30)	1.24 (1.11-1.39)	1.37 (1.01-1.86)	0.74 (0.50-1.10)	1.52 (1.03-2.24)	1.05 (0.91-1.22)
Comorbidity-adj OR (95% CI) ^d	1.0 (reference)	1.08 (0.89-1.32)	1.25 (1.12-1.39)	1.37 (1.01-1.85)	0.74 (0.49-1.10)	1.52 (1.03-2.24)	1.06 (0.92-1.23)
Comorbidity + occupational risk factors-adj OR (95% CI) ^e	1.0 (reference)	0.97 (0.80-1.19)	1.16 (1.04-1.30)	1.24 (0.91-1.69)	0.73 (0.49-1.09)	1.18 (0.80-1.73)	1.05 (0.91-1.22)
Comorbidity + occupational risk factors + personal contact with COVID-19-adj OR (95% CI) ^f	1.0 (reference)	0.96 (0.78-1.17)	1.14 (1.02-1.28)	1.27 (0.94-1.72)	0.67 (0.45-0.99)	1.09 (0.74-1.62)	1.04 (0.89-1.20)

Abbreviation: adj, adjusted; CI, confidence interval; OR, odds ratio.

^a All models were weighted according to the inverse probability of testing for COVID-19 calculated as a function of age, sex, date of study entry, race/ethnicity, symptoms (fatigue, headache, sore throat, chest pain, shortness of breath, persistent cough, diarrhea, abdominal pain, skipped meals/anorexia, hoarse voice, myalgias, delirium, loss of smell/taste, fever), and occupation as frontline healthcare worker (among overall participants).

^b Logistic regression model conditioned on age, sex, and date of study entry.

^c Additionally adjusted for body mass index (17-18.4, 18.5-24.9, 25-29.9, and ≥ 30 kg/m²), history of diabetes (no, yes), heart disease (no, yes), lung disease or asthma (no, yes), kidney disease (no, yes), cancer (active or in the past; no, yes), and smoking status (never/former smokers, current smokers).

^d Additionally adjusted for occupation as frontline healthcare worker (no, yes; among overall participants). For healthcare workers, the model was additionally adjusted for access to personal protective equipment (reuse or inadequate, adequate) and practice setting (inpatient, nursing homes, outpatient hospital clinics, home health sites, ambulatory clinics, other).

^e Additionally adjusted for personal contact with COVID-19 (no, suspected COVID-19, confirmed COVID-19).

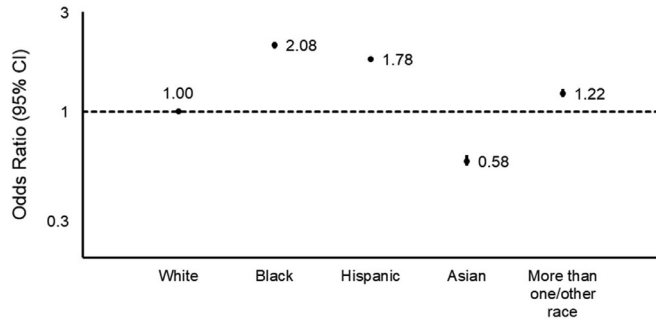
COVID-19 infection among minority populations in the US and the UK. In contrast, contact in the community with individuals with COVID-19 played an important role in the differential risk of COVID-19 among racial/ethnic minorities, as shown by the attenuation of the race/ethnicity-COVID-19 association and the mediation analysis.

An individual's likelihood of contact with COVID-19 can be influenced by structural and individual factors. On the structural level, communities of color are highly represented among the essential workforce [7] in the US and the UK and thus may be less able to effectively practice social distancing. In both countries, Asian and Hispanic populations are more likely than Whites to live in multigenerational households [3], and, like Black populations, are more likely to live in densely populated urban areas [4]. Moreover, due to residential segregation, particularly in the US, racial and ethnic minorities may live in neighborhoods with a higher prevalence of infection, thereby increasing their risk of coming into contact with infected members of the community [30]. Even when accounting for such factors, an individual's ability to adhere to social distancing guidelines could have an additional influence on the likelihood of infection. Structural barriers such as not being able to work remotely [5-7], lacking access to private transportation, or living in crowded communities likely limit the ability of minority populations to socially distance [4]. Furthermore, individuals from certain racial and ethnic minority groups may also face barriers to accessing healthcare, such as lack of health insurance, transportation, childcare, or ability to take time off work [31], which may cause a delay in the detection of SARS-CoV-2 infection and can further facilitate its transmission. We observed that the racial and ethnic disparities in COVID-19 risk were more pronounced in the initial waves of the pandemic in both countries. This disproportionate impact experienced by racial and ethnic minority groups might have improved throughout the course due to both countries' efforts to expand the testing capacity and improve diagnostic methods as the

