Social determinants of ethnic disparities in SARS-CoV-2 infection: UK Biobank SARS-CoV-2 Serology Study

Wemimo Omiyale,[1] Jo Holliday,[1,2] Nicola Doherty,[2] Howard Callen,[1,2] Natasha Wood ,[1,2] Edward Horn,[1,2] Frances Burnett,[1] Allen Young,[1,2] Sarah Lewington,[1,3] Dan Fry,[2] Jelena Bešević ,[1] Megan Conroy ,[1] Simon Sheard,[2] Qi Feng,[1] Samantha Welsh,[2] Mark Effingham,[2] Alan Young,[1,2] Rory Collins,[1,2] Ben Lacey ,[1,2] Naomi Allen[1,2]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jech-2023-220353).

[1]Nuffield Department of Population Health, University of Oxford, Oxford, UK
[2]UK Biobank, Stockport, UK
[3]MRC Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Correspondence to
Dr Ben Lacey, Nuffield Department of Population Health, University of Oxford, Oxford OX1 3AZ, UK; ben.lacey@ndph.ox.ac.uk

WO and JH contributed equally. BL and NA contributed equally.

Received 17 March 2023
Accepted 25 August 2023
Published Online First 12 September 2023

To cite: Omiyale W, Holliday J, Doherty N, *et al*. *J Epidemiol Community Health* 2024;**78**:3–10.

ABSTRACT

Background The social determinants of ethnic disparities in risk of SARS-CoV-2 infection during the first wave of the pandemic in the UK remain unclear.

Methods In May 2020, a total of 20 195 adults were recruited from the general population into the UK Biobank SARS-CoV-2 Serology Study. Between mid-May and mid-November 2020, participants provided monthly blood samples. At the end of the study, participants completed a questionnaire on social factors during different periods of the pandemic. Logistic regression yielded ORs for the association between ethnicity and SARS-CoV-2 immunoglobulin G antibodies (indicating prior infection) using blood samples collected in July 2020, immediately after the first wave.

Results After exclusions, 14 571 participants (mean age 56; 58% women) returned a blood sample in July, of whom 997 (7%) had SARS-CoV-2 antibodies. Seropositivity was strongly related to ethnicity: compared with those of White ethnicity, ORs (adjusted for age and sex) for Black, South Asian, Chinese, Mixed and Other ethnic groups were 2.66 (95% CI 1.94–3.60), 1.66 (1.15–2.34), 0.99 (0.42–1.99), 1.42 (1.03–1.91) and 1.79 (1.27–2.47), respectively. Additional adjustment for social factors reduced the overall likelihood ratio statistics for ethnicity by two-thirds (67%; mostly from occupational factors and UK region of residence); more precise measurement of social factors may have further reduced the association.

Conclusions This study identifies social factors that are likely to account for much of the ethnic disparities in SARS-CoV-2 infection during the first wave in the UK, and highlights the particular relevance of occupation and residential region in the pathway between ethnicity and SARS-CoV-2 infection.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Previous studies have found that ethnic disparities in risk of SARS-CoV-2 infection are not adequately explained by differences in underlying health conditions, health-related behaviours (including smoking and obesity) or various social factors (including household size, occupation and socioeconomic deprivation). Such studies have suggested that investigation of alternative biological susceptibilities, as well as a more comprehensive assessment of social and behavioural factors influencing possible exposure to SARS-CoV-2, is warranted.

WHAT THIS STUDY ADDS

⇒ The current study examined the association of ethnicity with risk of SARS-CoV-2 infection during the first wave of the COVID-19 pandemic in the UK, using detailed assessments of social factors to investigate which, if any, might explain the ethnic differences in risk of infection. The study found that social factors (particularly occupational factors and UK region of residence), rather than genetic predisposition or other biological susceptibilities, are likely to account for higher infection rates of SARS-CoV-2 among minority ethnic groups during the first wave of the pandemic in the UK.

INTRODUCTION

Many ethnic minority groups in the UK and elsewhere are at increased risk of SARS-CoV-2 infection, and subsequent hospitalisation and death from coronavirus disease (COVID-19), compared with White ethnic groups.[1–6] Prior to widespread vaccination, representative surveys in England found the prevalence of antibodies to SARS-CoV-2 (ie, indicating prior infection) among some ethnic minority groups, including people of Black or South Asian ethnicity, to be twofold to threefold higher than among White ethnic groups.[1 7]

Proposed explanations for these ethnic disparities have included differences in major comorbidities, health-related behaviours (including smoking and obesity) and social factors (such as household size, employment and living in areas of high socioeconomic deprivation). However, previous large-scale studies have found only modest impacts of these factors on the association of ethnicity with risk of SARS-CoV-2 infection.[1 2 8] As a result, there have been widespread calls for further investigation into alternative biological susceptibilities, as well as more comprehensive assessments of the social and

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This large population-based study indicates the primacy of social factors, as opposed to any biological susceptibility, in driving ethnic disparities in risk of SARS-CoV-2 infection. In addition to addressing the inequalities in social opportunity that ultimately underlie the differences in infection rates between ethnic groups, this study indicates that policies to mitigate the specific vulnerabilities of different regions (including major urban conurbations, given the higher proportion of the population from ethnic minority groups in such regions of the UK), as well as policies aimed at preventing occupational exposure to major respiratory infections, are likely to be particularly important in ensuring future pandemics do not further exacerbate existing ethnic inequalities in health.

behavioural determinants of the ethnic differences in the risk of SARS-CoV-2 infection.[8–10]

Understanding the key mechanisms driving ethnicity disparities in SARS-CoV-2 infection is crucial to mitigate the current and future impact of the COVID-19 pandemic (and future pandemics) on population health and health inequalities. The current study examines the association of ethnicity with the risk of SARS-CoV-2 infection in a large-scale population-based study in the UK, and uses detailed assessments of social factors to investigate the potential pathways between ethnicity and SARS-CoV-2 infection risk.

