Journal of Clinical Epidemiology 161 (2023) 94−103

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Revisiting ethnic discrepancies in COVID-19 hospitalized cohorts: a correction for collider bias

Annastazia E. Learoyd[a,*], Jennifer Nicholas[b], Nicholas Hart[c,d], Abdel Douiri[a]

[a]*School of Life Course and Population Sciences, King College London, London, UK*
[b]*Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK*
[c]*Lane Fox Clinical Respiratory Physiology Research Centre, Guy's & St Thomas' NHS Foundation Trust, London, UK*
[d]*Centre for Human and Applied Physiological Sciences, King's College London, London, UK*

Accepted 20 June 2023; Published online 28 June 2023

Abstract

Objectives: Studies from the first waves of the coronavirus disease 2019 (COVID-19) pandemic suggest that individuals from minority ethnicities are at an increased risk of worse outcomes. Concerns exist that this relationship is potentially driven by bias from analyzing hospitalized patients only. We investigate this relationship and the possible presence of bias.

Study Design and Setting: Using data from South London hospitals across two COVID-19 waves (February 2020 − May 2021), the relationship between ethnicity and COVID-19 outcomes was examined using regression models. Three iterations of each model were completed: 1) an unadjusted analysis, 2) adjusting for covariates (medical history and deprivation), and 3) adjusting for covariates and bias induced by conditioning on hospitalization.

Results: Among 3,133 patients, those who were Asian had a two-fold increased risk of death during the hospital stay that was consistent across the two COVID-19 waves and was not affected by correcting for conditioning on hospitalization. However, wave-specific effects demonstrate significant differences between ethnic groups until bias from using a hospitalized cohort was corrected for.

Conclusion: Worsened COVID-19 outcomes in minority ethnicities may be minimized by correcting for bias induced by conditioning on hospitalization. Consideration of this bias should be a key component of study design.

Keywords: Collider bias; COVID-19; Epidemiology; Ethnicity; Inequalities; Statistical methods; Observational studies

1. Introduction

From the beginning of the coronavirus disease 2019 (COVID-19) pandemic, greater disease severity and worsened outcomes have been noted in ethnic minorities [1−3]. As the pandemic has progressed, evidence has emerged that this relationship is not consistent. In the United Kingdom (UK) ethnic differences in outcomes changed between waves of COVID-19 [4,5]. This may be because this relationship is driven by social causes [6−8] including increased poverty/deprivation, health inequalities, and differences in occupation, rather than biological mechanisms.

Another factor that could explain the divergent findings is the populations being studied. Many studies that look into ethnic differences within COVID-19 focus on hospitalized patients [9−11] due to the easily accessible data contained in electronic health records (EHRs). This allowed studies to be completed within months of the discovery

Funding: The authors acknowledge funding support from the National Institute for Health Research (NIHR) Applied Research Collaboration South London at King's College Hospital NHS Foundation Trust and the Royal College of Physicians, as well as the King's Health Partners/Guy's and St Thomas Charity 'MLTC Challenge Fund' (grant number EIC180702). We also acknowledge support from the NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London.

Author contributions: AEL completed the analysis. JN, and AD provided statistical input. AEL wrote the manuscript. AEL and AD conceived and designed the study. NH sponsored the acquisition of data. AEL completed the statistical analysis, drafted the manuscript, and completed the revision after peer review. JN and AD assisted AEL with interpretation of the data. All other authors reviewed the manuscript, provided critical revisions to the initial manuscript, and provided approval to proceed with resubmission after peer review.

* Corresponding author. Faculty of Life Sciences and Medicine, School of Life Course and Population Sciences, 4th Floor, Addison House, King's College London; Guy's Campus, London SE1 1UL, UK.

E-mail address: annastazia.learoyd@kcl.ac.uk (A.E. Learoyd).

https://doi.org/10.1016/j.jclinepi.2023.06.014

What is new?

Key findings

- Bias caused by conditioning on covariates is a concern in Covid-19 studies.
- Statistical methods can account for this conditioning correcting aberrant results.

What this adds to what was known?

- This study is one of the first clear examples of the presence of sampling biases such as collider bias in Covid-19 research.
- This study provides a real life example of how statistical techniques utilising multiple data sources can correct for such sampling biases.

What is the implication and what should change now?

- Researchers should consider the potential for collider bias among other sampling biases in any observational studies that utilise data from a restricted population.

of SARS-CoV-2, but, as only the most severe cases of COVID-19 are typically admitted to hospital, it also provides a restricted and potentially biased sample.

At least two types of bias could be prevalent in COVID-19 research utilizing hospitalized cohorts. The first is overadjustment bias, where a mediator (or descending proxy) between the exposure and outcome of interest is controlled for and partially blocks the association between these variables (Fig. 1A). In COVID-19, severity is a potential mediator between ethnicity and increased mortality. Hospitalization is a descending proxy for severity, and therefore conditioning on this variable through the use of a hospitalization cohort means studies are at risk of overadjustment bias. Secondly, differences in risk of hospitalization with COVID-19 have been noted between ethnic groups [4,11,12]. If this association is via paths other than differences in COVID-19 severity, then hospitalization becomes a collider [13] — a variable influenced by both the exposure and a mediator/outcome and when conditioned on induces an association (Fig. 1B). Both biases can mean that the associations seen in hospitalized cohorts are not reflective of the causal effect of ethnicity on COVID-19 mortality.

This study aims to illustrate the potential impact of these biases in an analysis of the association of ethnicity and COVID-19 outcomes over the first two waves of COVID-19 using data from an ethnically diverse South London hospitalized cohort.