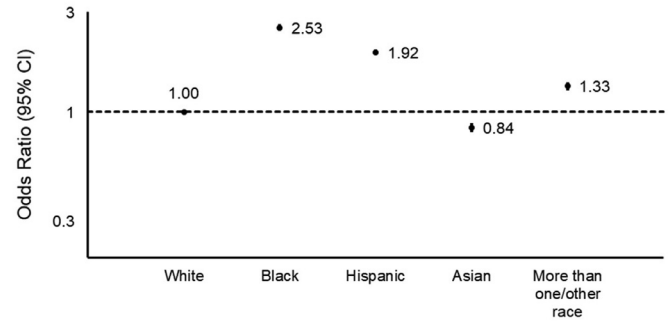
need for identifying the source of infection became the priority [31]. As the fundamental understanding of the COVID-19 pandemic evolved, additional factors may have also contributed to lessening these disproportionate risks including improved access to personal protective equipment and evolving work safety standards in essential work settings. Moreover, community-level awareness of the risks for COVID-19 especially among minority communities may have increased with time and additional protective measures may have been adopted.

Importantly, our results demonstrate that medical comorbidities do not explain the excess risk experienced by racial and ethnic minorities and support the central role that inequities in the social determinants of health play in COVID-19 infection. We showed that community-level socioeconomic factors mediated a significant portion of the COVID-19 risk in Black participants. Even after accounting for personal contact with COVID-19, occupation as frontline HCW, and comorbid conditions, community-level socioeconomic factors still mediated 16.6% and 7.7% of the excess COVID-19 risk compared to White participants in Black participants in the US and the UK, respectively. The role of genetics in the now established racial disparities in COVID-19 risk is controversial. Although some studies have demonstrated that the expression of the gene facilitating SARS-CoV-2 infection is greater among Black individuals compared to those of other races/ethnicities [32], so far there is little evidence to support the role of underlying genetics as a primary contributor to these observed differences. Race, as measured in this and other studies, is a social construct, self-defined by participants, that changes with time and between geographic locations, and has variable concordance with genetic ancestry [33]. Our results demonstrate increased COVID-19 risk across multiple racial/ethnic groups who are socially and economically marginalized to varying degrees yet are unlikely to share genetic variation in yet unidentified genes associated with COVID-19 risk.

Percent with a bachelor's degree (lowest quintile)



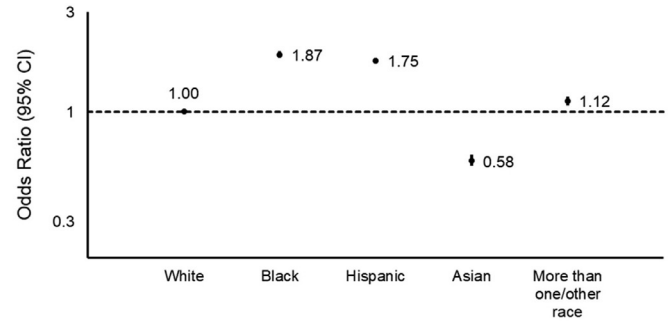
Percent males in management (lowest quintile)



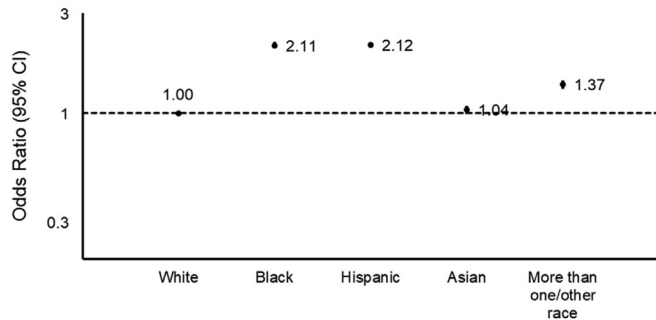
Percent females in management (lowest quintile)



