Ethnic differences in COVID-19 infection, hospitalisation, and mortality: an OpenSAFELY analysis of 17 million adults in England

The OpenSAFELY Collaborative: Rohini Mathur*+, PhD[1], Christopher T Rentsch*, PhD[1], Caroline E Morton*, MRCGP[2], William J Hulme, PhD[2], Anna Schultze, PhD[1], Brian MacKenna, MPharm[2], Rosalind M Eggo[1], PhD, Professor Krishnan Bhaskaran, PhD[1], Angel YS Wong, PhD[1], Professor Elizabeth J Williamson, PhD[1], Harriet Forbes, PhD[1], Kevin Wing, PhD[2], Helen I McDonald, PhD[14], Chris Bates, PhD[3], Seb Bacon, BA[2], Alex J Walker, PhD[2], David Evans, MPhil[2], Peter Inglesby, MPhil[2], Amir Mehrkar, MB[2], Helen J Curtis, DPhil[2], Nicholas J DeVito, MPH[2], Richard Croker, MSc[2], Henry Drysdale, MRCS[2], Jonathan Cockburn, BSc[3], John Parry, MRCGP[3], Frank Hester, BSc[3], Sam Harper, MSci[3], Professor Ian J Douglas, PhD[1], Laurie Tomlinson, PhD[1], Professor Stephen JW Evans, MSc[1], Professor Richard Grieve[1], David Harrison[5], Kathy Rowan[5], Professor Kamlesh Khunti, FMedSci[6], Professor Nishi Chaturvedi[7], Professor Liam Smeeth, FMedSci[1]†, Ben Goldacre, MRCPsych[2]†

*Authors contributed equally
1 London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT

2 The DataLab, Nuffield Department of Primary Care Health Sciences, University of Oxford, OX26GG

3 TPP, TPP House, 129 Low Lane, Horsforth, Leeds, LS18 5PX

4 NIHR Health Protection Research Unit (HPRU) in Immunisation, WC1E 7HT

5 Intensive Care National Audit and Research Centre (ICNARC), 24 High Holborn, London WC1V 6AZ

6 Diabetes Research Centre, University of Leicester, Leicester, LE5 4PW

7 MRC Unit for Lifelong Health and Ageing, University College London, 1 – 19 Torrington Place, London, WC1E 7HB

+Corresponding

Abstract: 300/300
Word count: 3,814/3,500
Tables/figures: 5
References: 58

Abstract

Background: COVID-19 has disproportionately impacted UK ethnic minority populations. Our aim was to quantify ethnic differences in SARS-CoV-2 infection and COVID-19 outcomes during the first and second waves of the coronavirus pandemic in England.

Methods: An observational cohort study using the OpenSAFELY platform. Multivariable Cox regression adjusted for socio-demographic, clinical and household factors examined ethnic differences in being tested and testing positive for SARS-CoV-2 and COVID-19-related hospitalisations, intensive care unit (ICU) admissions, and mortality between February and August 2020 (wave 1) and September and December 2020 (wave 2).

Findings: Of 17,288,532 adults, 63% were White, 5.9% south Asian, 2% Black, 1.8% other, 1% mixed, and 26.3% unknown. In wave 1, likelihood of testing did not differ markedly by ethnicity; however, risk of testing positive was doubled in south Asian groups (HR 1.99, 95%CI 1.94-2.04) and 1.69 times higher in Black groups (1.62-1.77). Compared to White groups, south Asian groups were at elevated risk of COVID-19-related hospitalisation (1.48, 1.41-1.55), ICU admission (2.18, 1.92-1.48), and mortality (1.26, 1.15-1.37). Similarly, Black groups were also at elevated risk of COVID-19-related hospitalisation (1.78, 1.67-1.90), ICU admission (HR 3.12, 2.65-1.90), and COVID-19 mortality (1.51, 1.33-1.71) compared to White groups. In wave 2, relative risks of hospitalisation, ICU admission, and death increased for south Asian groups and attenuated for Black groups relative to White. Disaggregation into 16 group ethnicity revealed important heterogeneity.