2. Methods

2.1. Patient population

Five thousand nine hundred and ninety-two patients were admitted to two hospitals within the Guy's and St. Thomas' National Health Service Foundation Trust (GSTT) and received a COVID-19 positive test between February 20, 2020, and May 24, 2021.

The following patients were excluded from this analysis (Fig. 2): those without a known date of admission and discharge; those with admission prior to January 28, 2020 (date of the first known COVID-19 cases within the UK) or with the only recorded COVID-19 positive test more than 28 days prior to admission (both indicators of non-COVID-19-related admission); those transferred from other National Health Service trusts for a higher level of care; those under 18 years old; and those without a known index of multiple deprivation (IMD) quintile. The analyzed cohort included 3,133 patients (52.3% of the initial population).

2.2. Data sources

Anonymized clinical, laboratory, and demographic data for patients with a positive reverse transcription-polymerase chain reaction COVID-19 test was collated from six linked EHR databases. Data management was performed using SQL, with analysis carried out on the secure King's Health Partners Rosalind high-performance computer infrastructure running Jupyter Notebook 6.0.3, R

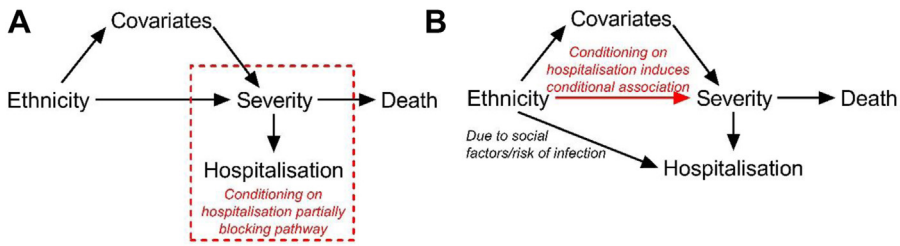

Fig. 1. Directed acyclic graph demonstrating how the relationship between ethnicity and COVID-19 severity/mortality may be influenced by (A) overadjustment bias and (B) collider bias. Overadjustment bias will reduce the association between ethnicity and COVID-19 mortality. Collider bias will induce an association between ethnicity and COVID-19 severity/mortality.

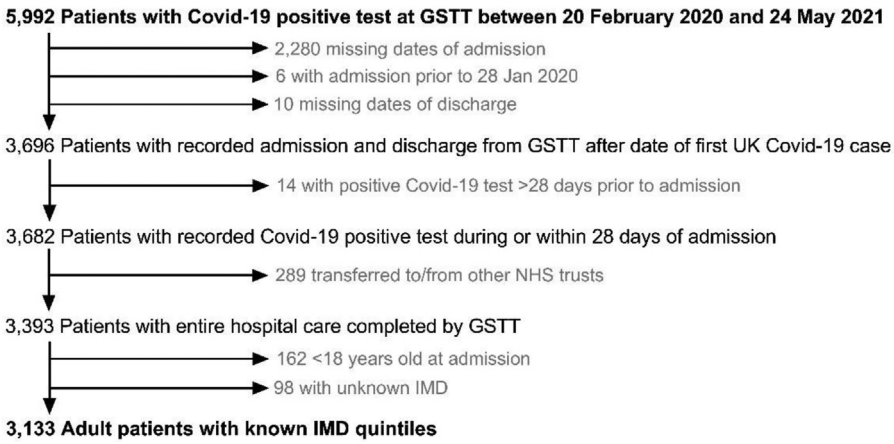

Fig. 2. Study population flowchart. GSTT, Guy's and St. Thomas' National Health Service Foundation Trust; IMD, index of multiple deprivation.

3.6.3, and Python 3.7.6. The study authors did not have access to the databases used in the initial data linkage.

Patient-level data and programming code is unavailable due to infrastructure changes in August 2022. Further summarized data on patient subgroups (ethnicity, sex, medical history etc.) is available upon request.

2.3. Ethics

Ethical approval was granted by The London Bromley Research Ethics Committee (reference (20/HRA/1871)) to the King's Health Partners Data analytics and modeling COVID-19 group to collect clinically relevant data points from patient's EHRs. Access to Lambeth DataNet was under a project-specific approval granted by Lambeth Public Health Caldicott Guardian, 26 June 2020. Individual patient informed consent was not required.

2.4. Exposures

The primary exposure was ethnicity categorized as White (British, European, and other), Black (African and Caribbean), Asian (South, South-East, and East Asian), Mixed/other (Middle Eastern, South American, and Mixed), and unknown (or not reported).

Other variables included in adjusted analyses include age, sex, IMD quintile, medical history of comorbidities, and "do not attempt resuscitation" (DNACPR) orders. No interactions were observed between variables during model comparisons.

Age and sex were extracted as recorded at admission. For analysis, age was centered on the mean of 59.7 years. A linear association between age and COVID-19 outcomes was deemed appropriate following inspection of model-specific residual plots.

IMD is a relative measure of deprivation for small regional areas in the UK based on seven domains of deprivation [14]. Patient addresses were linked to IMD and organized into quintiles, with 1 denoting the most deprived areas and five the least deprived ones. Quintile two was used as the reference group during analysis due to the prominence of this quintile within this population.

Comorbidities were extracted from a combination of three linked EHR databases using either International Classification of Diseases 10 codes or automated searches of free text data. Patients were categorized as having/not having a medical history of cardiovascular disease (stroke, transient ischemic attack, atrial fibrillation, congestive heart failure, ischemic heart disease, valve disease, peripheral artery disease, or atherosclerotic disease), diabetes mellitus, chronic kidney disease, chronic liver disease, and chronic obstructive pulmonary disease/emphysema. These comorbidities were included based on known association with COVID-19 outcomes and known links to ethnicity (Supplementary Fig. 1).

