Revisiting ethnic discrepancies in Covid-19 hospitalised cohorts: a correction for collider bias

Annastazia E Learoyd[1], Jennifer Nicholas[2], Nicholas Hart[3,4] and Abdel Douiri[1]

[1] School of Life Course and Population Sciences, King College London, London, UK

[2] Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

[3] Lane Fox Clinical Respiratory Physiology Research Centre, Guy's & St Thomas' NHS Foundation Trust, London, UK

[4] Centre for Human and Applied Physiological Sciences, King's College London, UK

Corresponding author:

Dr Annastazia E Learoyd

annastazia.learoyd@kcl.ac.uk

Present Address:
School of Life Course and Population Sciences
Faculty of Life Sciences and Medicine
4th Floor, Addison House
King's College London; Guy's Campus
London SE1 1UL, UK

Abstract

Objective: 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 potential driven by bias from analysing hospitalised 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 were examined using regression models. Three iterations of each model were completed: 1) an unadjusted analysis, 2) adjusting for covariates (medical history and deprivation), 3) adjusting for covariates and bias induced by conditioning on hospitalisation.

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 hospitalisation. However, wave-specific effects demonstrate significant differences between ethnic groups until bias from using a hospitalised cohort was corrected for.

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

Key words: Collider bias, Covid-19, Epidemiology, Ethnicity, Inequalities, Statistical methods

Running title: Bias in Covid-19 cohort studies

Word count: 3,000

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 into ethnic differences within Covid-19 focus on hospitalised patients[9]–[12] due to the easily accessible data contained in electronic health records (EHR). This allowed studies to be completed within months of the discovery 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 utilising hospitalised 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 (Figure 1A). In Covid-19, severity is a potential mediator between ethnicity and increased mortality. Hospitalisation is a descending proxy for severity, and therefore conditioning on this variable through use of a hospitalisation cohort means studies are at risk of overadjustment bias. Secondly, differences in risk of hospitalisation with Covid-19 have been noted between ethnic groups[4], [12], [13]. If this association is via paths other than differences in Covid-19 severity, then hospitalisation becomes a collider[14] – a variable influenced by both the exposure and a mediator/outcome and when conditioned on induces an association (Figure 1B). Both biases can mean that the associations seen in hospitalised 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 hospitalised cohort.

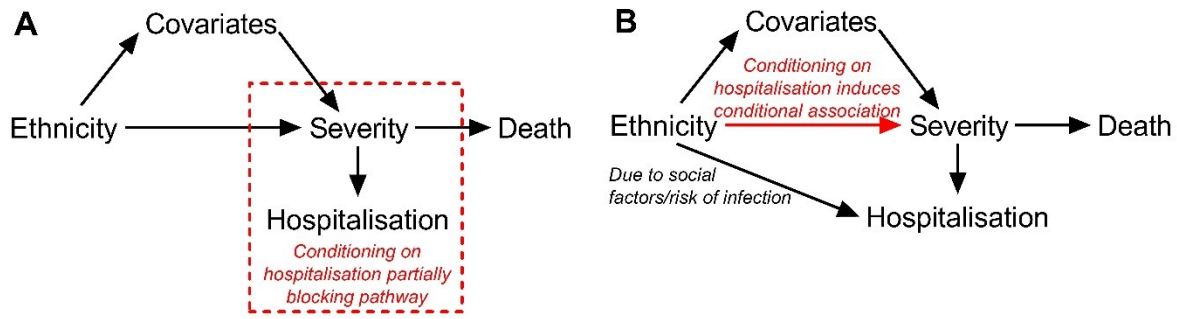

Figure 1: Directed acyclic graph demonstrating how the relationship between ethnicity and Covid-19 severity/mortality may be influenced by (A) overadjustment bias and (B) collider bias. Overadjustment bias will reduce the association between ethnicity and Covid-19 mortality. Collider bias will induce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

2. Methods

2.1 Patient population

5,992 patients were admitted to two hospitals within the Guy's and St. Thomas' NHS Foundation Trust (GSTT) and received a Covid-19 positive test between 20th February 2020 and 24th May 2021.