METHODS

Study design and data collection

The UK Biobank SARS-CoV-2 Serology Study is a longitudinal study of a subset of UK Biobank participants and their relatives (children or grandchildren >18 years), which aimed to assess the prevalence and determinants of SARS-CoV-2 infection across different subgroups in the UK. Details of UK Biobank and its data collection have been reported previously.[11] In brief, UK Biobank is a prospective cohort study of 502 000 adults, recruited from the general population in England, Scotland and Wales between 2006 and 2010. Eligible participants were men and women aged 40–69 years, who were registered with the National Health Service (NHS) and living within approximately 25 miles of the 22 assessment centres; centres were distributed throughout the UK in settings to allow heterogeneity in socioeconomic, ethnic and urban-rural characteristics.[12] Overall, 9.2 million individuals were invited, a response rate of 5.5%. Information on lifestyle and health-related characteristics was collected at recruitment, and health outcomes have been identified over time through linkage to electronic health records.

Between 7 May and 8 June 2020, all UK Biobank participants with a valid email address, resident in mainland UK and who had previously indicated willingness to be contacted about research activities were invited to join the Serology Study (see online supplemental figure S1). UK Biobank participants who had not provided a valid email address were able to view details of the study and register an interest via the UK Biobank participant website. Participants were asked to consent to taking a series of capillary blood samples and to answer questionnaires about potential symptoms of COVID-19. They were also asked to indicate if they would be willing to forward an email invitation from UK Biobank to their adult (ie, >18 years) children and/or grandchildren (if applicable).

A total of 116 435 individuals consented to participate in the study. From this pool of consented individuals, UK Biobank selected 20 203 participants (11 345 UK Biobank participants and 8858 of their relatives; eight participants subsequently withdrew, leaving 20 195 participants) using the following procedures. First, households with more than one consented participant (UK Biobank participants or their relatives) were identified, and a single study participant from each household was randomly selected for inclusion in the sampling pool. Participants were then selected within each postcode area of mainland UK (excluding Northern Ireland) using random stratified sampling by age, sex and socioeconomic status (online supplemental table S1). The number of participants selected from each area was proportional to the geographical distribution of the 2011 UK Census population. There was some oversampling of participants from ethnic minority populations and from urban settings, including London.

Between mid-May and mid-November 2020, at approximately monthly intervals, participants received a capillary blood sampling kit to collect a finger-prick blood sample (about 0.5 mL), together with a symptom questionnaire. Participants were requested to return the sample and questionnaire by post to the UK Biobank laboratory on the same day of collection (from October, participants could also complete the questionnaire online). Plasma was recovered from capillary blood samples and stored at −80°C at UK Biobank facilities (Cheadle, North West England) then shipped weekly on dry ice to the Target Discovery Institute (University of Oxford) for analysis. Serum immunoglobulin G to the spike (S) protein of SARS-CoV-2 was measured using a well-validated commercial diagnostic ELISA kit.[13 14]

In January 2021, participants were asked to complete a further questionnaire that included detailed information on the major social factors that might have affected their risk of exposure to SARS-CoV-2 at work, at home and in their community. Information was collected on: employment (including work sector, whether they worked mostly at home or outside home and whether their work involved close contact with public); household circumstances (including number, age and work status of coresidents); community factors (urban-rural classification, UK region and area deprivation); and major selected activities outside the home, which are to some extent socially determined, and may have increased participants' contact with others in the community (including usual mode of transportation and frequency of shopping).

Some questions in the questionnaire referred to specific time periods (mid-March to end of June; start of July to mid-September; mid-September to end of October; and November) to capture changes in social factors over the course of 2020. The time periods were chosen to broadly correspond to different phases of the pandemic in the UK. The first period is consistent with the first wave of SARS-CoV-2 infections and the associated government-mandated social restrictions (ie, 'lockdown') in England, Wales and Scotland. Restrictions during this period included the closure of schools and non-essential retail, the requirement of stay at home except for essential purposes, and a directive to work from home where possible (starting in May, these restrictions were slowly relaxed).[15 16]

Statistical analysis

Analysis was restricted to participants who returned a valid blood sample in July (ie, the month following the end of the first

wave of the pandemic in the UK; online supplemental table S2) and those who had completed the questionnaire on social factors administered at the end of the study.

Logistic regression models were used to estimate odds ratios (ORs) and corresponding 95% CIs for the association between ethnicity and SARS-CoV-2 seropositivity. Group-specific variances were used to calculate 95% CIs to enable comparisons between any two groups, rather than solely with the arbitrarily selected reference group.[17] ORs are reported with adjustment for age and sex only, and then with further adjustment for a range of social factors that may lie on the pathway between ethnicity and SARS-CoV-2 infection, including: UK region of residence, urban-rural classification of residence, work location, work sector, whether work involves close proximity with the public or coworkers, number of coresidents aged ≤18 years, number of coresidents aged >18 years, number of coresidents working outside the home, use of public transportation, frequency of shopping, Townsend Deprivation Score (derived from postcode) and highest educational level (see online supplemental table S3 for the operational definitions of these variables and the specific categories used in the analyses). Missing values for each covariate were assigned to a separate category.

To assess the extent to which each of these factors might explain the association between ethnicity and SARS-CoV-2 infection, we calculated the likelihood ratio (LR) χ^2 statistics for ethnicity in the relevant model.[18] The change in LR χ^2 statistics on addition of the ethnicity term to logistic models was assessed in models with and without potential mediators of the association between ethnicity and SARS-CoV-2 seropositivity. The percentage reduction in LR χ^2 statistics for ethnicity with further adjustment for each social factor was calculated as a measure of the extent to which the factor explains the association between ethnicity and SARS-CoV-2 seropositivity.[18 19] All analyses were performed using R (V.4.1.1; R Development Core Team, Vienna, Austria).