The application of DNACPR orders was extracted along with the date of application as an indicator of level of care. DNACPR was treated as a time-dependent covariate in the analysis of mortality and a binary covariate of application or no application by 30 days for analysis of intensive care unit (ICU) admission or death.

2.5. Outcomes

Two primary outcomes were assessed: 1) time from admission to all-cause in-hospital mortality and 2) a composite binary outcome ICU admission or death within 30 days [15]. Secondary outcomes examining respiratory measures recorded within 24 hours of admission (as a marker of COVID-19 severity) were also examined (Supplementary Materials, page 7−14).

2.6. Statistical analysis

Demographic characteristics, medical history, and outcomes were compared between ethnicities using Kruskal−Wallis tests (continuous measures) or chi-squared tests (counts).

Time to death was analyzed using a competing risk regression model with discharge as a competing event [16]. The proportional subhazards assumption was checked using log-log plots and estimated survivor curves comparing levels of demographic characteristics.

ICU admission/death within 30 days [15] was analyzed using logistic regression. Linear predictor specification and model fit were deemed reasonable using scatter plots and Q-Q plots of residuals.

The association of ethnicity with each outcome was assessed using four model iterations: 1) unadjusted (no covariates), 2) adjusted for age and sex, 3) adjusted for all covariates (age, sex, IMD, and medical history), 4) adjusted for all covariates, and 5) using inverse probability weighting (IPW) to account for conditioning on hospitalization [17]. All models were completed for each outcome using data from all patients, those admitted during wave 1 of COVID-19 (February 2020−August 2020) and those admitted during wave two (September 2020−May 2021).

IPWs were calculated from estimated probabilities of hospitalization with COVID-19 based on each individual's ethnicity, admission date, and survival status utilising summary data from the OpenSAFELY database as a reference [4]. Full details of this weighting method are published separately [18]. The probabilities of hospitalization were derived using results from analysis of COVID-19 outcomes in a UK national cohort between February 1, and December 31, 2020 [4]. Robust standard errors were used for all models to prevent biased variance estimates from using IPW while maintaining model comparison.

3. Results

3.1. Cohort characteristics

Of the 3,133 patients admitted to the hospital, 53.4% were male, and the mean age was 59.71 ± 18.80 (Table 1). Overall, 63.8% (n = 2,000) of patients had a medical history of at least 1 comorbidity of interest, most commonly cardiovascular conditions, followed by diabetes and chronic kidney disease. The representation of ethnic groups were as follows: 40.4% White, 24.9% Black, 8.6% Asian, 8.6% mixed/other, and 17.7% unknown. Ethnic groups differed in most patient characteristics, including age, sex, and presence of comorbidities (Table 1). There was a trend toward a difference in ethnic distributions between COVID-19 waves (P = 0.061) with a higher proportion of Black people in wave 1 and more people from mixed/other and unknown ethnicities in wave two. Patient characteristics in each wave are described in Supplementary Tables 1 and 2.

3.2. Analysis of in-hospital survival over time

Deaths occurred steadily over time from hospital admission until ~20 days with the mortality rate declining thereafter (Supplementary Fig. 2A). Survival curves for each ethnic group (Supplementary Fig. 2B) suggest lower risks of death in Black and mixed/other ethnicities and a comparable risk of death in Asians and Whites. This is matched by unadjusted hazard ratios (HR) (Table 2) with a 30% (8−47%) and a 55% (25−73%) lower risk of death in Black (P = 0.012) and mixed/other ethnic groups (P = 0.002) compared to White and a nonsignificant increased risk of death in Asians (P = 0.234). The decreased risk of death in these minority ethnicities was driven by the first wave for Black individuals (HR = 0.58 [0.39, 0.87], P = 0.008) and the second wave for mixed/other individuals (HR = 0.32[0.14, 0.74], P = 0.007).

Adjusting for sex and age created comparable hazard rates for in-hospital mortality in Black, mixed/other, and White ethnicities overall, which were modified slightly by the inclusion of additional covariates describing medical history (Table 2). A slightly lower risk of death remained in the first wave for Black individuals (adjusting for sex and age: HR = 0.72 [0.48, 1.07], P = 0.107); adjusted for all covariates: HR = 0.63 [0.39, 1.00], P = 0.049). Adjusting for covariates also revealed a significant increase in risk of death for Asian individuals, which was similar across both waves (adjusting for sex and age: HR = 1.96 [1.39, 2.77], P < 0.001; adjusting for all covariates: HR = 1.94 [1.28, 2.93], P = 0.002) (Table 2).

Correcting for conditioning on hospitalization using IPW (in addition to covariate adjustment) indicated that, in the wider population, the increased risk in Asians was still present (HR = 2.06 [1.15, 3.67], P = 0.014), but there were comparable hazard rates in Black and White ethnic groups in the first wave (HR = 1.06 [0.56, 2.00], P = 0.85).

Differences in risk of death across both waves of COVID-19 seemed to be primarily driven by male sex (HR = 1.47 [1.15, 1.87], P = 0.002), the application of DNACPR orders (HR = 1.36 [1.23, 1.50], P < 0.001), and chronic kidney disease (HR = 1.55 [1.20, 2.01], P = 0.001). Neither age nor IMD quintile were contributing factors (Supplementary Table 4), after accounting for other predictors and correcting for conditioning on hospitalization.