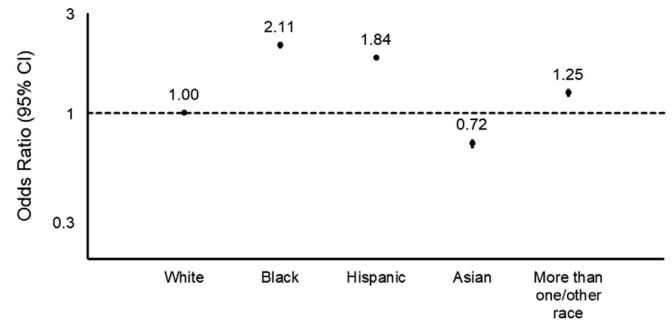
Percent males in professional occupations (lowest quintile)



Percent females in professional occupations (lowest quintile)



Percent essential worker (highest quintile)



Median household value (lowest quintile)

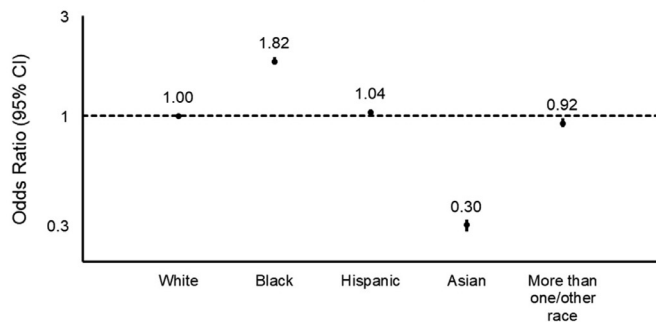


Fig. 2. Risk of living in community with specific measures of deprivation according to race and ethnicity in the United States. Data points represent the odds ratios with 95% confidence intervals. White participants were used as the reference group. "Hispanic" was defined as any race of Hispanic or Latino ancestry, while other racial categories were defined as each respective race not of Hispanic or Latino ancestry. Census data from the US Census Bureau were assigned to each participant based on ZIP Code Tabulation Areas. Each domain categorized based on these cutoffs was associated with an increased risk of personal contact with COVID-19 and testing positive for COVID-19 in **Supplementary Fig. 1**.

Our study has several strengths. The bi-national data from a common survey instrument provided a unique opportunity to compare findings in racial/ethnic minority groups with shared ancestry yet disparate social, economic, and cultural experiences. Notably, racial and ethnic disparities in COVID-19 risk exist in both countries despite social and policy differences between the two areas, such as a

universal healthcare system in the UK versus the hybrid private and public healthcare system in the US. Other strengths include examining COVID-19 in a population-wide sample, overcoming limitations related to capturing only more severe cases through hospitalization records or death reports; accounting for a wide range of personal risk factors for COVID-19 generally not available at a population scale;

Table 5
Risk of personal contact with COVID-19 and testing positive for COVID-19 according to community-level deprivation.^a

	Personal contact with COVID-19		COVID-19 infection	
	No. of cases	OR (95% CI) ^b	No. of cases	OR (95% CI) ^b
United States				
Quintile 1 (least deprived)	6721	1 (reference)	1281	1 (reference)
Quintile 2	6732	1.01 (0.97-1.04)	1556	1.20 (1.12-1.30)
Quintile 3	7199	1.07 (1.03-1.10)	1810	1.36 (1.26-1.46)
Quintile 4	7977	1.14 (1.10-1.17)	2056	1.46 (1.36-1.57)
Quintile 5 (most deprived)	9356	1.24 (1.20-1.28)	2654	1.71 (1.60-1.84)
<i>P</i> _{trend}		<0.001		<0.001
United Kingdom				
Quintile 1 (least deprived)	52716	1 (reference)	8274	1 (reference)
Quintile 2	50106	1.04 (1.03-1.05)	7560	1.05 (1.01-1.08)
Quintile 3	52851	1.04 (1.02-1.05)	8582	1.04 (1.01-1.07)
Quintile 4	53159	1.06 (1.05-1.08)	8437	1.09 (1.05-1.12)
Quintile 5 (most deprived)	56063	1.11 (1.10-1.12)	9775	1.26 (1.22-1.29)
<i>P</i> _{trend}		<0.001		<0.001

Abbreviations: CI, confidence interval; OR, odds ratio.