Interpretation: Ethnic minority populations in England have excess risks of testing positive for SARS-CoV-2 and COVID-19 outcomes even after accounting for differences in socio-demographic, clinical, and household characteristics. Causes are likely to be multifactorial. Delineating exact mechanisms is crucial. Tackling ethnic inequalities will require action across many fronts including reducing structural inequalities, addressing barriers to equitable care, and improving uptake of testing and vaccination.

Funding: Medical Research Council (MR/V015737/1)

Keywords: Coronavirus; COVID-19; SARS-CoV-2; ethnicity; UK; England; inequalities; primary care; ICU; mortality

Research in context

Evidence before this study

We searched PubMed for population-based studies examining the association between ethnicity and COVID-19; keywords included (ethnic* OR race) AND (COVID OR coronavirus OR SARS-CoV-2) AND (UK or England) AND (risk OR rate OR odds)". Results were filtered to those conducted on humans, published from 2019 onwards with abstracts available. We identified six studies examining ethnic differences in the COVID-19 infection and outcomes in population-based samples. Five studies from the UK Biobank reported increased risk of COVID-19 infection and hospitalization in Black and south Asian groups. As the UK Biobank cohort is known to be healthier, less deprived and less ethnically diverse than the UK population, findings are not wholly generalizable to the wider UK population. Our previous study in the OpenSAFELY platform, reported increased risk of COVID-19 mortality in ethnic minority groups, but did not examine the role of household size or examine ethnic differences in COVID-19 infection and outcomes earlier in the care pathway.

Added value of this study

This is the largest study in the UK to examine ethnic inequalities in testing positive for SARS-CoV-2 and COVID-19 related outcomes in a cohort covering 40% of the population in England. Additionally, it is the only population representative study to date which accounts for household size in addition to socio-demographic characteristics and clinical co-morbidities. Examining ethnicity according to both high-level and detailed ethnic groupings, we highlight important ethnic differences in risks of testing positive for SARS-CoV-2, COVID-19 related hospital and ICU admissions and COVID-19 related mortality. We show that multiple factors contribute to ethnic inequalities in COVID-19 and the importance of these factors varies by ethnic group. Compared to wave 1, wave 2 risks of COVID-19 hospitalisation and death were magnified for south Asian groups and reduced in all other ethnic minority groups.

Implications of all the available evidence

The risks of COVID-19 infection and severe outcomes are disproportionately increased amongst ethnic minority groups, both in the UK and internationally. Reducing ethnic inequalities in COVID-19 risks requires action on social determinants including addressing disadvantage and discrimination, reducing risk of infection and transmission, improving quality of and access to quality clinical care and improving management of pre-existing clinical conditions. The appropriate balance of these actions needs tailoring for different ethnic groups.

Background

The risks of SARS-CoV-2 infection and COVID-19 disease have been reported to be disproportionately increased amongst ethnic minority groups, both in the UK and internationally.[1–6] It is hypothesized that ethnic differences are driven by factors such as living in deprived areas, working in high-exposure or frontline occupations, living in large, multigenerational households, higher burden of underlying conditions, experiences of discrimination, and access to health or community services.[7–10]

In the UK, collection of ethnicity data is considered essential for identifying and reducing ethnic inequalities.[11,12] Though there is no single universally accepted definition of ethnicity, it serves as an important social construct and surrogate marker for shared exposures or risks for people with similar social, biological, and cultural characteristics.[13–15]

To date, many COVID-19 studies have reported findings according to high-level ethnic groupings, such as, white, south Asian, and black, rather than considering disaggregated ethnic groupings. Furthermore, most evidence has been derived from populations with severe disease requiring hospitalisation, making it difficult to extrapolate findings to the general population.[16–21] Finally, while previous studies have accounted for health status, socio-economic deprivation, or household composition none yet have considered these factors in conjunction.[22,23]

The aim of this study was to estimate the effect of ethnicity on being tested and testing positive for SARS-CoV-2, and COVID-19 related hospitalisation, ICU admission and mortality, recognizing the potential role of socio-demographic, clinical, and household related factors.