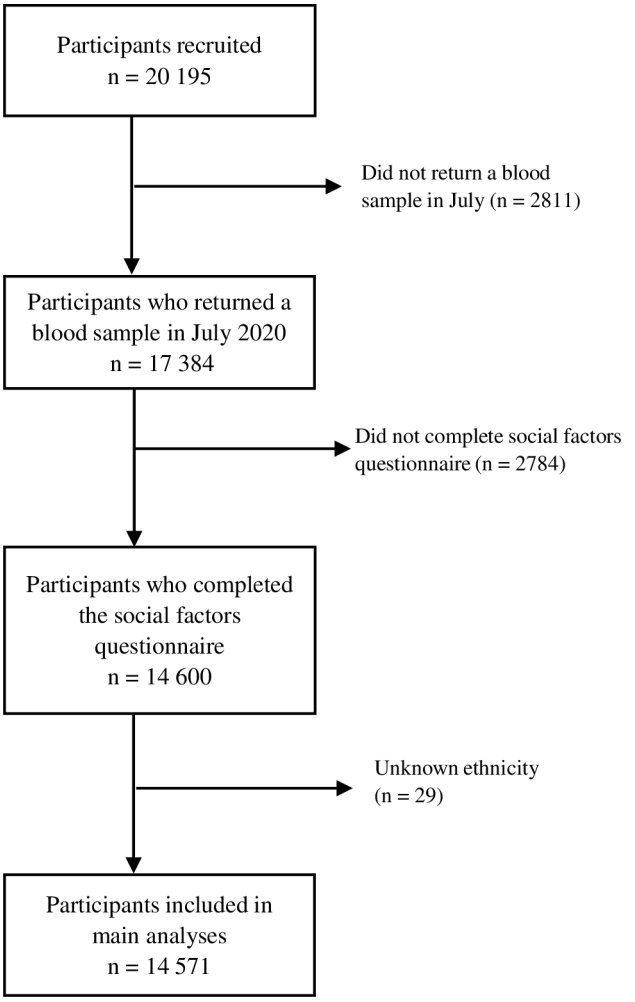

Figure 1 Flow diagram of participants included in the analysis.

Patient and public involvement

Participants were not involved in the design, conduct, reporting or dissemination of the research reported here. However, UK Biobank participants were involved in developing the ethics and governance framework for UK Biobank and have been engaged in the progress of UK Biobank through follow-up questionnaires and additional assessment visits.

RESULTS

Of the 20 195 participants of the Serology Study, 17 384 returned a blood sample in July 2020, of whom 14 600 also completed the questionnaire on social factors. We excluded a further 29 participants with missing information on ethnicity, leaving 14 571 participants available for analysis (figure 1). Overall, 87% of participants self-reported their ethnicity as White, 3% as Black, 3% as South Asian, 1% as Chinese, 4% as Mixed and 3% as Other ethnicity. The mean age of participants was 56 years and 58% were women. Participants were from all major regions of UK (online supplemental table S4), most lived in urban areas (86%) and just under two-thirds (63%) held a university or college degree. Compared with the 2021 UK Census, the study population was on average slightly older, with a higher proportion of female participants and a lower proportion from minority ethnic groups; there was also a higher proportion of participants living in London, reflecting the oversampling from this region (online supplemental table S4).

The characteristics of the study population varied by ethnicity (table 1; online supplemental table S5). Compared with those of White ethnicity, all other ethnic minority groups were more likely to live in urban areas (White 85%, other than White 89–96%), with a notably higher proportion living in London. Those of Black ethnicity were much more likely to live in more deprived areas than other ethnic groups (Townsend Deprivation Score ≥2: Black 40%, other than Black 19–25%), and had lower levels of education (university or college degree: Black 52%, other than Black 62–79%), whereas those of Chinese ethnicity were least likely to live in more deprived areas (19%) and had higher levels of education (79%) than other ethnic groups.

There were also differences in occupational and household factors. Black, South Asian, Chinese and Other ethnic groups were all less likely to work from home than those of White or Mixed ethnicity (White 32%, Mixed ethnicity 42%, other than White or Mixed ethnicity 23–29%); and those of Black ethnicity, in particular, were more likely to work in the NHS or social care (Black 17%, other than Black 6–11%), and more likely to work in close proximity with the public or coworkers (Black 39%, other than Black 20–30%), than all other ethnic groups.

With respect to household factors, compared with those of White, Mixed or Other ethnicity, participants of South Asian ethnicity were more likely to have one or more adult coresidents, whereas those of Black ethnicity and of Chinese ethnicity were less likely to have an adult coresident (White, Mixed and Other ethnicities 76–79%, South Asian 82%, Black 67%, Chinese

Table 1 Participant characteristics and their social circumstances during the first wave of the pandemic, by ethnicity