3.3. Analysis of ICU admission/death within 30 days of admission

Median hospital stay length was 5.8 days (range: 0.01−243.6 days). 2,806 (89.6%) patients stayed in hospital for 30 days or less (Supplementary Fig. 3). During hospitalization, 356 (11.4%) patients died (304 within 30 days of admission), and 527 (16.8%) patients were admitted to ICU (513 within 30 days of admission). Median ICU stay was 8.2 days (range: 0.0−136.2 days). Some patients (112 [3.6%]) were readmitted to hospital, including 26 patients readmitted to ICU within 30 days of initial admission. 190 (6.1%) patients died after discharge − 58 within 30 days of initial admission. As a result, there is a

Table 1. Patient characteristics by ethnic group

Patient characteristics	All patients	Ethnicity	
		White	Black
Total num. (%) of patients	3,133	1,265 (40.4%)	779 (24.9%)
Num. (%) patients - Wave 1	1,010	415 (41.1%)	274 (27.1%)
Num. (%) patients - Wave 2	2,123	850 (40.0%)	505 (23.8%)
Covariates			
Index of multiple deprivation (IMD)			
Rank [median (IQR)]	9,324 (6,591–13,839)	9,929 (7,088–14,996)	8,105 (6,008–11,635)
Quintile (%):			
1 (Most)	775 (24.7%)	267 (21.1%)	236 (30.3%)
2	1,528 (48.8%)	617 (48.8%)	402 (51.6%)
3	498 (15.9%)	221 (17.5%)	111 (14.2%)
4	227 (7.2%)	100 (7.9%)	27 (3.5%)
5 (Least)	105 (3.4%)	60 (4.7%)	3 (0.4%)
Age (mean ± SD)	59.71 ± 18.80	64.06 ± 18.81	57.56 ± 18.33
Male sex (%)	1,674 (53.4%)	691 (54.6%)	369 (47.4%)
DNACPR applied			
Number (%)	693 (22.1%)	348 (27.5%)	139 (17.8%)
Time to DNACPR (days)	1.3 (0.2–9.8)	1.1 (0.2–10.3)	1.2 (0.3–8.5)
Cardiovascular conds (%)	1,710 (54.6%)	776 (61.3%)	454 (58.3%)
Chronic Obstructive Pulmonary Disease /Emphysema (%)	262 (8.4%)	176 (13.9%)	27 (3.5%)
Diabetes (%)	875 (27.9%)	326 (25.8%)	276 (35.4%)
Kidney conditions (%)	631 (20.1%)	263 (20.8%)	210 (27.0%)
Liver conditions (%)	82 (2.6%)	42 (3.3%)	20 (2.6%)
Death during hospital stay			
Number (%)	356 (11.4%)	166 (13.1%)	72 (9.2%)
Time to death (days)	11.1 (5.7–21.8)	11.8 (5.9–22.2)	10.8 (5.3–18.9)
Time to censor (days)	5.0 (1.4–12.7)	6.1 (1.9–14.9)	4.8 (1.4–12.6)
ICU admission/Death within 30 days			
Total number (%)	790 (25.2%)	333 (26.3%)	192 (24.7%)
Contribution:			
ICU admission	428 (54.2%)	160 (48.1%)	120 (62.5%)
Death	251 (31.8%)	120 (36.0%)	42 (21.9%)
Both	111 (14.1%)	53 (15.9%)	30 (15.6%)

Abbreviations: IQR, interquartile range; DNACPR, "do not attempt resuscitation" order; COPD, chronic obstructive pulmonary disease.
Times reported as median (IQR).
Censored means patient discharged without experiencing event.
P values come from Kruskal–Wallis (continuous measures) or chi-squared tests (counts).
Bold *P*-values indicate significance at the 5% level.

total of 539 (17.2%) recorded ICU admissions and 362 (11.6%) recorded deaths within 30 days of admission.

Unadjusted analysis indicated no difference in the odds of ICU admission/death within 30 days between ethnicities with similar findings after accounting for covariates (Table 3). Correcting for conditioning on hospitalization suggested an increased odds of ICU admission/death in mixed/other individuals relative to White (OR = 1.49 (1.02, 2.21), $P = 0.047$).

While ethnicity was not associated with a change in the odds of ICU admission/death, male sex (OR = 1.57 (1.31, 1.87), $P < 0.001$), cardiovascular disease (OR = 1.77 (1.43, 2.19), $P < 0.001$), diabetes (OR = 1.31 (1.07, 1.60), $P = 0.010$), chronic kidney disease (OR = 1.61 (1.30, 1.99), $P < 0.001$), and the application of DNACPR orders (OR = 3.80 (3.03, 4.77), $P < 0.001$) were.

4. Discussion

This study found that the use of a hospitalized cohort influenced the magnitude and direction of the association between ethnicity and COVID-19 outcomes. A reduction in mortality in Blacks in wave 1 was seen in the hospitalized

Ethnicity			Comparison
Asian	Mixed/Other	Unknown	*P*-value
268 (8.6%)	268 (8.6%)	553 (17.7%)	
87 (8.6%)	72 (7.1%)	162 (16.0%)	
181 (8.5%)	196 (9.2%)	391 (18.4%)	
8,161 (6,422–12,982)	8,894 (6,731–12,694)	9,939 (6,731–15,898)	<**0.001**
75 (28.0%)	64 (23.9%)	133 (24.1%)	<**0.001**
128 (47.8%)	141 (52.6%)	240 (43.4%)	
40 (14.9%)	38 (14.2%)	88 (15.9%)	
19 (7.1%)	19 (7.1%)	62 (11.2%)	
6 (2.2%)	6 (2.2%)	30 (5.4%)	
56.13 ± 17.71	53.38 ± 17.98	57.58 ± 18.29	<**0.001**
153 (57.1%)	134 (50.0%)	327 (59.1%)	<**0.001**
56 (20.9%)	32 (11.9%)	118 (21.3%)	<**0.001**
1.8 (0.2–15.2)	0.8 (0.2–5.3)	2.8 (0.3–11.0)	0.367
136 (50.7%)	123 (45.9%)	221 (40.0%)	<**0.001**
9 (3.4%)	18 (6.7%)	32 (5.8%)	<**0.001**
96 (35.8%)	56 (20.9%)	121 (21.9%)	<**0.001**
53 (19.8%)	32 (11.9%)	73 (13.2%)	<**0.001**
9 (3.4%)	4 (1.5%)	7 (1.3%)	0.080
42 (15.7%)	16 (6.0%)	60 (10.8%)	<**0.001**
11.0 (5.5–23.2)	8.2 (3.1–10.9)	11.8 (7.4–23.1)	0.245
3.9 (1.1–8.7)	3.3 (0.8–10.1)	4.7 (1.2–11.9)	
75 (28.0%)	63 (23.5%)	127 (23.0%)	0.413
40 (53.3%)	41 (65.1%)	67 (52.8%)	**0.012**
27 (36.0%)	18 (28.6%)	44 (34.7%)	
8 (10.7%)	4 (6.4%)	16 (12.6%)	