Methods

Study design and population

We conducted a population-based, observational cohort study using the OpenSAFELY platform, for which NHS England is the data controller.[24] OpenSAFELY holds electronic health records data for 24 million people registered with primary care practices using TPP software, representing approximately 40% of the English population.[29]

Individuals-level primary care data were linked to SARS-CoV-2 testing data from the Second Generation Surveillance System (SGSS), COVID-19 related hospital admissions from the secondary uses service (SUS), COVID-19 related ICU admissions from the Intensive Care National Audit & Research Centre (ICNARC)[25], and mortality data from the Office for National Statistics (ONS). The study population comprised adults aged 18 older, registered with a primary care practice on 1 February 2020. The study period ranged from 1 February

2020 to 3 August 2020 for wave 1 and from 1 September 2020 to 31 December 2020 for wave 2. A minimum of twelve months of continuous registration prior to the start date of each wave was required for inclusion, to ensure that baseline factors were adequately captured. Individuals residing in care homes were excluded from the main analyses as we hypothesized that the role of socio-demographic, clinical, and household characteristics would be systematically different for care home residents than for the general population.

Study variables

The primary exposure was self-reported ethnicity as captured on the primary care record, collapsed into the five high-level and 16 detailed census categories of white (White British, Irish, other white), south Asian (Indian, Pakistani, Bangladeshi, other south Asian), black (African, Caribbean, other black), other (Chinese, all other), and mixed (white and Asian, white and African, white and Caribbean, other mixed), and unknown. Comparisons were reported for the five high-level ethnic groups with the white group as reference, and for the 16 disaggregated groups, with the white British group as the reference.

Outcomes of interest included receiving a PCR test or testing positive for SARS-CoV-2 and COVID-19 related hospital admission, ICU admissions, and mortality, the latter defined as the presence of ICD-10 codes U07.1 (confirmed COVID-19) and U07.2 (suspected COVID-19) anywhere on the death certificate. Testing outcomes were obtained from the UK's Pillar 1 (NHS and Public Health England laboratories) and Pillar 2 (commercial partners) testing strategies and included results from PCR swab tests used to identify symptomatic individuals.[26,27]

Demographic characteristics included age, sex, deprivation, household size (number of people living in a household, categorised as 1-2 people; 3-5 people; 6-10 people; 11 or more people), number of primary care consultations in the 12 months prior, and geographic region, defined by the sustainability and transformation partnership (STP, a National Health Service administrative area). Deprivation was defined using quintiles of the Index of Multiple Deprivation (IMD), an area level composite measure of seven domains including income; employment, education, skills and training, health and disability; crime; barriers to housing services and living environment.[28] Household size was determined using the number of individuals (of all ages) in OpenSAFELY residing at the same address on 1 February 2020.

Clinical covariates were identified using the Read clinical classification system[29] and included body mass index (BMI), glycated haemoglobin (HbA1c), and blood pressure (BP). BMI in kg/m^2 was grouped into six categories using the World Health Organisation classification with adjustments for south Asian ethnicity: underweight (<18 kg/m^2), normal 18.5– 24 (23.5 if south Asian), overweight 25-30 kg/m^2 (23.6-27.5); obese I 30-34.9 (27.5-32.4); obese II 35-39.9 (32.5-37.4); obese III 40+ (37.5+). HbA1c was grouped into five categories: <6.5%, 6.5-

7.4%, 7.5-7.9%, 8-8.9%, >=9%. BP was grouped into four categories of normal (<120/80), elevated (120-130/80), high stage I (131-140/80-90), and high stage II (>140/90). Smoking status was grouped into current, former and never smokers. Those with missing smoking status were grouped as never smokers. Those with missing BMI, HbA1c and BP were grouped into a separate category of 'unknown'.