	White (n=12 725)	Black (n=369)	South Asian (n=396)	Chinese (n=126)	Mixed (n=526)	Other (n=429)	All (n=14 571)
Age (years), mean (SD)	55 (16.1)	60.9 (7.3)	61.8 (9.2)	62.5 (8.0)	54.0 (14.0)	63.5 (8.8)	55.6 (15.6)
Female, n (%)	7298 (57.4)	224 (60.7)	189 (47.7)	95 (75.4)	366 (69.6)	259 (60.4)	84 313 (57.9)
Resident of urban area, n (%)*	10 802 (84.9)	355 (96.2)	371 (93.7)	120 (95.2)	466 (88.6)	405 (94.4)	12 519 (85.9)
Resident of more deprived area, n (%)†	2890 (22.7)	149 (40.4)	83 (21.0)	24 (19.0)	133 (25.3)	136 (31.7)	3415 (23.4)
Educated to university or college level, n (%)	8020 (63.0)	193 (52.3)	246 (62.1)	74 (58.7)	330 (62.7)	268 (62.5)	9131 (62.7)
Works mostly from home, n (%)	4122 (32.4)	85 (23.0)	106 (26.8)	36 (28.6)	219 (41.6)	107 (24.9)	4675 (32.1)
Works in NHS or social care, n (%)	1038 (8.2)	61 (16.5)	39 (9.8)	7 (5.6)	57 (10.8)	33 (7.7)	1235 (8.5)
Works in close proximity with public/coworkers, n (%)	2841 (22.3)	142 (38.5)	117 (29.5)	25 (19.8)	140 (26.6)	107 (24.9)	3372 (23.1)
One or more coresidents aged ≤18 years, n (%)	3082 (24.2)	83 (22.5)	78 (19.7)	14 (11.1)	134 (25.5)	60 (14.2)	3452 (23.7)
One or more coresident aged >18 years, n (%)	10 000 (78.6)	248 (67.2)	325 (82.1)	88 (69.8)	405 (77.0)	329 (76.7)	11 395 (78.2)
Has a coresident who works outside the home, n (%)	3807 (29.9)	134 (36.3)	144 (36.4)	31 (24.6)	171 (32.5)	143 (33.3)	4430 (30.4)
Public transport is usual mode of transport, n (%)	377 (3.0)	38 (10.3)	9 (2.3)	5 (4.0)	25 (4.8)	20 (4.7)	474 (3.3)
Shopping frequency (≥1 day/week), n (%)	6491 (51.0)	160 (43.4)	138 (34.8)	49 (38.9)	272 (51.7)	171 (39.9)	7281 (50.0)

Analyses among 14 571 participants; exclusions as in figure 1.
*UK Biobank urban and rural area classification, based on postcode information matched to census information on population density.
†More deprived area defined as postcode with Townsend Deprivation Score ≥2. See online supplemental table S2 for further details on characteristics.
NHS, National Health Service.

70%). Participants of Black ethnicity were also more likely than other ethnic groups to use public transport as their usual mode of transport (Black 10%, other than Black 2–5%), and there was also evidence of differences in shopping frequency with those of White ethnicity and Mixed ethnicity shopping more frequently than other ethnic groups (shopping frequency greater than once per week: White 51%, Mixed 52%, other than White or Mixed ethnicity 35–43%).

Overall, 997 (7%) participants had antibodies to SARS-CoV-2 in blood samples taken in July 2020. In analyses adjusted for age and sex, ethnic minority groups (with the exception of Chinese and those of Mixed ethnicity) had a significantly higher odds of having antibodies to SARS-CoV-2 compared with those of White ethnicity (figure 2), with ORs (95% CIs) for Black, South Asian, Mixed and Other ethnic groups of 2.66 (1.98–3.59), 1.66 (1.18–2.35), 1.42 (1.05–1.91) and 1.79 (1.30–2.48), respectively. There were fewer participants of Chinese ethnicity than other ethnic groups, and hence substantial statistical uncertainty about the OR estimate for this group (0.99 (0.46–2.13)).

Table 2 shows the ORs for the association between ethnicity and SARS-CoV-2 seropositivity, and the percentage reduction in

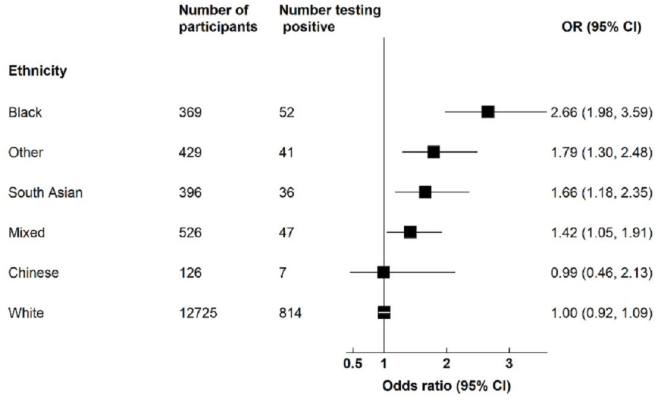

Figure 2 ORs for SARS-CoV-2 seropositivity by ethnicity, adjusted for age and sex. Analysis among 14 571 participants; exclusions as in figure 1. ORs, adjusted for age and sex, relative to participants of White ethnicity; group-specific variances used to calculate 95% CIs.

LR χ^2 statistics after adjustment for each social factor, or set of factors in a given category. Following full adjustment, the risk of SARS-CoV-2 seropositivity associated with ethnicity was substantially reduced compared with those of White ethnicity; the ORs (95% CIs) for Black, South Asian, Chinese, Mixed and Other ethnic groups were 1.83 (1.34–2.50), 1.20 (0.84–1.71), 0.89 (0.41–1.93), 1.31 (0.97–1.77) and 1.34 (0.96–1.87), respectively.

Adjustment for all the social factors reduced the LR χ^2 statistic for ethnicity by 67%, indicating that about two-thirds of the excess risk of SARS-CoV-2 seropositivity among ethnic minority groups was explained by these mediating factors. Adjustment for geographical factors (UK region, and whether the participant lived in an urban or rural area) resulted in the largest reduction in the LR χ^2 statistic (38%), followed by occupational factors (21%), other measures of socioeconomic status (education and Townsend Deprivation Score; 13%), use of public transport and shopping frequency (10%) and household factors (9%).

Of the geographical factors assessed, UK region alone reduced the LR χ^2 statistic by 36% and urban-rural classification of residence by 11%. For the individual occupational factors, working outside the home and working in close proximity with public or with coworkers reduced the LR χ^2 statistic by 13% and 15%, respectively, whereas working in the NHS or in social care reduced the LR χ^2 statistic by 7%.