The following patients were excluded from this analysis (Figure 2): those without a known date of admission and discharge; those with admission prior to 28th January 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 NHS trusts for a higher level of care; those under 18 years old; and those without a known Index of Multiple Deprivation (IMD) quintile. The analysed cohort included 3,133 patients (52.3% of the initial population).

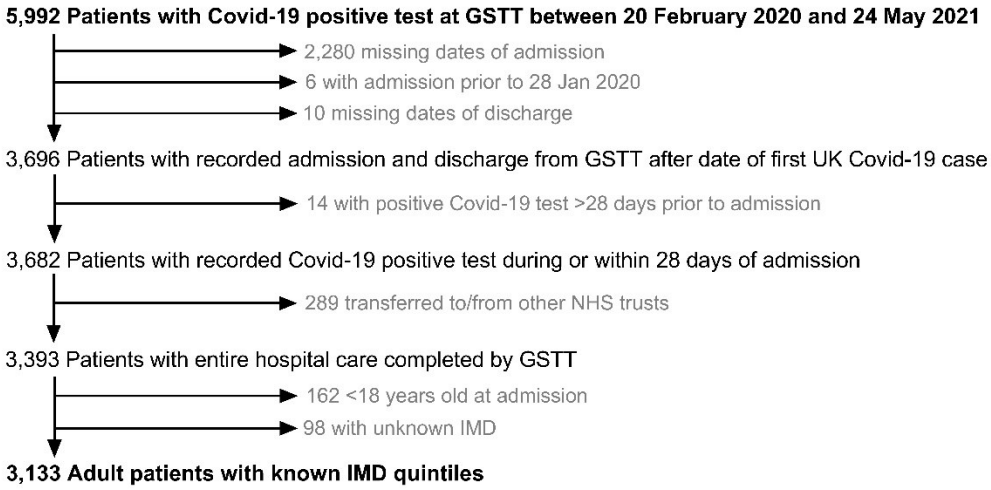

Figure 2: Study population flowchart. GSTT= Guy's and St. Thomas' NHS Foundation Trust. IMD= Index of Multiple Deprivation.

2.2 Data sources

Anonymised clinical, laboratory and demographic data for patients with a positive RT-PCR 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 3.6.3 and Python 3.7.6. The study authors did not have access to the databases used in initial data linkage.

Patient-level data and programming code is unavailable due to infostructure changes in August 2022. Further summarised 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 Modelling COVID-19 Group to collect clinically relevant data points from patient's EHR. 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, Other), Black (African, 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 centred 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 7 domains of deprivation[15]. Patient addresses were linked to IMD and organised into quintiles with one 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 ICD10 codes or automated searches of free text data. Patients werecategorised as having/not having a medical history of cardiovascular disease (stroke, transient ischaemic attack, atrial fibrillation, congestive heart failure, ischaemic 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 Figure 1).

The application of DNARCPR orders was extracted along with the date of application as an indicator of level of care. DNARCPR was treated as a time-dependent covariate in analysis of mortality and a binary covariate of application/no application by 30 days for analysis of ICU admission/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[16]. Secondary outcomes examining respiratory measures recorded within 24hrs 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 analysed using a competing risk regression model with discharge as a competing event[17]. 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[16] was analysed using logistic regression. Linear predictor specification and model fit was 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 using inverse probability weighting (IPW) to account for conditioning on hospitalisation[18]. All models were completed for each outcome using data from all patients, those admitted during wave one of Covid-19 (February 2020-August 2020), and those admitted during wave two (September 2020-May 2021).