Clinical comorbidities were considered present at baseline if recorded any time prior to 1 February 2020 for wave 1 or 1 September 2020 for wave 2. These included: hypertension, asthma, chronic respiratory disease, chronic heart disease, type 1 and type 2 diabetes mellitus, cancer, chronic liver disease, stroke, dementia, other chronic neurological diseases, chronic kidney disease (CKD, defined as eGFR<60 ml/min/1.73m^2), end stage renal failure, common autoimmune diseases (rheumatoid arthritis, systemic lupus erythematosus, or psoriasis), and immunosuppression (HIV, sickle cell disease, organ transplant, asplenia). All codelists are available for review and re-use.[30]

Statistical Analysis

Socio-demographic and clinical characteristics at baseline were summarised using descriptive statistics, stratified by ethnic group. Follow-up began on 1 February 2020 for wave 1 and 1 September 2020 for wave 2 and ended at the earliest of experiencing the outcome of interest, death, de-registration from a primary care practice, or the censoring date for the dataset capturing the outcome of interest (between July 30 and August 3, 2020 for wave 1 and December 31, 2020 for wave 2).

Multivariable Cox proportional hazards regression was used to estimate ethnic differences in the cause-specific hazard of each outcome in the whole denominator population.[31] All analyses were adjusted for age (using restricted cubic splines), sex, deprivation quintile, diagnosed co-morbidities, BMI, HbA1c, blood pressure, number of primary care consultations in the preceding 12 months, household size. To investigate the extent to which age-sex adjusted ethnic differences could further be explained by deprivation, co-morbidities, and household size, we sequentially adjusted for age and sex in the first model, adding deprivation in the second, co-morbidities, clinical factors and GP consultations in the third, and household size in the fourth. All models were stratified by STP to account for clustering by geographical region. All analyses we conducted separately for wave 1 and wave 2.

Secondary and Sensitivity Analyses

First, we estimated ethnic differences in the risk of non-COVID-19 death (defined as any death without a COVID-19 diagnosis code anywhere on the death certificate). Second, we used logistic regression adjusting for all covariates to examine ethnic differences in the odds of testing positive amongst those tested for SARS-CoV-2. Third, we estimated ethnic

differences in all outcomes for care home residents, adjusting for all covariates except for household size. Sensitivity analyses included using multiple imputation to account for missing ethnicity data, examining ethnic differences in the risk of death where COVID-19 was the underlying cause (rather than any cause) and exploring the impact of regional variation on ethnic differences in all outcomes. Proportional hazards assumptions were assessed by testing for a zero slope in the scaled Schoenfeld residuals and graphical inspection of Kaplan-Meier plots.

Data management was performed using Python 3.8 and SQL, and analysis was carried out using Stata 16.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. CTR and CEM had full access to all of the data and the corresponding author had final responsibility for the decision to submit for publication.

Results

From a total of 23,600,617 individuals in OpenSAFELY on 1 February 2020, 17,288,532 adults aged 18 or over were included in the study (**Figure 1**). The ethnic breakdown of the study population was 63% White, 5.9% south Asian, 2% Black, 1.8% other, 1% mixed, and 26.3% unknown. Compared with the White population, ethnic minority groups were, on average, ten years younger and over-represented in deprived neighbourhoods, large households, and diabetic populations (**Table 1, S1**).

Ethnic differences in being tested and testing positive for SARS-CoV-2

Between 1 Feb 2020 and 3 Aug 2020, 7% of the study population received a test for SARS-CoV-2 infection (n=1,216,801), and 0.4% tested positive (n=71,246) (**Table 2**). The ethnic breakdown of individuals receiving a test was similar to that of the general population, though test recipients were slightly older with more co-morbid chronic conditions than the general population (**Table S2**). After accounting for all measured explanatory variables, south Asian, Black, and mixed groups were more likely to be tested and test positive (south Asian HR 1.99, 95% CI 1.94-2.04; Black HR 1.69, 95% CI 1.62-1.77; mixed HR 1.49, 95% CI 1.39-1.59; **Figure 2)**. Patterns across the 16 categories of ethnicity were similar, except for the Chinese group, for whom risks of being tested and testing positive were lower than for White groups. When restricted to the population ever receiving a test, ethnic patterning remained unchanged except for the Chinese group, who had equivalent risk of testing positive (OR 1.13, 95%CI 0.95-1.34; **Figure S1).**