In analyses that progressively adjusted for categories of factors (table 3), the effect of adjustment for geographical and occupational factors on the association of SARS-CoV-2 seropositivity with ethnicity was found to be largely independent of each other; the LR χ^2 statistic for ethnicity was reduced by 56% after adjustment for both these factors. By contrast, the LR χ^2 statistic declined by only a further 4% after additional adjustment for household characteristics, and a further 7% for all other measures combined.

Sensitivity analyses were conducted to include the results of blood samples returned in May, June or July. Cases were defined as those who tested positive in any month, and all other participants who returned a sample during this period were defined as negative for SARS-CoV-2. This increased the number of SARS-CoV-2 cases identified but meant including some participants

Table 2 ORs for SARS-CoV-2 seropositivity by ethnicity, with adjustment for age, sex and social factors

	OR (95% CI)						Per cent reduction in LR statistics for ethnicity
	White	Black	South Asian	Chinese	Mixed	Other	
Baseline model							
Ethnicity+age+sex	1.00 (0.92 to 1.09)	2.66 (1.98 to 3.59)	1.66 (1.18 to 2.35)	0.99 (0.46 to 2.13)	1.42 (1.05 to 1.91)	1.79 (1.30 to 2.48)	
Further adjustment							
Geographical factors							
Urban-rural classification	1.00 (0.92 to 1.09)	2.53 (1.88 to 3.41)	1.60 (1.13 to 2.25)	0.95 (0.44 to 2.04)	1.39 (1.03 to 1.88)	1.73 (1.25 to 2.39)	11
UK region	1.00 (0.92 to 1.09)	2.24 (1.66 to 3.02)	1.44 (1.02 to 2.03)	0.89 (0.41 to 1.92)	1.35 (1.00 to 1.82)	1.54 (1.11 to 2.14)	36
All	1.00 (0.92 to 1.09)	2.21 (1.64 to 2.99)	1.42 (1.01 to 2.01)	0.88 (0.41 to 1.90)	1.34 (0.99 to 1.81)	1.53 (1.11 to 2.12)	38
Occupational factors							
Workplace location*	1.00 (0.92 to 1.09)	2.48 (1.84 to 3.34)	1.59 (1.12 to 2.24)	1.00 (0.47 to 2.16)	1.40 (1.04 to 1.89)	1.76 (1.28 to 2.44)	13
Work in NHS or social care	1.00 (0.92 to 1.09)	2.54 (1.88 to 3.43)	1.64 (1.16 to 2.32)	1.02 (0.47 to 2.19)	1.42 (1.05 to 1.91)	1.80 (1.30 to 2.48)	7
Work in close proximity with public/coworkers	1.00 (0.92 to 1.09)	2.45 (1.82 to 3.30)	1.58 (1.12 to 2.23)	1.00 (0.46 to 2.15)	1.42 (1.05 to 1.91)	1.73 (1.25 to 2.40)	15
All	1.00 (0.92 to 1.09)	2.37 (1.76 to 3.21)	1.54 (1.09 to 2.18)	1.03 (0.48 to 2.21)	1.39 (1.03 to 1.88)	1.72 (1.25 to 2.39)	21
Household factors							
Number of coresidents aged ≤18 years	1.00 (0.92 to 1.09)	2.67 (1.98 to 3.59)	1.66 (1.18 to 2.35)	0.99 (0.46 to 2.12)	1.42 (1.05 to 1.92)	1.80 (1.30 to 2.48)	0
Number of coresidents aged >18 years	1.00 (0.92 to 1.09)	2.65 (1.97 to 3.57)	1.52 (1.07 to 2.15)	0.97 (0.45 to 2.09)	1.41 (1.05 to 1.91)	1.73 (1.25 to 2.39)	8
Has a coresident who works outside the home	1.00 (0.92 to 1.09)	2.61 (1.94 to 3.51)	1.61 (1.14 to 2.28)	1.01 (0.47 to 2.16)	1.42 (1.05 to 1.92)	1.75 (1.26 to 2.42)	6
All	1.00 (0.92 to 1.09)	2.63 (1.95 to 3.54)	1.53 (1.08 to 2.17)	0.99 (0.46 to 2.12)	1.42 (1.05 to 1.92)	1.71 (1.23 to 2.36)	9
Shopping frequency/use of public transport							
Frequency of shopping	1.00 (0.92 to 1.09)	2.62 (1.95 to 3.53)	1.63 (1.16 to 2.30)	0.95 (0.44 to 2.05)	1.41 (1.04 to 1.9)	1.77 (1.28 to 2.45)	5
Use of public transport	1.00 (0.92 to 1.09)	2.53 (1.88 to 3.41)	1.62 (1.15 to 2.28)	0.95 (0.44 to 2.05)	1.40 (1.04 to 1.89)	1.75 (1.27 to 2.42)	11
All	1.00 (0.92 to 1.09)	2.54 (1.88 to 3.42)	1.63 (1.15 to 2.30)	0.95 (0.44 to 2.03)	1.39 (1.03 to 1.88)	1.75 (1.27 to 2.42)	10
Deprivation and education							
Deprivation level	1.00 (0.92 to 1.09)	2.49 (1.85 to 3.36)	1.67 (1.18 to 2.35)	1.01 (0.47 to 2.16)	1.41 (1.04 to 1.90)	1.74 (1.26 to 2.40)	11
Highest level of education	1.00 (0.92 to 1.09)	2.62 (1.95 to 3.53)	1.65 (1.17 to 2.33)	0.99 (0.46 to 2.12)	1.42 (1.05 to 1.91)	1.79 (1.30 to 2.47)	3
All	1.00 (0.92 to 1.09)	2.46 (1.82 to 3.31)	1.66 (1.18 to 2.34)	1.01 (0.47 to 2.16)	1.41 (1.04 to 1.9)	1.73 (1.25 to 2.39)	13
All variables	1.00 (0.91 to 1.10)	1.83 (1.34 to 2.50)	1.2 (0.84 to 1.71)	0.89 (0.41 to 1.93)	1.31 (0.97 to 1.77)	1.35 (0.97 to 1.87)	67

Analysis among 14 571 participants; exclusions as in figure 1. ORs relative to participants of White ethnicity, adjusted for age and sex (baseline model), with further adjustment for social factors where indicated; group-specific variances used to calculate 95% CIs. See online supplemental table S3 for further details on characteristics. LR statistic for ethnicity in the baseline model=48.21.
*Workplace location (mid-March to end of June 2020): working from home, working outside home, other.
LR, likelihood ratio; NHS, National Health Service.