IPWs were calculated from estimated probabilities of hospitalisation with Covid-19 based on each individuals ethnicity, admission date, and survival status. Full details of this weighting method are published separately[19]. The probabilities of hospitalisation were derived using results from analysis of Covid-19 outcomes in a UK national cohort between 1st February and 31st December 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 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 one comorbidity of interest, most commonly cardiovascular conditions, followed by diabetes and chronic kidney disease. 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 towards a difference in ethnic distributions between Covid-19 waves (p=0.061) with a higher proportion of Black people in wave one and more people from Mixed/Other and Unknown ethnicities in wave two. Patient characteristics in each wave are described in Supplementary Tables 1&2.

Patient Characteristics	All patients	Ethnicity					Comparison
		White	Black	Asian	Mixed/Other	Unknown	p-value
Total num. (%) of patients	3,133	1,265 (40.4%)	779 (24.9%)	268 (8.6%)	268 (8.6%)	553 (17.7%)	
Num. (%) patients - Wave 1	1,010	415 (41.1%)	274 (27.1%)	87 (8.6%)	72 (7.1%)	162 (16.0%)	
Num. (%) patients - Wave 2	2,123	850 (40.0%)	505 (23.8%)	181 (8.5%)	196 (9.2%)	391 (18.4%)	
Covariates							
Index of Multiple Deprivation (IMD)							
Rank [Median (IQR)]	9,324 (6591-13,839)	9,929 (7,088-14,996)	8,105 (6,008-11,635)	8,161 (6,422-12,982)	8,894 (6,731-12,694)	9,939 (6,731-15,898)	**<0.001**
Quintile (%): 1 (Most)	775 (24.7%)	267 (21.1%)	236 (30.3%)	75 (28.0%)	64 (23.9%)	133 (24.1%)	**<0.001**
2	1,528 (48.8%)	617 (48.8%)	402 (51.6%)	128 (47.8%)	141 (52.6%)	240 (43.4%)	
3	498 (15.9%)	221 (17.5%)	111 (14.2%)	40 (14.9%)	38 (14.2%)	88 (15.9%)	
4	227 (7.2%)	100 (7.9%)	27 (3.5%)	19 (7.1%)	19 (7.1%)	62 (11.2%)	
5 (Least)	105 (3.4%)	60 (4.7%)	3 (0.4%)	6 (2.2%)	6 (2.2%)	30 (5.4%)	
Age (Mean±SD)	59.71±18.80	64.06±18.81	57.56±18.33	56.13±17.71	53.38±17.98	57.58±18.29	**<0.001**
Male Sex (%)	1,674 (53.4%)	691 (54.6%)	369 (47.4%)	153 (57.1%)	134 (50.0%)	327 (59.1%)	**<0.001**
DNARCPR applied							
Number (%)	693 (22.1%)	348 (27.5%)	139 (17.8%)	56 (20.9%)	32 (11.9%)	118 (21.3%)	**<0.001**
Time to DNARCPR (days)	1.3 (0.2-9.8)	1.1 (0.2-10.3)	1.2 (0.3-8.5)	1.8 (0.2-15.2)	0.8 (0.2-5.3)	2.8 (0.3-11.0)	0.367
Cardiovascular conds (%)	1,710 (54.6%)	776 (61.3%)	454 (58.3%)	136 (50.7%)	123 (45.9%)	221 (40.0%)	**<0.001**
COPD/Emphysema (%)	262 (8.4%)	176 (13.9%)	27 (3.5%)	9 (3.4%)	18 (6.7%)	32 (5.8%)	**<0.001**
Diabetes (%)	875 (27.9%)	326 (25.8%)	276 (35.4%)	96 (35.8%)	56 (20.9%)	121 (21.9%)	**<0.001**
Kidney conditions (%)	631 (20.1%)	263 (20.8%)	210 (27.0%)	53 (19.8%)	32 (11.9%)	73 (13.2%)	**<0.001**
Liver conditions (%)	82 (2.6%)	42 (3.3%)	20 (2.6%)	9 (3.4%)	4 (1.5%)	7 (1.3%)	0.080
Death during hospital stay							
Number (%)	356 (11.4%)	166 (13.1%)	72 (9.2%)	42 (15.7%)	16 (6.0%)	60 (10.8%)	**<0.001**
Time to death (days)	11.1 (5.7-21.8)	11.8 (5.9-22.2)	10.8 (5.3-18.9)	11.0 (5.5-23.2)	8.2 (3.1-10.9)	11.8 (7.4-23.1)	0.245
Time to censor (days)	5.0 (1.4-12.7)	6.1 (1.9-14.9)	4.8 (1.4-12.6)	3.9 (1.1-8.7)	3.3 (0.8-10.1)	4.7 (1.2-11.9)	
ICU admission/Death within 30 days							
Total Number (%)	790 (25.2%)	333 (26.3%)	192 (24.7%)	75 (28.0%)	63 (23.5%)	127 (23.0%)	0.413
Contribution: ICU admission	428 (54.2%)	160 (48.1%)	120 (62.5%)	40 (53.3%)	41 (65.1%)	67 (52.8%)	**0.012**
Death	251 (31.8%)	120 (36.0%)	42 (21.9%)	27 (36.0%)	18 (28.6%)	44 (34.7%)	
Both	111 (14.1%)	53 (15.9%)	30 (15.6%)	8 (10.7%)	4 (6.4%)	16 (12.6%)	