Ethnic differences in COVID-19 related hospitalisation, ICU admissions and mortality

Between 1 Feb 2020 and 3 Aug 2020, 0.2% of the study population were admitted to hospital for COVID-19 (n=32,473), <0.1% were admitted to ICU for COVID-19 (n=3,096), and 0.1% had a COVID-19-related death (n=11,649) (**Table 2**). After accounting for all measured explanatory factors, risk of hospitalisation was increased in all ethnic minority groups relative to White (south Asian HR 1.48, 95% CI 1.41-1.55; Black HR 1.78, 95% CI 1.67-1.90; mixed HR 1.63, 95% CI 1.45-1.83; other HR 1.54, 95%CI 1.41-1.69; **Figure 3**). Risk ICU admission was increased 2 to 5 fold in ethnic minority groups relative to White (south Asian HR 2.18, 95% CI 1.92-2.48; Black HR 3.12, 95% CI 2.65-3.67; mixed HR 2.96, 95% CI 2.26-3.87; other HR 3.18, 95%CI 2.58-3.93; **Figure 3**). Risk of COVID-19 death was increased by 22-51% in ethnic minority groups relative to white (south Asian HR 1.26, 95% CI 1.15-1.37; Black HR 1.51, 95% CI 1.31-1.71; mixed HR 1.41, 95% CI 1.11-1.81; other HR, 1.22, 95% CI 1.00-1.48) (**Figure 4**).

Role of deprivation, clinical characteristics and household size

After accounting for age and sex, further adjustment had little impact on likelihood of being tested for COVID-19. In south Asian groups, adjustment for clinical characteristics led to the largest reduction in the hazard ratios in testing positive for SARS-CoV-2, hospitalisation and ICU admission, while adjustment for deprivation and household size made equivalent reductions in the hazard ratio for COVID-19 mortality. In all other ethnic minority groups, adjustment for social deprivation led to the largest reduction in the hazard ratio for all outcomes after accounting for age and sex (Table 2, S5).

Ethnic differences in wave 2 vs. wave 1

Between September 1 and December 31, 2020, 15% of the study population received a test (n=2,647,756), 2.9% tested positive (n=506,773), 0.1% were admitted to hospital for COVID-19 (n=18,885), and <0.1% had a COVID-19 related ICU admission (n=3,351) or COVID-19 related death (n=7,366). In contrast to the wave 1, all ethnic minority groups were less likely to be tested than White groups (**Figure 2).** South Asian groups remained at higher risk of testing positive (HR 1.32, 95%CI 1.31-1.33), with relative risks of COVID-19 related hospitalization, ICU admission and mortality greater in magnitude in wave 2 compared to wave 1 (hospitalization HR 1.89, 95%CI 1.79-2.00, ICU HR 2.68, 95%CI 2.39-3.01, mortality HR 1.87, 95%CI 1.68-2.07; **Figure 2, 3, 4)**. In contrast to wave 1, Black groups were less likely than White groups to test positive (HR 0.85, 95%CI 0.84-0.87), though risk of testing positive remained elevated amongst those ever tested (HR 1.03, 95%CI 1.02-1.06; **Figure 2, S1**). Risks of hospitalization and ICU admission remained higher for Black groups compared to White in wave 2, though attenuated in magnitude compared to wave 1 (hospitalisation HR 1.23, 95%CI 1.11-1.37; ICU HR 1.67, 95%CI 1.37-2.05; **Figure 3).** Risk of COVID-19 death was attenuated for Black groups compared to white (HR 0.92, 95%CI 0.73-1.16; **Figure 4**).