Table 3 ORs for SARS-CoV-2 seropositivity by ethnicity, progressively adjusted for social factors

	OR (95% CI)						Per cent reduction in LR statistics for ethnicity
	White	Black	South Asian	Chinese	Mixed	Other	
Baseline model (ie, ethnicity+age+sex)	1.00 (0.92 to 1.09)	2.66 (1.98 to 3.59)	1.66 (1.18 to 2.35)	0.99 (0.46 to 2.13)	1.42 (1.05 to 1.91)	1.79 (1.30 to 2.48)	
+geographical factors	1.00 (0.92 to 1.09)	2.21 (1.64 to 2.99)	1.42 (1.01 to 2.01)	0.88 (0.41 to 1.90)	1.34 (0.99 to 1.81)	1.53 (1.11 to 2.12)	38
+occupational factors	1.00 (0.91 to 1.09)	1.93 (1.43 to 2.62)	1.31 (0.92 to 1.85)	0.92 (0.43 to 1.99)	1.32 (0.98 to 1.79)	1.46 (1.05 to 2.03)	56
+household factors	1.00 (0.91 to 1.10)	1.92 (1.41 to 2.61)	1.22 (0.86 to 1.73)	0.92 (0.43 to 1.99)	1.33 (0.98 to 1.80)	1.39 (1.00 to 1.93)	60
+shopping frequency and use of public transport	1.00 (0.91 to 1.10)	1.88 (1.38 to 2.56)	1.2 (0.84 to 1.71)	0.89 (0.41 to 1.92)	1.31 (0.97 to 1.78)	1.36 (0.98 to 1.89)	64
+deprivation and education	1.00 (0.91 to 1.10)	1.83 (1.34 to 2.50)	1.2 (0.84 to 1.71)	0.89 (0.41 to 1.93)	1.31 (0.97 to 1.77)	1.35 (0.97 to 1.87)	67

Analysis among 14 571 participants; exclusions as in figure 1. ORs relative to participants of White ethnicity, adjusted for age and sex (baseline model), with progressive adjustment for social factors where indicated; group-specific variances used to calculate 95% CIs. See online supplemental table S3 for further details on characteristics. LR statistic for ethnicity in the baseline model=48.2.
LR, likelihood ratio.

with unknown SARS-CoV-2 serology status at the end of the first wave (ie, those who returned negative samples in May or June but did not provide a sample in July). The analyses produced similar results to the main findings (adjustment for all the social factors reduced the LR χ^2 statistic for ethnicity by 64%; online supplemental table S6).

DISCUSSION

This large population-based study examined the association of ethnicity with risk of SARS-CoV-2 infection during the first wave of pandemic in the UK, and used detailed assessments of social factors to explore the pathways between ethnicity and risk of SARS-CoV-2 infection. Some ethnic minority groups (including Black, South Asian, Mixed and Other ethnic groups) had a much higher risk of SARS-CoV-2 infection compared with those of White ethnicity. This excess risk was reduced by more than half after adjustment for occupational and geographical factors (mainly UK region of residence), with modest additional effects from household composition, transport, shopping or other measures of socioeconomic status; more precisely measured social factors (or other unmeasured social factors) may have further reduced the strength of the LR χ^2 statistic.[19] As such, the effect of any hypothesised biological differences in susceptibilities between ethnic groups (such as differences in expression of transmembrane serine protease 2 in lung tissue, which may facilitate entry and spread of SARS-CoV-2 in host cell[20]) is likely to be modest, at best.

This study suggests that ethnic disparities in SARS-CoV-2 infection risk during the first wave of the pandemic in the UK are largely a function of how ethnicity shapes social opportunities and inequities, and highlights the particular relevance of occupational and geographical factors. In the first wave of the pandemic in the UK, the highest infection rates were in major urban conurbations (particularly London, which had two to three times higher cumulative SARS-CoV-2 seropositivity by June 2020 than other UK regions), and ethnic minority groups are more likely than White groups to live in such regions of the UK.[21] With respect to occupational factors, working outside the home, and working in close proximity with public or with coworkers, changed the association with ethnicity more than working specifically in health or social care. Not all work in health and social care involves close proximity to others, and such environments are likely to have provided some personal protective equipment. The present study indicates the critical importance of infection control policies and equipment in all high-risk occupations in the context of a pandemic, not just in health and social care.

Of the various other social factors, use of public transport and Townsend Deprivation Score were the only other variables that reduced the LR χ^2 statistic for ethnicity by 10% or more. However, the effect of these variables was reduced when differences in UK region and occupation were adjusted for. Adjustment for household factors had only a modest impact on the association. Household transmission had previously been identified as a major cause of infection in the UK population as a whole,[1 22 23] but the present study suggests that differences in household factors (such as household size or age composition) between ethnicities are unlikely to drive much of the ethnic disparities in SARS-CoV-2 infection during the first wave. There is some evidence from other studies, however, that those in large, multigenerational household (with both school-age children and older adults) were at higher risk of infection during the second wave of the pandemic in the UK when schools remained open for several months, and it is thought this may explain the particularly high rates of COVID-19 deaths during the second wave among Bangladeshi and Pakistani communities in the UK.[3 24] There were too few new cases of SARS-CoV-2 among ethnic minority groups after the first wave in the present study for reliable analyses in subsequent time periods.