^a Community-level deprivation was represented by the Neighborhood Deprivation Index in the United States and the Index of Multiple Deprivation in the United Kingdom. Higher scores represented more deprived neighborhood.

^b Logistic regression model adjusted for age, sex, and date of study entry.

and accounting for community-level sociodemographic variables for each participant.

We acknowledge several limitations. While the use of syndromic surveillance to better understand the COVID-19 pandemic has great strengths in flexibility, speed, and sample size, this methodology is dependent upon self-reported data, which is susceptible to measurement error, and voluntary participation, which is prone to selection (collider) bias [34,35]. Racial and ethnic minority groups and those who are older, have lower income, or have lower health literacy may be less likely to participate in a smartphone-based study [36]. Although the proportion of racial and ethnic minority participants in our study was lower than national demographics, we were still able to enroll a considerable number of participants from these underrepresented groups [18]. In addition, given the nature of the study, we recruited primarily individuals from different racial and ethnic minority groups with similar levels of access to technology. This may minimize differences in community-level socioeconomic factors observed in the general population, which would tend to underestimate broader racial and ethnic disparities in COVID-19 risk compared to population samples that do not incorporate voluntary participation through a common data collection instrument. In fact, the use of a smartphone application for data collection allowed us to demonstrate racial and ethnic disparities that persisted despite uniform access to technology. Although it is possible that symptomatic or high-risk individuals may be more likely to participate, it is unlikely that this difference would vary by race or ethnicity. Furthermore, our findings from the mediation analysis should be interpreted with caution due to the potential for unmeasured individual-level variables that might correlate with community-level factors. We used community-level data to estimate socioeconomic factors which may not capture more specific structural inequities experienced by a participant. However, assessing sensitive personal data on such a large scale would not have been feasible. Finally, the exclusion of participants who declined to provide data on race and ethnicity might fail to account for a severely underrepresented population, potentially biasing results towards the null.

In conclusion, within a large population-wide sample of individuals in the US and the UK, we demonstrate a significantly increased risk of COVID-19 in certain racial and ethnic minority groups compared to White individuals, which appeared to be mediated in part by community-level socioeconomic factors, especially in Black individuals. Our findings stress the importance of allocating resources to

specific communities and the need to build robust public health infrastructure accessible to all.

Data sharing statement

Data collected in the app are being shared with other health researchers through the NHS-funded Health Data Research UK (HDRUK)/SAIL consortium, housed in the UK Secure e-Research Platform (UKSeRP) in Swansea. Anonymized data collected by the COVID Symptom Study app can be shared with bonafide researchers via HDRUK, provided the request is made according to their protocols and is in the public interest (see <https://healthdatagateway.org/detail/9b604483-9cdc-41b2-b82c-14ee3dd705f6>). US investigators are encouraged to coordinate data requests through the COPE Consortium (www.monganinstitute.org/cope-consortium). Data updates can be found at <https://covid.joinzoe.com>.

Funding

LHN is supported by the American Gastroenterological Association Research Scholars Award and the National Institute of Diabetes and Digestive and Kidney Diseases (K23 DK125838). LHN and DAD are supported by the American Gastroenterological Association AGA-Takeda COVID-19 Rapid Response Research Award (AGA2021-5102). DAD is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (K01 DK120742). ATC is the Stuart and Suzanne Steele MGH Research Scholar and Stand Up to Cancer scientist. The Massachusetts Consortium on Pathogen Readiness (MassCPR) and Mark and Lisa Schwartz supported MGH investigators (LHN, DAD, ADJ, CGG, WM, RSM, C-HL, SK, ATC). King's College of London investigators (MG, CHS, MJC, SO, CJS, TDS) were supported by the Wellcome Trust and EPSRC (WT212904/Z/18/Z, WT203148/Z/16/Z, T213038/Z/18/Z), the NIHR GSTT/KCL Biomedical Research Centre, the Medical Research Council/British Heart Foundation (MR/M016560/1), UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare, and the Alzheimer's Society (AS-JF-17-011). CMA is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (K23 DK120899) and the Boston Children's Hospital Office of Faculty Development Career Development Award. The Multiethnic Cohort investigators (LRW, CAH, LLM) were supported by the National Institutes of Health (U01 CA164973). The Black Women's Health Study investigators (JRP, YCC,