Table 1: Patient characteristics by ethnic group. 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). IQR=Interquartile Range. DNARCPR= “do not attempt resuscitation” order. COPD=chronic obstructive pulmonary disease.

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 Figure 2A). Survival curves for each ethnic group (Supplementary Figure 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 (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 non-significant 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 was 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 hospitalisation 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 DNARCPR 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 or IMD quintile were contributing factors (Supplementary Table 4), after accounting for other predictors and correcting for conditioning on hospitalisation.

Death	Unadjusted estimates		*Adjusted for:* Sex and age		All covariates		All covariates plus IPW for hospitalisation	
	Hazard ratio (95% CI)	p-value	Hazard ratio (95% CI)	p-value	Hazard ratio (95% CI)	p-value	Hazard ratio (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

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). Unadjusted analysis represents biased estimates restricted to hospitalised populations. Sequential adjustment allows for consideration of confounding and collider bias. “All covariates” includes sex, age, medical history (cardiovascular, kidney and liver conditions, COPD/emphysema, diabetes, and DNARCPR), and IMD quintile. P-values are derived from univariate Wald tests of the relevant hazard ratio. CI=Confidence Interval. IPW=Inverse probability weighting.

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 Figure 3). During hospitalisation, 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 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 hospitalisation 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 DNARCPR orders (OR=3.80 (3.03, 4.77), p<0.001) were.

Composite ICU admission/ death within 30 days	Unadjusted estimates		*Adjusted for:* Sex and age		All covariates		All covariates plus IPW for hospitalisation	
	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

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). Unadjusted analysis represents biased estimates restricted to hospitalised populations. Sequential adjustment allows for consideration of confounding and collider bias. “All covariates” includes sex, age, medical history (cardiovascular, kidney and liver conditions, COPD/emphysema, diabetes, and DNARCPR), and IMD quintile. P-values are derived from univariate Wald tests of the relevant odds ratio. CI=Confidence Interval. IPW=Inverse probability weighting.

4. Discussion

This study found that use of a hospitalised cohort influenced the magnitude and direction of association between ethnicity and Covid-19 outcomes. A reduction in mortality in Blacks in wave one was seen in the hospitalised cohort, but this was no longer apparent when accounting for differences in risk of hospitalisation. However, an increased risk of mortality in Asians was seen in both the hospitalised cohort and when correcting for probability of hospitalisation. 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[12] 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 hospitalisation is unusual. Community-based studies[3], [4], [20] have demonstrated an increased risk of mortality in Black ethnic groups in wave one 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. 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[21]. Accounting for the conditioning on hospitalisation 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[22], [23]. 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[24]–[26]. Even within these occupations, non-White individuals are demonstrated to have a greater risk of Covid-19 infection[27], [28] 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 which independently increase the risk of Covid-19 associated mortality[3], [29], [30]. This analysis is not the first to show this effect of covariate adjustment[2], [31] and supports the idea that other patient characteristics are of more relevance than ethnicity. Interestingly, compared to other studies focusing on cohorts hospitalised with Covid-19[10], [11], 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 non-hospitalised populations. These regional differences necessitate the careful consideration of covariates relevant to the assessed cohort. Here, this consideration necessitated the inclusion of application of DNARCPR orders as an important covariate due to their high prevalence in White patients.