This study was designed to recruit individuals across a wide range of age, sex, ethnicity and socioeconomic deprivation, and while not fully representative of the UK general population, the prevalence of SARS-CoV-2 seropositivity across the different ethnic groups is consistent with contemporary UK national surveys.[1 7] The Real Time Assessment of Community Transmission (REACT-2) study, a representative survey of SARS-CoV-2 serological status among 100 000 adults in the UK, conducted at the same time as the current study, described similar excess risks among ethnic minority groups, with relative risks of infection among those of Black, South Asian and Mixed ethnicities of about 3, 2 and 1.5 times higher than among those of White ethnicity, respectively.[1] Other studies with less representative sampling, or that used routinely collected health data (for which unbiased assessments rely on equal likelihood of being tested for SARS-CoV-2 infection among subgroups of the population), have tended to find somewhat lower excess risks of SARS-CoV-2 infection among ethnic minority groups.[2 5]

A number of previous studies have investigated potential causes of ethnic disparities in SARS-CoV-2 infection in the UK. Several studies have found area-level measures of deprivation to explain some of the ethnic differences in prevalence of SARS-CoV-2 seropositivity, but the effects have largely been modest, and most of these studies have lacked the detailed information on the exposures needed to understand which aspects of deprivation might be contributing to these disparities.[2 25] Other studies have investigated the effect of more specific social factors (such as employment and household composition) on ethnic differences in SARS-CoV-2 infection, but the relative importance of these factors has previously not been well described.[1 8] Although the relevance of specific social factors is most generalisable to the target UK population, the study did not find any evidence to support the hypothesis that genetic predisposition or other biological susceptibilities are the main reasons for higher infection rates of SARS-CoV-2 among some ethnic groups, and this finding is likely generalisable to many other populations.

The detailed assessments of social factor are a key strength of the present study. This information allowed for exploration of the independent contribution of occupational, household and other social factors in explaining ethnic differences in the risk of SARS-CoV-2 infection, although we acknowledge that this information was self-reported at the end of the study. Another strength of this study was that vaccination status did not influence behaviour or the risk of infection because the study was conducted before the widespread vaccination of the UK population. In addition, as SARS-CoV-2 status was measured in all participants, the impact of collider bias was likely to be limited.[26]

The size of the study enabled robust analysis of the association of ethnicity with risk of SARS-CoV-2 infection. However, greater sample size would have allowed exploration of differences in risk within the broad ethnic groups used in the present study, such as between individuals who identify as Black Caribbean, Black African or any other Black background, or among certain smaller ethnicity groups. Ethnicity is a complex social construct, with groups reflecting diverse aspects of culture, ancestry, religion, language, nationality and other characteristics.[2 27 28] There may well be important differences in risk of SARS-CoV-2 infection within the ethnic categories defined in the present study,

and understanding these differences would allow more targeted public health action.

It is also a limitation that it was not possible to assess the consistency of the social structural experiences (such as crowded housing) by ethnicity between the study population and the wider UK population, which would also have been helpful in assessing generalisability of the findings. Greater sample size would also have allowed investigation of the associations by level of other factors (such as within region) to assess further the generalisability. In addition, although efforts were made to recruit participants with similar demographic characteristics to the wider UK population, it is important to acknowledge that selected individuals (who were all volunteers) may differ in other ways that it was not possible to control for, which might also affect risk of infection (such as trust in authorities), with the potential to cause bias.

The analyses used changes in the LR χ^2 statistic to assess the impact of social factors as potential variables in the pathway between ethnicity and SARS-CoV-2 infection risk. Although this and other adjustment methods have been widely used to assess the impact of potential mediators, various assumptions are needed for the effects to be interpreted causally, notably that regression models control for exposure-outcome, mediator-outcome and exposure-mediator confounding. Alternative methods, such as causal mediation analysis, might be useful in future analyses to assess further the causal impact of selected variables on the association of specific ethnicities with SARS-CoV-2 infection risk, although larger data sets might be required. Future research should also explore whether the findings varied for different waves of the pandemic, as the data set did not have sufficient number of cases among those of non-White ethnicity after the first wave. Lastly, we were limited in our ability to assess the impact of more specific occupations (given the number of participants in each occupational group) on differences in risk of SARS-CoV-2 infection between ethnic groups, and this should also be an important focus of future research.

In conclusion, this large-scale population-based study suggests that higher infection rates of SARS-CoV-2 among ethnic minority groups are largely a function of how ethnicity shapes social opportunity and inequities. The study identifies a wide range of social factors driving these inequalities, but highlights the particular relevance of occupational and geographical factors. In addition to addressing the inequalities in social opportunity that ultimately underlie the differences in infection rates between ethnic groups, this study indicates that policies to mitigate the specific vulnerabilities of different regions (including major urban conurbations, given the higher proportion of the population from ethnic minority groups in such regions of the UK), as well as policies which address occupational exposure to major respiratory infections, are likely to be particularly important in ensuring future pandemics do not further exacerbate existing ethnic inequalities in health.

Acknowledgements We thank the study participants. We also acknowledge the work of UK Biobank staff, including Maggie Gregory, Jon Scattergood, Maureen Dawson, Lynn Pank, Christine Ryan and Rob Nicholls.

Contributors NA initiated the study. WO had full access to all of the data in the study and takes responsibility for the accuracy of the data analysis. WO, JH, BL and NA drafted the manuscript and did the statistical analysis. All authors contributed to the design of the study; the acquisition, analysis and interpretation of data; and the critical revision of the manuscript for important intellectual content. WO, JH, BL and NA are the guarantors of this study.