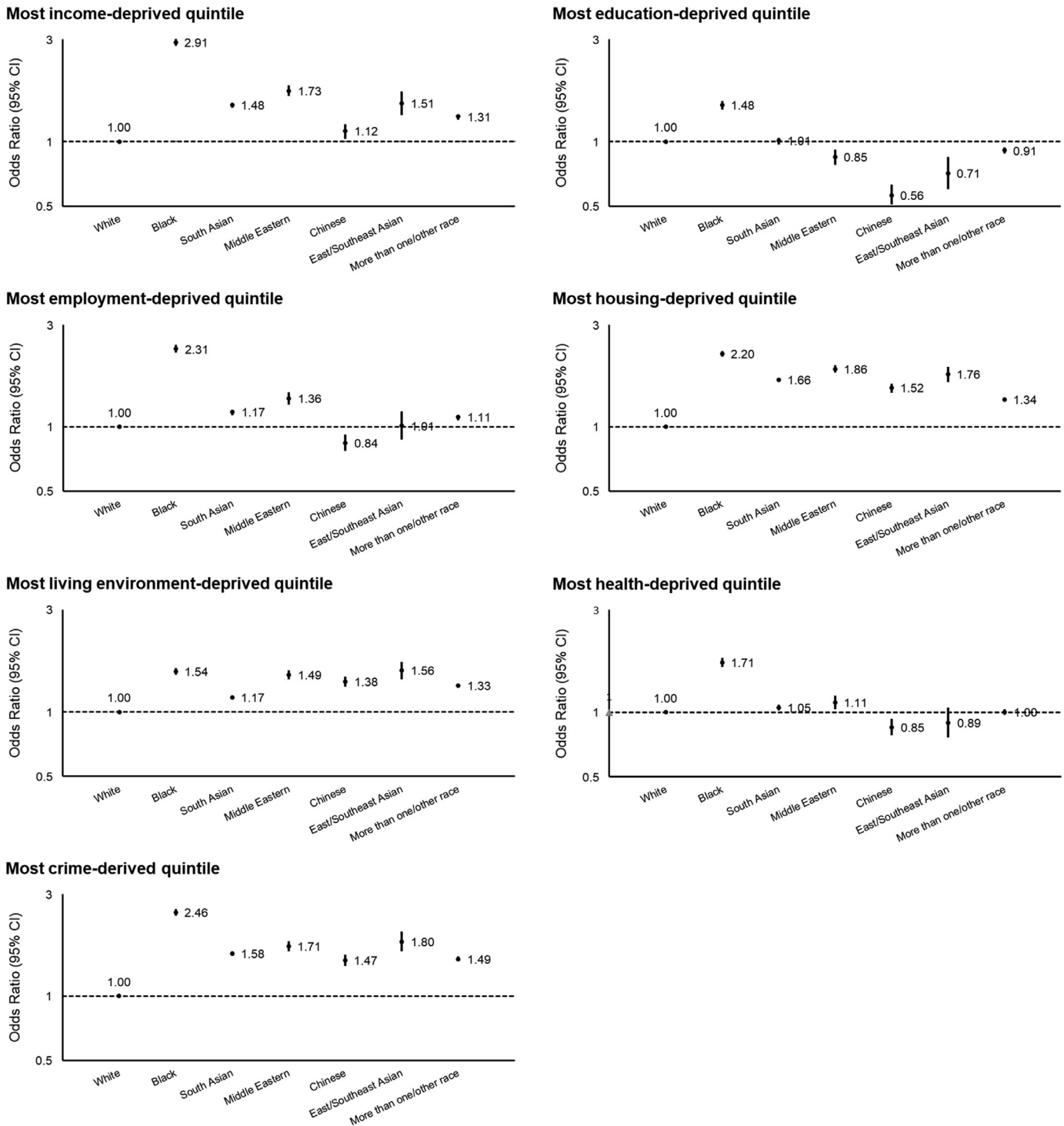


Fig. 3. Risk of living in community with specific measures of deprivation according to race and ethnicity in the United Kingdom. Data points represent the odds ratios with 95% confidence intervals. White participants were used as the reference group. Data from the Office for National Statistics (England), the Welsh Government (Wales), the Scottish Government (Scotland), and the Northern Ireland Statistics and Research Agency (Northern Ireland) were assigned to each participant based on Lower Layer Super Output Areas. Each domain except for housing categorized based on these cutoffs was associated with an increased risk of personal contact with COVID-19 and testing positive for COVID-19 in **Supplementary Fig. 2**.