Notably this analysis did not include Covid-19 severity at admission as a covariate – despite this factor being a contributor to inducing collider bias (Figure 1). Secondary outcomes assessing respiratory measures within 24hrs of admission suggested heightened severity in Asian individuals (Supplementary Table 5-8). But these measures are prone to ethnic discrepancies[32] 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[33], [34].

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 come from IMD quintile two, rather than the even spread across all quintiles expected in the national population. By restricting to a semi-homogenous population, the effect of deprivation has been minimised. 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 of this analysis was in the use of a national cohort, rather than London-based cohort, to correct for risk of hospitalisation. Typically, corrections for sample restriction, here hospitalisation with Covid-19, are used within nested samples whereby the probability of inclusion into the analysed cohort is directly known. Here, the probability of hospitalisation

due to Covid-19 has been estimated from OPENSAFELY[4] – a national database of primary healthcare data that allows the increased risk of hospitalisation to be identified. We cannot guarantee that the estimates used accurately represent the true probabilities of hospitalisation with Covid-19 within this London-based cohort – a population that will have an increased likelihood of contracting Covid-19 compared to rural populations[35], [36]. However, assessment of the methodology used demonstrates little change in the results obtained when the probability of hospitalisation was adjusted[19]. Likewise, the estimates used are unlikely to reflect the true risk of hospitalisation or Covid-19 outcomes in other countries. National cohorts relevant to each country 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 hospitalised cohort. Acknowledging inherent biases induced by this restriction, corrections for the probability of hospitalisation 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.

Funding

The authors acknowledge funding support from the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) 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. All other authors reviewed the manuscript.

Acknowledgements

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

Competing interest

There are no conflicts of interest to declare.

References

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

[2] S. Sze *et al.*, "Ethnicity and clinical outcomes in COVID-19: A systematic review and meta-analysis," *EClinicalMedicine*, vol. 29–30, p. 100630, 2020, doi: 10.1016/j.eclinm.2020.100630.

[3] E. J. Williamson *et al.*, "Factors associated with COVID-19-related death using OpenSAFELY," *Nature*, vol. 584, no. 7821, pp. 430–436, 2020, doi: 10.1038/s41586-020-2521-4.

[4] R. Mathur *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*, vol. 397, no. 10286, pp. 1711–1724, May 2021, doi: 10.1016/S0140-6736(21)00634-6/ATTACHMENT/4CDE4253-9BAB-4607-86DC-F915D324B547/MMC1.PDF.

[5] V. Nafilyan *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.*, vol. 36, no. 6, pp. 605–617, Jun. 2021, doi: 10.1007/S10654-021-00765-1/FIGURES/1.

[6] M. S. Razai, H. K. N. Kankam, A. Majeed, A. Esmail, and D. R. Williams, "Mitigating ethnic disparities in covid-19 and beyond," *BMJ*, vol. 372, p. m4921, Jan. 2021, doi: 10.1136/BMJ.M4921.

[7] A. E. Learoyd, A. Douiri, and N. Hart, "COVID-19 and ethnicity: has history repeated itself?," *Thorax*, vol. 76, no. 6, pp. 537–538, Jun. 2021, doi: 10.1136/THORAXJNL-2021-216992.