cohort, but this was no longer apparent when accounting for differences in risk of hospitalization. However, an increased risk of mortality in Asians was seen in both the hospitalized cohort and when correcting for the probability of hospitalization. The increased risk of mortality in Asians across the first two waves of the COVID-19 pandemic matches the effect described within other local cohorts [11] and larger national cohorts [3–5]. The consistency of this effect across studies is reassuring, supporting the notion of a relationship between this ethnicity and COVID-19 mortality by an unspecified mechanism. Meanwhile, the decreased risk of death in those who are Black during the first wave of COVID-19 seen prior to correcting for risk of hospitalization is unusual. Community-based studies [3,4,19] have demonstrated an increased risk of mortality in Black ethnic groups in wave 1 specifically. The disparity between this analysis and other studies highlights the issue being addressed here: the use of hospital-based data for opportunistic/retrospective analysis introduces bias into the relationship between ethnicity and COVID-19 outcomes. These patients represent only the most severe cases of COVID-19. Additionally, ethnic minorities within the UK had an increased risk of contracting COVID-19 [4,8] through societal/cultural pressures, meaning these individuals are overrepresented within hospital cohorts.

Table 2. Association of ethnicity with the risk of death if infected with COVID-19 during time periods: Feb 2020–May 2021 (both waves), Feb 2020–Aug 2020 (Wave 1), and Sept 2020–May 2021 (Wave 2)

			Adjusted for:					
	Unadjusted estimates		Sex and age		All covariates		All covariates plus IPW for hospitalization	
Death	**HR (95% CI)**	***P* value**	**HR (95% CI)**	***P* value**	**HR (95% CI)**	***P* value**	**HR (95% CI)**	***P* value**
Both Waves (n = 3,133)								
White (ref)	1	-	1	-	1	-	1	-
Black	0.70 (0.53, 0.92)	**0.012**	0.98 (0.74, 1.30)	0.895	0.91 (0.65, 1.27)	0.575	1.29 (0.87, 1.93)	0.211
Asian	1.23 (0.88, 1.72)	0.234	1.96 (1.39, 2.77)	**<0.001**	1.94 (1.28, 2.93)	**0.002**	2.06 (1.15, 3.67)	**0.014**
Mixed/other	0.45 (0.27, 0.75)	**0.002**	0.79 (0.46, 1.34)	0.382	0.86 (0.48, 1.52)	0.598	0.79 (0.39, 1.64)	0.533
Unknown	0.82 (0.61, 1.11)	0.198	1.15 (0.85, 1.56)	0.352	1.06 (0.76, 1.49)	0.719	0.96 (0.64, 1.44)	0.838
Wave 1 (n = 1,010)								
White (ref)	1	-	1	-	1	-	1	-
Black	0.58 (0.39, 0.87)	**0.008**	0.72 (0.48, 1.07)	0.107	0.63 (0.39, 1.00)	**0.049**	1.06 (0.56, 2.00)	0.851
Asian	1.36 (0.85, 2.15)	0.197	1.85 (1.16, 2.94)	**0.010**	1.73 (0.99, 3.01)	0.052	2.43 (1.05, 5.62)	**0.038**
Mixed/other	0.67 (0.34, 1.30)	0.234	1.09 (0.55, 2.15)	0.810	1.01 (0.46, 2.20)	0.980	0.87 (0.29, 2.62)	0.811
Unknown	0.67 (0.42, 1.06)	0.086	0.79 (0.49, 1.28)	0.331	0.66 (0.39, 1.11)	0.116	0.50 (0.26, 0.96)	**0.039**
Wave 2 (n = 2,123)								
White (ref)	1	-	1	-	1	-	1	-
Black	0.80 (0.54, 1.18)	0.265	1.23 (0.83, 1.84)	0.302	1.23 (0.77, 1.96)	0.380	1.54 (0.90, 2.63)	0.114
Asian	1.13 (0.69, 1.86)	0.627	2.02 (1.21, 3.35)	**0.007**	2.06 (1.10, 3.85)	**0.024**	2.18 (0.99, 4.81)	0.054
Mixed/other	0.32 (0.14, 0.74)	**0.007**	0.59 (0.25, 1.38)	0.222	0.72 (0.29, 1.77)	0.474	0.72 (0.25, 2.03)	0.533
Unknown	1.02 (0.69, 1.50)	0.940	1.59 (1.07, 2.37)	**0.022**	1.69 (1.08, 2.64)	**0.023**	1.61 (0.93, 2.79)	0.091

Abbreviations: CI, confidence interval; IPW, inverse probability weighting.

Unadjusted analysis represents biased estimates restricted to hospitalized populations.