LR) were supported by the National Institutes of Health (U01 CA164974). The sponsors had no role in study design, analysis, and interpretation of data, report writing, and the decision to submit for publication. The corresponding author had full access to data and the final responsibility to submit for publication. Zoe Limited, American Gastroenterological Association, National Institutes of Health, Wellcome Trust, Medical Research Council/British Heart Foundation,

Alzheimer’s Society, and Massachusetts Consortium on Pathogen Readiness.

Declaration of Competing Interest

JC reports personal fees from Zoe Ltd. during the conduct of the study and outside the submitted work and is an employee of Zoe Ltd.

Table 6

Proportion of excess COVID-19 risk in racial/ethnic minority participants compared to White participants mediated by community-level deprivation in the United States and the United Kingdom.

	Excess risk attributable to community-level deprivation ^a	Excess risk attributable to personal contact with COVID-19 ^b	Excess risk attributable to occupation as frontline healthcare worker	Excess risk attributable to comorbid conditions ^c	Total excess risk mediated by community-level deprivation and personal risk factors
United States					
White	Reference	Reference	Reference	Reference	Reference
Black	16.6%	6.6%	<1.0%	<1.0%	24.0%
Hispanic	5.9%	3.0%	<1.0%	<1.0%	9.2%
United Kingdom					
White	Reference	Reference	Reference	Reference	Reference
Black	7.7%	3.8%	6.5%	<1.0%	18.2%
South Asian	2.0%	1.0%	1.0%	<1.0%	4.0%
Middle Eastern	2.2%	<1.0%	1.3%	<1.0%	4.3%

^a Community-level deprivation, not otherwise explained by personal risk factors, as represented by the Neighborhood Deprivation Index in the United States and the Index of Multiple Deprivation in the United Kingdom.

^b Personal contact with COVID-19 is represented by a report of contact with individuals in the community with suspected or confirmed COVID-19.

^c Comorbidities include overweight/obesity, diabetes, heart disease, lung disease/asthma, and smoking.

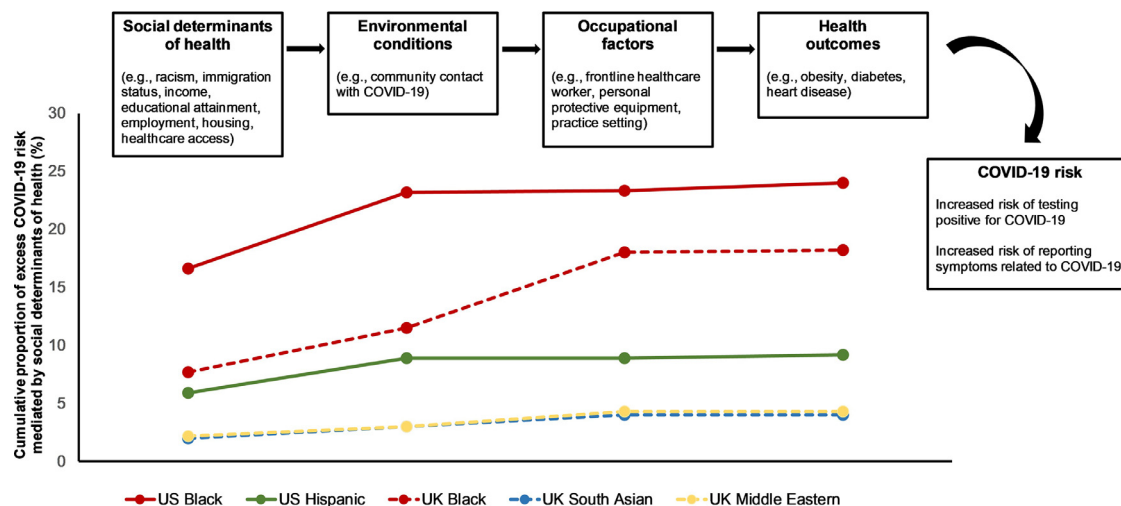


Fig. 4. Connections and consequences of social determinants of health in the context of COVID-19 risk. We propose that social determinants of health are the root causes of health disparities at the population level. Upstream social determinants contribute to midstream factors, such as environmental conditions and occupational factors, which in turn mediate the effects of social determinants of health on downstream health outcomes, including obesity and diabetes, among others. Finally, excess downstream COVID-19 risk is proposed to result from the upstream and midstream factors. Examples of each determinant and their contribution to COVID-19 risk are presented.