[8] H. B. Gershengorn *et al.*, "Association of RACE and ethnicity with covid-19 test positivity and hospitalization is mediated by socioeconomic factors," *Ann. Am.*

Thorac. Soc., vol. 18, no. 8, pp. 1326–1334, Aug. 2021, doi: 10.1513/ANNALSATS.202011-1448OC/SUPPL_FILE/DISCLOSURES.PDF.

[9] M. Dashtban *et al.*, "Predicting and Validating Risk of Pre-Pandemic and Excess Mortality in Individuals With Chronic Kidney Disease," *SSRN Electron. J.*, Dec. 2021, doi: 10.2139/SSRN.3970707.

[10] P. Ferrando-Vivas *et al.*, "Prognostic Factors for 30-Day Mortality in Critically Ill Patients With Coronavirus Disease 2019," *Crit. Care Med.*, vol. 49, no. 1, pp. 102–111, 2020, doi: 10.1097/ccm.0000000000004740.

[11] B. M. Singh *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*, vol. 11, no. 2, p. e046556, Feb. 2021, doi: 10.1136/BMJOPEN-2020-046556.

[12] Y. I. Wan *et al.*, "Ethnic disparities in hospitalisation and hospital-outcomes during the second wave of COVID-19 infection in east London," *Sci. Rep.*, vol. 12, no. 1, p. 3721, Dec. 2022, doi: 10.1038/S41598-022-07532-6.

[13] C. Lassale, B. Gaye, M. Hamer, C. R. Gale, and G. D. Batty, "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.*, vol. 88, p. 44, Aug. 2020, doi: 10.1016/J.BBI.2020.05.074.

[14] G. J. Griffith *et al.*, "Collider bias undermines our understanding of COVID-19 disease risk and severity," *Nat. Commun. 2020 111*, vol. 11, no. 1, pp. 1–12, Nov. 2020, doi: 10.1038/s41467-020-19478-2.

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

[16] B. C. Kahan *et al.*, "Treatment estimands in clinical trials of patients hospitalised for COVID-19: ensuring trials ask the right questions," *BMC Med.*, vol. 18, p. 286, Sep.

2020, doi: 10.1186/S12916-020-01737-0.

[17] X. Xue, O. Saeed, F. Castagna, U. P. Jorde, and I. Agalliu, “The analysis of COVID-19 in-hospital mortality: A competing risk approach or a cure model?:,” *Stat. Methods Med. Res.*, p. ., Jun. 2022, doi: 10.1177/09622802221106300.

[18] C. A. Thompson and O. A. Arah, “Selection bias modeling using observed data augmented with imputed record-level probabilities,” *Ann. Epidemiol.*, vol. 24, no. 10, pp. 747–753, Oct. 2014, doi: 10.1016/J.ANNEPIDEM.2014.07.014.

[19] A. E. Learoyd, J. Nicholas, N. Hart, and A. Douiri, “Application of information from external data to correct for collider bias in a Covid-19 hospitalised cohort,” *Res. Sq.*, 2023.

[20] J. Elliott *et al.*, “COVID-19 mortality in the UK Biobank cohort: revisiting and evaluating risk factors,” *Eur. J. Epidemiol.*, vol. 36, no. 3, pp. 299–309, Mar. 2021, doi: 10.1007/S10654-021-00722-Y/FIGURES/2.

[21] Office of National Statistics, “Population by Ethnic Group by borough - ONS Annual Population Survey,” *2022*. https://data.london.gov.uk/dataset/ethnic-groups-borough (accessed Jul. 27, 2022).

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

[23] “Timeline of UK government coronavirus lockdowns and restrictions - The Institute for Government,” *2022*. https://www.instituteforgovernment.org.uk/charts/uk-government-coronavirus-lockdowns (accessed Aug. 18, 2022).

[24] P. Anand *et al.*, “Work-related and personal predictors of COVID-19 transmission: evidence from the UK and USA,” *J Epidemiol Community Heal.*, vol. 76, pp. 152–157, 2022, doi: 10.1136/jech-2020-215208.