Sequential adjustment allows for consideration of confounding and collider bias.

"All covariates" includes sex, age, medical history (cardiovascular, kidney and liver conditions, chronic obstructive pulmonary disorder/emphysema, diabetes, and DNACPR), and IMD quintile.

P values are derived from univariate Wald tests of the relevant HRs.

Bold *P*-values indicate significance at the 5% level.

Notably here, only 40.4% of the cohort is White (49.0% excluding unknown ethnicity), rather than 61.5% White as estimated for the GSTT catchment area [20]. Accounting for the conditioning on hospitalization has corrected the biased finding of a reduced risk of mortality in Black patients, producing something closer to the effects estimated in community-based studies.

Omitting Asian ethnicity, most associations between ethnic group and COVID-19 outcomes are specific to individual waves. This matches other studies [4,5] and suggests that the relationship between ethnicity and COVID-19 outcomes is unlikely to be driven by biological factors. Societal pressures may be behind this ethnic vulnerability. Government guidelines changed dramatically throughout the pandemic [21,22]. Its onset changed access to healthcare, disproportionately impacting those already experiencing health inequalities. Certain occupations stereotypically associated with ethnic minorities were also greatly impacted at differing timepoints (e.g., during lockdowns) with continued requirements to work outside the home and increased infection rates [23–25]. Even within these occupations, non-White individuals are demonstrated to have a greater risk of COVID-19 infection [26,27], suggesting other contributors, such as social discrimination.

Another element highlighted here is the importance of adjusting for covariates. Unadjusted analysis suggested a reduced risk of death in mixed/other patients compared to White patients. This is due to the older, frequently male White patients with multiple comorbidities – all factors that independently increase the risk of COVID-19-associated mortality [3,28,29]. This analysis is not the first to show this effect of covariate adjustment [2,30] and supports the idea that other patient characteristics are of more relevance than ethnicity. Interestingly, compared to other studies focusing on cohorts hospitalized with COVID-19 [9,10] patients included in this analysis were younger, less likely to be male, less likely to have medical comorbidities, and more ethnically diverse. This is probably due to differences within the local nonhospitalized populations. These regional differences necessitate the careful consideration of covariates relevant to the assessed cohort. Here, this consideration necessitated the inclusion of application of DNACPR orders as an important covariate due to their high prevalence in White patients.

Table 3. Association of ethnicity with the odds of composite ICU admission and death within a given period after COVID-19 infection during time periods: Jan 2020–May 2021 (both waves), Jan 2020–Aug 2020 (Wave 1), and Sept 2020–May 2021 (Wave 2)

Composite ICU admission/death within 30 days	Unadjusted estimates		Adjusted for: Sex and age		Adjusted for: All covariates		Adjusted for: All covariates plus IPW for hospitalization	
	Odds ratio (95% CI)	*P*-value	Odds ratio (95% CI)	*P*-value	Odds ratio (95% CI)	*P*-value	Odds ratio (95% CI)	*P*-value
Both Waves (n = 3,133)								
White (ref)	1	-	1	-	1	-	1	-
Black	0.92 (0.75, 1.12)	0.399	1.11 (0.89, 1.38)	0.346	1.06 (0.84, 1.34)	0.621	1.05 (0.79, 1.41)	0.729
Asian	1.09 (0.81, 1.46)	0.576	1.33 (0.98, 1.79)	0.064	1.27 (0.93, 1.74)	0.136	1.36 (0.73, 2.55)	0.330
Mixed/Other	0.86 (0.63, 1.17)	0.339	1.15 (0.84, 1.58)	0.383	1.30 (0.94, 1.79)	0.115	1.49 (1.02, 2.21)	**0.047**
Unknown	0.83 (0.66, 1.05)	0.130	0.96 (0.76, 1.23)	0.765	1.06 (0.82, 1.38)	0.637	1.07 (0.77, 1.48)	0.681
Wave 1 (n = 1,010)								
White (ref)	1	-	1	-	1	-	1	-
Black	0.92 (0.67, 1.28)	0.637	1.03 (0.74, 1.44)	0.856	0.94 (0.65, 1.37)	0.757	1.21 (0.76, 1.93)	0.428
Asian	1.17 (0.73, 1.90)	0.509	1.28 (0.67, 1.95)	0.311	1.25 (0.75, 2.07)	0.398	1.42 (0.62, 3.25)	0.403
Mixed/Other	0.96 (0.57, 1.63)	0.884	1.14 (0.67, 1.95)	0.630	1.27 (0.74, 2.19)	0.390	1.27 (0.61, 2.67)	0.523
Unknown	0.99 (0.67, 1.45)	0.952	1.07 (0.72, 1.58)	0.741	1.10 (0.71, 1.72)	0.667	1.60 (0.95, 2.69)	0.075
Wave 2 (n = 2,123)								
White (ref)	1	-	1	-	1	-	1	-
Black	0.88 (0.67, 1.16)	0.371	1.13 (0.85, 1.50)	0.397	1.15 (0.84, 1.56)	0.383	1.04 (0.73, 1.48)	0.818
Asian	1.04 (0.71, 1.53)	0.830	1.36 (0.92, 2.01)	0.121	1.32 (0.88, 1.98)	0.177	1.44 (0.68, 3.05)	0.339
Mixed/Other	0.86 (0.58, 1.26)	0.434	1.20 (0.81, 1.80)	0.365	1.36 (0.90, 2.05)	0.140	1.57 (1.00, 2.46)	**0.049**
Unknown	0.78 (0.58, 1.05)	0.105	0.93 (0.68, 1.27)	0.667	1.05 (0.76, 1.45)	0.778	0.96 (0.64, 1.44)	0.852

Abbreviations: CI, confidence interval; IPW, inverse probability weighting.