TDS is a consultant to Zoe Ltd. LLM reports grants from National Cancer Institute during the conduct of the study. SO reports grants from Wellcome Trust, grants from Innovate UK (UKRI), grants from Chronic Disease Research Foundation (CDRF) outside the submitted work. DAD reports grants from National Institutes of Health, grants from American Gastroenterological Association during the conduct of the study and that he previously served as an investigator for a clinical trial of diet and lifestyle using a separate mobile application that was supported by Zoe Ltd. RD reports grants from Department of Health and Social Care (UK), personal fees from Zoe Ltd. during the conduct of the study and outside the submitted work and is an employee of Zoe Ltd. CJS reports grants from Chronic Disease Research Foundation during the conduct of the study. CHS reports grants from Alzheimer's Society during the conduct of the study. CMA reports grants from National Institutes of Health K23 DK120899 and Boston Children's Hospital Office of Faculty Development Career Development Award during the conduct of the study. ATC reports personal fees from Bayer Pharma AG, Pfizer Inc., Boehringer Ingelheim, and grants from Zoe Ltd. outside the submitted work and that he previously served as an investigator for a clinical trial of diet and lifestyle using a separate mobile application that was supported by Zoe Ltd. JW is the CEO of Zoe Ltd. Other authors have no conflict of interest to declare.

Acknowledgments

We express our sincere thanks to all of the participants who entered data into the app, including study volunteers enrolled in cohorts within the Coronavirus Pandemic Epidemiology (COPE) consortium. We thank the staff of Zoe Ltd., the Department of Twin Research at King's College London, and the Clinical and Translational Epidemiology Unit at Massachusetts General Hospital for their tireless work. Zoe Ltd. provided in kind support for all aspects of building, running and supporting the app and service to users worldwide.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.eclinm.2021.101029](https://doi.org/10.1016/j.eclinm.2021.101029).

References

- [1] Sze S, Pan D, Nevill CR, Gray IJ, Martin CA, Nazareth J, et al. Ethnicity and clinical outcomes in COVID-19: a systematic review and meta-analysis. *EClinicalMedicine* 2020;29–30:100630.
- [2] Iceland J, Mateos P, Sharp G. Ethnic residential segregation by nativity in Great Britain and the United States. *J Urban Aff* 2011;33(4):409–29.