Secondary and sensitivity analyses

A total of 71,920 non-COVID related deaths occurred over wave 1. The risk of non-COVID-related death was 15-32% lower in all non-White ethnic groups compared with White groups (south Asian HR 0.85, 95%CI 0.81-0.90; Black HR 0.85, 95%CI 0.78-0.92; mixed HR 0.81, 95%CI 0.70-0.93; other HR 0.68, 95%CI 0.61-0.77; **Table S3**). In wave 2, risk of non-COVID death remained lower for south Asian, Black, and other groups compared to White groups (**Table S5**).

In wave 1, amongst the 78,124 care home residents, 59% individuals were tested for SARS-CoV-2, 8% tested positive, 3% were admitted to hospital and 5% died from COVID-19. While ethnic differences in being tested for or testing positive for COVID-19 were apparent, people of Black and other ethnicity were more likely to die from COVID-19 than people of White ethnicity (Black HR 1.43, 95%CI 1.02-2.00; other HR 1.73, 95%CI 1.19-2.50). In wave 2, no ethnic differences among care home populations were evident (**Figure S2).** Due to small numbers, we were unable to explore ethnic differences in ICU admissions or differences according to ethnicity in 16 categories among care home residents.

Using multiple imputation to account for unknown ethnicity did not materially change any of the associations observed in the complete case analysis (**Figure S6**), nor did restricting the definition of COVID-19 death to underlying cause only (**Figure S7**) or removing adjustment for STP region (**Figure S8**). We detected no evidence of deviations from the proportional-hazards assumption (**Table S7)**.

Discussion

Summary

In a population-based cohort study of 17 million adults in England we found that, while ethnic differences in testing were small, ethnic minority groups were at increased risk of testing positive for SARS-CoV-2 and COVID-19 related hospitalisation, ICU admission, and death. Disaggregation into detailed ethnic categories revealed important within-group heterogeneity, emphasizing the importance of disaggregated reporting wherever possible. In wave 2, ethnic minority groups were less likely to be tested than White groups, and risks of severe COVID-19 outcomes increased for south Asian groups whilst attenuated in all other ethnic groups compared to wave 1.

Strengths and limitations

In the largest UK-based study to date, we captured high quality clinical data across a range of healthcare settings and linked individual-level COVID-19 datasets which enabled us to generate timely insights into ethnic disparities at different stages of COVID-19 severity prior to mortality. We were able to report findings according to self-reported ethnicity in 16 categories whereas other UK-based studies have aggregated ethnicity into higher-level groups due to small numbers. Finally, we reported differences in outcomes using a general

population-based sample, which allowed us to overcome issues commonly faced by studies limited to individuals with SARS-CoV-19 infection, or hospitalization, whereby populations under study may not represent the true general population at risk.[32]

Our inability to capture all potential explanatory factors of ethnic disparities in COVID-19 outcomes is likely to have impacted our observed associations. For example, we were unable to account for ethnic differences in ancestry[33,34], occupation[35], experiences of racism or structural discrimination[9,36,37], and health-related behaviour[38,39]. Due to invalid address information, we were unable to estimate household size for 13% of our population. We may have underestimated household size for homes including people registered at non-TPP primary care practices and over-estimated it for individuals living in large apartment blocks, or for people who have not updated their address after moving. In recognition of these limitations, we grouped household size into four levels rather than considering it as a continuous measure. Furthermore, it is possible that cause of death may have been misclassified on death certificates, and that the extent of this misclassification may have differed by time period and ethnicity. A limitation of SARS-CoV-2 test data included the selective opportunity to be tested, which was skewed towards healthcare workers and people with severe or symptomatic disease, particularly during the first wave of the pandemic. Whilst OpenSAFELY is broadly representative of the English population, it includes data from a single software system which is known to have lower coverage in London compared to other regions of the UK. However, our results mirror other studies conducted in the UK[1] and in the US[5,40], suggesting that potential mechanisms underpinning ethnic differences in COVID-19 may be common across countries with similar population structures. OpenSAFELY data are collected prospectively in real time by clinicians and practice staff and are subject to the same strengths and biases as other UK- based EHR databases.