Unadjusted analysis represents biased estimates restricted to hospitalized populations.

Sequential adjustment allows for consideration of confounding and collider bias.

"All covariates" includes sex, age, medical history (cardiovascular, kidney and liver conditions, chronic obstructive pulmonary disorder/emphysema, diabetes, and DNACPR), and IMD quintile.

P values are derived from univariate Wald tests of the relevant odds ratio.

Bold *P*-values indicate significance at the 5% level.

Notably, this analysis did not include COVID-19 severity at admission as a covariate, despite this factor being a contributor to inducing collider bias (Fig. 1). Secondary outcomes assessing respiratory measures within 24 hours of admission suggested heightened severity in Asian individuals (Supplementary Tables 5–8). But these measures are prone to ethnic discrepancies [31], and their collection differed between ethnicities (Supplementary Table 4), suggesting that the inclusion of these measures as covariates would induce further bias. Likewise, vaccination status is not considered due to lack of data. Timings of the start of the UK vaccination programme mean that this could have impacted severity during wave two and complicated the relationship between ethnicity and severity due to ethnic differences in vaccine uptake [32,33].

Unlike larger national cohort studies [3,4], this analysis did not find significant ethnic differences in the risk of ICU admission and did not find any effect of deprivation on mortality. This may be a result of restricting this analysis to a single, unique population. In addition to the increased ethnic diversity, 48.2% of this patient cohort comes from IMD quintile two, rather than the even spread across all quintiles expected in the national population. By restricting to a semihomogenous population, the effect of deprivation has been minimized. The unique characteristics of this local population may also have knock-on effect on other relationships, such as that between ethnicity and ICU admission.

A limitation to this analysis was in the use of a national cohort, rather than London-based cohort, to correct for risk of hospitalization. Typically, corrections for sample restriction, here hospitalization with COVID-19, are used within nested samples whereby the probability of inclusion into the analyzed cohort is directly known. Here, the probability of hospitalization due to COVID-19 has been estimated from published data from OpenSAFELY [4] – a national database of English primary health care data that allows the increased risk of hospitalization to be identified. We cannot guarantee that the estimates used accurately represent the true probabilities of hospitalization with COVID-19 within this London-based cohort—a population that will have an increased likelihood of contracting COVID-19 compared to rural populations [34,35]. However, assessment of the methodology used demonstrates little change in the results obtained when the probability of hospitalization was adjusted [18]. Likewise, the estimates used are unlikely to reflect the true risk of hospitalization or COVID-

19 outcomes in other countries. National cohorts relevant to each country are needed to apply this methodology to the country of interest.

4.1. Summary

This study has investigated the relationship between ethnicity and COVID-19 outcomes within a South London hospitalized cohort. Acknowledging the inherent biases induced by this restriction, corrections for the probability of hospitalization with COVID-19 were made and found to reduce the observed associations between ethnicity and COVID-19 severity but did not affect the increased risk of mortality in Asian patients. This highlights the importance of considering bias induced by study design, which may impact study results. Causal thinking should be supported by directed acyclic graphs and consideration for confounders and overadjustment/collider bias when assessing restricted cohorts.

Declaration of competing interest

There are no conflicts of interest to declare.

Acknowledgments

We would like to acknowledge Marc Delord who provided additional statistical advice.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2023.06.014.

References

[1] Office of National Statistics. Coronavirus (COVID-19) related deaths by ethnic group, England and Wales. Available at https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/coronaviruscovid19relateddeathsbyethnicgroupenglandandwales/2march2020to15may2020. Accessed October 22, 2022.

[2] Sze S, Pan D, Nevill CR, Gray LJ, Martin CA, Nazareth J, et al. Ethnicity and clinical outcomes in COVID-19: a systematic review and meta-analysis. EClinicalMedicine 2020;29:100630.

[3] Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020;584:430−6.

[4] Mathur R, Rentsch CT, Morton CE, Hulme WJ, Schultze A, MacKenna B, et al. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. Lancet 2021;397:1711−24.

[5] Nafilyan V, Islam N, Mathur R, Ayoubkhani D, Banerjee A, Glickman M, et al. Ethnic differences in COVID-19 mortality during the first two waves of the Coronavirus Pandemic: a nationwide cohort study of 29 million adults in England. Eur J Epidemiol 2021;36(6):605−17.

[6] Razai MS, Kankam HKN, Majeed A, Esmail A, Williams DR. Mitigating ethnic disparities in COVID-19 and beyond. BMJ 2021;372:m4921. https://doi.org/10.1136/BMJ.M4921.

[7] Learoyd AE, Douiri A, Hart N. COVID-19 and ethnicity: has history repeated itself? Thorax 2021;76(6):537−8.

[8] Gershengorn HB, Patel S, Shukla B, Warde PR, Bhatia M, Parekh D, et al. Association of RACE and ethnicity with COVID-19 test positivity and hospitalization is mediated by socioeconomic factors. Ann Am Thorac Soc 2021;18(8):1326−34.

[9] Ferrando-Vivas P, Doidge J, Thomas K, Gould DW, Mouncey P, Shankar-Hari M, et al. Prognostic factors for 30-day mortality in critically ill patients with coronavirus disease 2019. Crit Care Med 2020;49(1):102−11.

[10] Singh BM, Bateman J, Viswanath A, Klaire V, Mahmud S, Nevill A, et al. Risk of COVID-19 hospital admission and COVID-19 mortality during the first COVID-19 wave with a special emphasis on ethnic minorities: an observational study of a single, deprived, multiethnic UK health economy. BMJ Open 2021;11(2):e046556.