- [3] U.S. Census Bureau. Multigenerational households: 2009–2011 Internet. 2012 Oct 1. Available from: <https://www.census.gov/library/publications/2012/acs/acsbr11-03.html>
- [4] Caldwell JT, Ford CL, Wallace SP, Wang MC, Takahashi LM. Intersection of living in a rural versus urban area and race/ethnicity in explaining access to health care in the United States. *Am J Public Health* 2016;106(8):1463–9.
- [5] U.S. Bureau of Labor Statistics. Workers who could work at home, did work at home, and were paid for work at home, by selected characteristics, averages for the period 2017–2018 Internet. 2019 Sep 24. Available from: <https://www.bls.gov/news.release/flex2.t01.htm>
- [6] U.S. Bureau of Labor Statistics. Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity Internet. 2021 Jan 22. Available from: <https://www.bls.gov/cps/cpsaat11.htm>
- [7] Lymperopoulou K, Finney N. Socio-spatial factors associated with ethnic inequalities in districts of England and Wales, 2001–2011. *Urban Stud* 2017;54(11):2540–60.
- [8] Figueroa JF, Wadhwa RK, Lee D, Yeh RW, Sommers BD. Community-level factors associated with racial and ethnic disparities in COVID-19 rates in Massachusetts. *Health Aff* 2020;39(11):1984–92.
- [9] Seligman B, Ferranna M, Bloom DE. Social determinants of mortality from COVID-19: a simulation study using NHANES. *PLoS Med* 2021;18(1):e1003490.
- [10] Hatfe E, Chang HY, Kitchen C, Weiner JP, Kharrazi H. Assessing the impact of neighborhood socioeconomic characteristics on COVID-19 prevalence across seven states in the United States. *Front Public Health* 2020;8:571808.
- [11] Figueroa JF, Wadhwa RK, Mehtsun WT, Riley K, Phelan J, Jha AK. Association of race, ethnicity, and community-level factors with COVID-19 cases and deaths across U.S. counties. *Healthcare* 2021;9(1):100495.
- [12] Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and mortality among black patients and white patients with Covid-19. *N Engl J Med* 2020;382(26):2534–43.
- [13] Azar KMJ, Shen Z, Romanelli RJ, Lockhart SH, Smits K, Robinson S, et al. Disparities in outcomes among COVID-19 patients in a large health care system in California. *Health Aff* 2020;39(7):1253–62.
- [14] Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020;368(6497):1362–7.
- [15] Chan AT, Drew DA, Nguyen LH, Joshi AD, Ma W, Guo CG, et al. The COronavirus Pandemic Epidemiology (COPE) consortium: a call to action. *Cancer Epidemiol Biomark Prev* 2020;29(7):1283–9.
- [16] National Institutes of Health. Racial and ethnic categories and definitions for NIH diversity programs and for other reporting purposes [Internet]. 2015 Apr 8. Available from: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-089.html>
- [17] Office for National Statistics. Ethnic group, national identity and religion [Internet]. Available from: <https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion#ethnic-group>
- [18] U.S. Census Bureau. Quick Facts in the United States [Internet]. 2019 Jul 1. Available from: <https://www.census.gov/quickfacts/fact/table/US/PST045219>
- [19] Office for National Statistics. Population of England and Wales [Internet]. 2018 Aug 1. Available from: <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest>
- [20] U.S. Census Bureau. Explore Census Data [Internet]. Available from: <https://data.census.gov/cedsci/>
- [23] Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, et al. The development of a standardized neighborhood deprivation index. *J Urban Health* 2006;83(6):1041–62.
- [24] Ministry of Housing, Communities & Local Government. English indices of deprivation 2019 [Internet]. GOV.UK. 2019 Sep 26. Available from: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>
- [25] Welsh Government. Welsh Index of Multiple Deprivation [Internet]. 2019 Nov 27. Available from: <https://gov.wales/welsh-index-multiple-deprivation-full-index-update-ranks-2019>
- [26] Scottish Government. Scottish Index of Multiple Deprivation [Internet]. 2020. Available from: <https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/>
- [27] Northern Ireland Statistics and Research Agency. Northern Ireland Multiple Deprivation Measure [Internet]. 2017 Nov 23. Available from: <https://www.nisra.gov.uk/statistics/deprivation/northern-ireland-multiple-deprivation-measure-2017-nimdm2017>
- [28] Nguyen LH, Drew DA, Graham MS, Joshi AD, Guo CG, Ma W, et al. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Health* 2020;5(9):e475–83.
- [29] Lee SF, Nikšić M, Rached B, Sanchez MJ, MA LF. Socioeconomic inequalities and ethnicity are associated with a positive COVID-19 test among cancer patients in the UK Biobank cohort. *Cancers* 2021;13(7):1514.
- [30] Yancy CW. COVID-19 and African Americans. *JAMA* 2020;323(19):1891.
- [31] Thakur N, Lovinsky-Desir S, Bime C, Wisnivesky JP, Celedón JC. The structural and social determinants of the racial/ethnic disparities in the U.S. COVID-19 pandemic. what's our role? *Am J Respir Crit Care Med* 2020;202(7):943–9.
- [32] Bunyavanich S, Grant C, Vicencio A. Racial/ethnic variation in nasal gene expression of transmembrane serine protease 2 (TMPRSS2). *JAMA* 2020;324(15):1567.
- [33] Gravelle CC. How race becomes biology: embodiment of social inequality. *Am J Phys Anthropol* 2009;139(1):47–57.
- [34] Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15(5):615–25.
- [35] Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun* 2020;11(1):5749.
- [36] Jones J, Sullivan PS, Sanchez TH, Guest JL, Hall EW, Luisi N, et al. Similarities and differences in COVID-19 awareness, concern, and symptoms by race and ethnicity in the United States: cross-sectional survey. *J Med Internet Res* 2020;22(7):e20001.
- [37] Chambers L. Data show COVID-19 is hitting essential workers and people of color hardest [Internet]. ACLU of Massachusetts Data for Justice. 2020. Available from: <https://data.aclum.org/2020/04/07/covid-19-disproportionately-affects-vulnerable-populations-in-boston/>