Despite these limitations, this study represents the most comprehensive examination of ethnic inequalities in England during the coronavirus pandemic in 2020. Using the OpenSAFELY data analytics platform, we capitalised on the rapid real-time linkage of routine datasets in a highly secure environment to explore a range of urgent questions around patterning of ethnic inequalities in the UK.

Findings in Context

In this study we build on previous research in several ways. Firstly, we confirm ethnic differences in COVID-19 mortality and provide novel data across a range of outcomes prior to death (testing, hospitalisation, and intensive care admission). Secondly, we explore whether household size has an effect over and above socio-demographic and clinical characteristics. Finally, we report on both general population and care home residents during the first and second waves of the pandemic in England.

We find that, though some ethnic minority groups are less likely to be tested for SARS-CoV-2, all non-White groups are more likely to test positive, even when restricted to those ever tested. This may suggest that White populations may be tested more frequently with mild or asymptomatic disease and/or that ethnic minority groups get tested at more severe stages of the disease. Disparities in testing may relate to lack of access to testing sites, poorer health literacy, lack of tailored and accessible health communications, or differences in testing related behaviours.[41] Emerging evidence suggests that individuals may avoid seeking a test for fear of losing income or employment if required to quarantine after testing positive.[42] Given that ethnic minority groups are more likely to work in insecure jobs with poor workplace protections, and in essential or key-worker roles associated with higher risk of COVID-19 mortality,[43–45] it is likely that social and economic barriers to testing are greater in ethnic minority groups.

Our finding that ethnic minority groups have higher risks of COVID-19 related hospitalisation, ICU admission, and death after accounting for clinical co-morbidities suggests that improving equity of clinical care and understanding potential interactions between COVID-19 and underlying conditions are essential for mitigating inequalities in the downstream effects of SARS-CoV-2 infection. The fact that inequalities worsened for South Asian groups in wave 2 compared to wave 1 suggests that more aggressive and tailored interventions are needed to meet the needs in these communities.[46] However, our finding of attenuated risks in all other ethnic groups is a potential positive finding; further investigation is warranted into which public health actions were most influential in mitigating health disparities for these groups.

Our finding that the magnitude of wave 1 ethnic differences in testing positive are similar to those of COVID-19 related mortality suggests that ethnic differences in death may be mediated through exposure or susceptibility to infection, rather than susceptibility to severe disease once infected. This hypothesis is supported by recent findings from the REACT-2 study which found higher levels of SARS-CoV-2 antibodies in ethnic minority groups, but no ethnic differences in the infection-to-mortality ratio.[47]

We show that after accounting for socio-demographic and clinical factors, household size further explained differences in COVID-19 outcomes for south Asian groups. This finding is consistent with an ONS study which found that multigenerational living was causally associated with increased risk of COVID-19 mortality in south Asian women, but not in any other ethnic groups.[48] Data from the 2011 census reports that 21% of south Asian groups live in multi-generational households compared to 6.8% of White groups.[25,49] We hypothesise that household size and deprivation may proxy viral exposure by capturing aspects of occupational and community level exposure. While multigenerational living may increase risk of exposure and transmission (from children or working age adults to older or vulnerable family members), such households and extended communities also offer

valuable informal care networks and facilitate engagement with health and community services.[50] In light of emerging evidence that ethnic minority groups are less likely to take up the COVID-19 vaccine, co-designing culturally competent and non-stigmatising engagement strategies with these communities is increasingly important.[51,52]

National data from England and Scotland have shown that some ethnic minority groups have both better overall health and lower rates of all-cause mortality than White groups.[53,54] We were able to confirm this pattern in our sensitivity analyses, thus, our findings of disparities in SARS-CoV-2 positivity and COVID-19-related outcomes, some of which have continued to widen over the course of the epidemic in the UK, are particularly concerning.