[25] S. Rhodes *et al.*, “Occupational differences in SARS-CoV-2 infection: Analysis of the UK ONS Coronavirus (COVID-19) Infection Survey,” *J Epidemiol Community Heal.*, vol. 0, pp. 1–6, 2022, doi: 10.1136/jech-2022-219101.

[26] C. Atchison *et al.*, "Early perceptions and behavioural responses during the COVID-19 pandemic: a cross-sectional survey of UK adults," *BMJ Open*, vol. 11, no. 1, p. e043577, Jan. 2021, doi: 10.1136/BMJOPEN-2020-043577.

[27] A. Shields *et al.*, "SARS-CoV-2 seroprevalence and asymptomatic viral carriage in healthcare workers: A cross-sectional study," *Thorax*, vol. 75, no. 12, pp. 1089–1094, 2020, doi: 10.1136/thoraxjnl-2020-215414.

[28] M. Mutambudzi *et al.*, "Occupation and risk of severe COVID-19: prospective cohort study of 120 075 UK Biobank participants," *Occup. Environ. Med.*, vol. 78, no. 5, pp. 307–314, May 2021, doi: 10.1136/OEMED-2020-106731.

[29] M. Biswas, S. Rahaman, T. K. Biswas, Z. Haque, and B. Ibrahim, "Association of Sex, Age, and Comorbidities with Mortality in COVID-19 Patients: A Systematic Review and Meta-Analysis," *Intervirology*, vol. 64, no. 1, pp. 36–47, Jan. 2021, doi: 10.1159/000512592.

[30] F. V. Gerayeli *et al.*, "COPD and the risk of poor outcomes in COVID-19: A systematic review and meta-analysis," *EClinicalMedicine*, vol. 33, p. 100789, Mar. 2021, doi: 10.1016/J.ECLINM.2021.100789/ATTACHMENT/F03D048F-3DBC-4E0B-B45B-4BB63E42A0EA/MMC2.PDF.

[31] E. Harrison, A. Docherty, and C. Semple, "Investigating associations between ethnicity and outcome from COVID-19," 2020, Accessed: May 08, 2022. [Online]. Available: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-

[32] M. Knight, C. Subbe, and M. Inada-Kim, "Racial discrepancies in oximetry: where do we stand?," *Anaesthesia*, vol. 77, no. 2, pp. 129–131, 2022, [Online]. Available: https://www.researchgate.net/profile/Matthew-Knight-11/publication/356620177_Racial_discrepancies_in_oximetry_where_do_we_stand/links/61adea6150e22929cd4eb8ed/Racial-discrepancies-in-oximetry-where-do-we-stand.pdf

[33] T. Dolby *et al.*, "Monitoring sociodemographic inequality in COVID-19 vaccination uptake in England: a national linked data study," *J Epidemiol Community Heal.*, vol.

76, no. 7, pp. 646–652, Jul. 2022, doi: 10.1136/JECH-2021-218415.

[34] R. E. Watkinson, R. Williams, S. Gillibrand, C. Sanders, and M. Sutton, “Ethnic inequalities in COVID-19 vaccine uptake and comparison to seasonal influenza vaccine uptake in Greater Manchester, UK: A cohort study,” *PLOS Med.*, vol. 19, no. 3, p. e1003932, Mar. 2022, doi: 10.1371/JOURNAL.PMED.1003932.

[35] N. Ali *et al.*, “Exposure to air pollution and COVID-19 severity: A review of current insights, management, and challenges,” *Integr. Environ. Assess. Manag.*, vol. 17, no. 6, pp. 1114–1122, 2021, doi: 10.1002/ieam.4435.

[36] K. Y. Lai, C. Webster, S. Kumari, and C. Sarkar, “The nature of cities and the Covid-19 pandemic,” *Curr. Opin. Environ. Sustain.*, vol. 46, pp. 27–31, Oct. 2020, doi: 10.1016/J.COSUST.2020.08.008.