[11] Wan YI, Apea VJ, Dhairyawan R, Puthucheary ZA, Pearse RM, Orkin CM, et al. Ethnic disparities in hospitalisation and hospital-outcomes during the second wave of COVID-19 infection in east London. Sci Rep 2022;12(1):3721.

[12] Lassale C, Gaye B, Hamer M, Gale CR, Batty GD. Ethnic disparities in hospitalisation for COVID-19 in England: the role of socioeconomic factors, mental health, and inflammatory and pro-inflammatory factors in a community-based cohort study. Brain Behav Immun 2020;88:44.

[13] 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):1−12.

[14] Ministry of Housing Communities & Local Government. The English Indices of Deprivation 2019 (IoD2019) 2019. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/835115/IoD2019_Statistical_Release.pdf. Accessed June 19, 2023.

[15] Kahan BC, Morris TP, White IR, Tweed CD, Cro S, Dahly D, et al. Treatment estimands in clinical trials of patients hospitalised for COVID-19: ensuring trials ask the right questions. BMC Med 2020;18:286.

[16] Xue X, Saeed O, Castagna F, Jorde UP, Agalliu I. The analysis of COVID-19 in-hospital mortality: a competing risk approach or a cure model? Stat Methods Med Res 2022;31(10):1976−91.

[17] Thompson CA, Arah OA. Selection bias modeling using observed data augmented with imputed record-level probabilities. Ann Epidemiol 2014;24:747−53.

[18] Learoyd AE, Nicholas J, Hart N, Douiri A. Application of information from external data to correct for collider bias in a Covid-19 hospitalised cohort. BMC Med Res Methodol 2023. in press.

[19] Elliott J, Bodinier B, Whitaker M, Delpierre C, Vermeulen R, Tzoulaki I, et al. COVID-19 mortality in the UK Biobank cohort: revisiting and evaluating risk factors. Eur J Epidemiol 2021;36(3):299−309.

[20] Office of National Statistics. Population by Ethnic Group by borough - ONS Annual Population Survey. 2022. Available at https://data.london.gov.uk/dataset/ethnic-groups-borough. Accessed July 27, 2022.

[21] Morrissey K, Spooner F, Salter J, Shaddick G. Area level deprivation and monthly COVID-19 cases: the impact of government policy in England. Soc Sci Med 2021;289:114413.

[22] Timeline of UK government coronavirus lockdowns and restrictions - The Institute for Government. 2022. Available at https://www.instituteforgovernment.org.uk/charts/uk-government-coronavirus-lockdowns. Accessed August 18, 2022.

[23] Anand P, Allen HL, Ferrer RL, Gold N, Gonzales Martinez RM, Kontopentelis E, et al. Work-related and personal predictors of COVID-19 transmission: evidence from the UK and USA. J Epidemiol Community Health 2022;76:152−7.

[24] Rhodes S, Wilkinson J, Pearce N, Mueller W, Cherrie M, Stocking K, et al. Occupational differences in SARS-CoV-2 infection: analysis of

the UK ONS coronavirus (COVID-19) infection survey. J Epidemiol Community Health 2022;76:841−6.

[25] Atchison C, Bowman LR, Vrinten C, Redd R, Pristerà P, Eaton J, et al. Early perceptions and behavioural responses during the COVID-19 pandemic: a cross-sectional survey of UK adults. BMJ Open 2021;11(1):e043577.

[26] Shields A, Faustini SE, Perez-Toledo M, Jossi S, Aldera E, Allen JD, et al. SARS-CoV-2 seroprevalence and asymptomatic viral carriage in healthcare workers: a cross-sectional study. Thorax 2020;75(12):1089−94.

[27] Mutambudzi M, Niedwiedz C, Macdonald EB, Leyland A, Mair F, Anderson J, et al. Occupation and risk of severe COVID-19: prospective cohort study of 120 075 UK Biobank participants. Occup Environ Med 2021;78(5):307−14.

[28] Biswas M, Rahaman S, Biswas TK, Haque Z, Ibrahim B. Association of sex, age, and comorbidities with mortality in COVID-19 patients: a systematic review and meta-analysis. Intervirology 2021;64(1):36−47.

[29] Gerayeli FV, Milne S, Cheund C, Li X, Yang CWT, Tam A, et al. COPD and the risk of poor outcomes in COVID-19: a systematic review and meta-analysis. EClinicalMedicine 2021;33:100789.

[30] Harrison E, Docherty A, Semple C. Investigating associations between ethnicity and outcome from COVID-19. 2020. Available at https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-. Accessed May 8, 2022.

[31] Knight M, Subbe C, Inada-Kim M. Racial discrepancies in oximetry: where do we stand? Anaesthesia 2022;77(2):129−31.

[32] Dolby T, Finning K, Baker A, Fowler-Dowd L, Khunti K, Razieh C, et al. Monitoring sociodemographic inequality in COVID-19 vaccination uptake in England: a national linked data study. J Epidemiol Community Health 2022;76(7):646−52.

[33] Watkinson RE, Williams R, Gillibrand S, Sanders C, Sutton M. Ethnic inequalities in COVID-19 vaccine uptake and comparison to seasonal influenza vaccine uptake in Greater Manchester, UK: a cohort study. PLoS Med 2022;19(3):e1003932.

[34] Ali N, Fariha KA, Islam F, Mishu MA, Mohanto NC, Hosen MJ, et al. Exposure to air pollution and COVID-19 severity: a review of current insights, management, and challenges. Integr Environ Assess Manag 2021;17(6):1114−22. https://doi.org/10.1002/ieam.4435.

[35] Lai KY, Webster C, Kumari S, Sarkar C. The nature of cities and the Covid-19 pandemic. Curr Opin Environ Sustain 2020;46:27−31.