Funding The funding for this substudy of UK Biobank was provided by the UK Department of Health and Social Care. The core funding for UK Biobank is provided by the UK Medical Research Council, Wellcome, British Heart Foundation, Cancer Research UK and the National Institute for Health Research (grant ref: 223600/Z/21/Z).

Disclaimer The funders had no role in the design; in the collection, analysis or interpretation; in the writing of the report; or in the decision to submit the paper for publication.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the North West (Haydock) Research Ethics Committee (ref: 16/NW/0274). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. UK Biobank is available for open access, without the need for collaboration, to any bona fide researcher who wishes to use it to conduct health-related research for the benefit of the public.

ORCID iDs
Natasha Wood http://orcid.org/0000-0002-3791-2957
Jelena Bešević http://orcid.org/0000-0003-4927-1840
Megan Conroy http://orcid.org/0000-0002-3847-6202
Ben Lacey http://orcid.org/0000-0003-0139-2934

REFERENCES

1 Ward H, Atchison C, Whitaker M, *et al*. SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nat Commun* 2021;12:905.
2 Mathur R, Rentsch CT, Morton CE, *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.
3 Nafilyan V, Islam N, Mathur R, *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;36:605–17.
4 Lo C-H, Nguyen LH, Drew DA, *et al*. Race, ethnicity, community-level socioeconomic factors, and risk of COVID-19 in the United States and the United Kingdom. *EClinicalMedicine* 2021;38:101029.
5 Sze S, Pan D, Nevill CR, *et al*. Ethnicity and clinical outcomes in COVID-19: a systematic review and meta-analysis. *EClinicalMedicine* 2020;29:100630.
6 Batty GD, Gaye B, Gale CR, *et al*. Explaining ethnic differentials in COVID-19 mortality: cohort study. *Am J Epidemiol* 2022;191:275–81.
7 Pouwels KB, House T, Pritchard E, *et al*. Community prevalence of SARS-Cov-2 in England from April to November, 2020: results from the ONS Coronavirus infection survey. *Lancet Public Health* 2021;6:e30–8.
8 Raisi-Estabragh Z, McCracken C, Bethell MS, *et al*. Greater risk of severe COVID-19 in black, Asian and minority ethnic populations is not explained by cardiometabolic, socioeconomic or behavioural factors, or by 25(OH)-Vitamin D status: study of 1326 cases from the UK biobank. *J Public Health (Oxf)* 2020;42:451–60.
9 Mude W, Oguoma VM, Nyanhanda T, *et al*. Racial disparities in COVID-19 pandemic cases, hospitalisations, and deaths: a systematic review and meta-analysis. *J Glob Health* 2021;11:05015.
10 Pan D, Sze S, Martin CA, *et al*. Covid-19 and ethnicity: we must seek to understand the drivers of higher transmission. *BMJ* 2021;375:2709.
11 Sudlow C, Gallacher J, Allen N, *et al*. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779.
12 Fry A, Littlejohns TJ, Sudlow C, *et al*. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol* 2017;186:1026–34.
13 Adams ER, Ainsworth M, Anand R, *et al*. Antibody testing for COVID-19: a report from the National COVID scientific advisory panel. *Wellcome Open Res* 2020;5:139.

14 The National SARS-CoV-2 Serology Assay Evaluation Group. Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. *Lancet Infect Dis* 2020;20:1390–400.
15 Knock ES, Whittles LK, Lees JA, *et al*. Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Sci Transl Med* 2021;13:eabg4262.
16 Public Health England. *NDNS: diet and physical activity – a follow-up study during COVID-19, P.H. England, Editor*. 2021.
17 Plummer M. Improved estimates of floating absolute risk. *Stat Med* 2004;23:93–104.
18 Parish S, Peto R, Palmer A, *et al*. The joint effects of apolipoprotein B, apolipoprotein A1, LDL cholesterol, and HDL cholesterol on risk: 3510 cases of acute myocardial infarction and 9805 controls. *Eur Heart J* 2009;30:2137–46.
19 Floud S, Balkwill A, Moser K, *et al*. The role of health-related behavioural factors in accounting for inequalities in coronary heart disease risk by education and area deprivation: prospective study of 1.2 million UK women. *BMC Med* 2016;14:145.
20 Bunyavanich S, Grant C, Vicencio A. Racial/ethnic variation in nasal gene expression of transmembrane serine protease 2 (TMPRSS2). *JAMA* 2020;324:1567–8.
21 Murray CJL, Abbafati C, Abbas KM, *et al*. Five insights from the global burden of disease study 2019. *Lancet* 2020;396:1135–59.
22 Miller E, Waight PA, Andrews NJ, *et al*. Transmission of SARS-CoV-2 in the household setting: a prospective cohort study in children and adults in England. *J Infect* 2021;83:483–9.
23 Koh WC, Naing L, Chaw L, *et al*. What do we know about SARS-Cov-2 transmission? A systematic review and meta-analysis of the secondary attack rate and associated risk factors. *PLoS One* 2020;15:e0240205.
24 Nafilyan V, Islam N, Ayoubkhani D, *et al*. Ethnicity, household composition and COVID-19 mortality: a national linked data study. *J R Soc Med* 2021;114:182–211.
25 Niedzwiedz CL, O'Donnell CA, Jani BD, *et al*. Ethnic and socioeconomic differences in SARS-Cov-2 infection: prospective cohort study using UK biobank. *BMC Med* 2020;18:160.
26 Griffith GJ, Morris TT, Tudball MJ, *et al*. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun* 2020;11:5749.
27 Flanagin A, Frey T, Christiansen SL, *et al*. The reporting of race and ethnicity in medical and science journals: comments invited. *JAMA* 2021;325:1049–52.
28 Lee C. "Race" and "Ethnicity" in biomedical research: how do scientists construct and explain differences in health" *Soc Sci Med* 2009;68:1183–90.