Our findings mirror large studies in the US, which have found that minority racial and ethnic communities have elevated risks of testing positive, hospitalisation, and death that differentially vary over time, even after accounting for socio-demographic characteristics and underlying health conditions.[5,40] These parallel findings suggest that mechanisms underpinning ethnic differences in COVID-19 outcomes in England may be common in other settings, and that learnings across settings should be shared.

Improving the quality and completeness of ethnicity data across health and administrative datasets is essential for building a complete picture of ethnic disparities.[55] Furthermore, though the recording of ethnicity on death certificates has been the norm in Scotland for the past decade, it is only now being considered for use in England.[56–58] Prioritizing linkage between health, social and employment data will be essential in building a complete picture of ethnic differences in COVID-19 risk and outcomes.

Conclusions

Ethnic minority groups in the UK have experienced disproportionately high levels of poor COVID-19 outcomes, with disparities increasing even within the course of the epidemic for some groups. Reducing ethnic inequalities will need action across a broad range of measures such as addressing the wider adverse effects of disadvantage and structural discrimination, reducing within- and between-household transmission, and improving control of clinical conditions. The relative importance of each of these measures will differ by both ethnic group and stage of COVID-19 progression. Equality is difficult to achieve, but structural and persistent inequalities must be addressed in a civilised society.

Data sharing

All data were linked, stored, and analysed securely within the OpenSAFELY platform. Detailed pseudonymised patient data are potentially re-identifiable and therefore not shared. We rapidly delivered the OpenSAFELY data analysis platform without previous funding to deliver timely analyses of urgent research questions in the context of the global COVID-19 health emergency: now that the platform is established, we are developing a formal process for external users to request access in collaboration with NHS England. Details of this process will be published in the near future on the OpenSAFELY website.

Acknowledgements

We are very grateful for all the support received from the TPP Technical Operations team throughout this work, and for generous assistance from the information governance and database teams at NHS England / NHSX.

Conflicts of Interest

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare the following: BG has received research funding from Health Data Research UK (HDRUK), the Laura and John Arnold Foundation, the Welcome Trust, the NIHR Oxford Biomedical Research Centre, the NHS National Institute for Health Research School of Primary Care Research, the Mohn-Westlake Foundation, the Good Thinking Foundation, the Health Foundation, and the World Health Organisation; he also receives personal income from speaking and writing for lay audiences on the misuse of science. IJD has received unrestricted research grants and holds shares in GlaxoSmithKline (GSK). KK is Director for the University of Leicester Centre for BME Health, Trustee of the South Asian Health Foundation, national NIHR ARC lead for Ethnicity and Diversity and a member of Independent SAGE and Chair for the SAGE Ethnicity Subgroup. RM, BG, and RME are members of the SAGE Ethnicity Subgroup.

Funding

This work was supported by the Medical Research Council MR/V015737/1. TPP provided technical expertise and infrastructure within their data centre pro bono in the context of a national emergency.

RM holds a fellowship funded by the Wellcome Trust. BG's work on better use of data in healthcare more broadly is currently funded in part by: NIHR Oxford Biomedical Research Centre, NIHR Applied Research Collaboration Oxford and Thames Valley, the Mohn-Westlake Foundation, NHS England, and the Health Foundation; all DataLab staff are supported by BG's grants on this work. LS reports grants from Wellcome, MRC, NIHR, UKRI, British Council, GSK, British Heart Foundation, and Diabetes UK outside this work. AS is employed by LSHTM on a fellowship sponsored by GSK. KB holds a Sir Henry Dale fellowship

jointly funded by Wellcome and the Royal Society. HIM is funded by the National Institute for Health Research (NIHR) Health Protection Research Unit in Immunisation, a partnership between Public Health England and LSHTM. AYSW holds a fellowship from BHF. EW holds grants from MRC. RG holds grants from NIHR and MRC. ID holds grants from NIHR and GSK. HF holds a UKRI fellowship. RE is funded by HDR-UK and the MRC. KK is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands (ARC EM) and the NIHR Leicester Biomedical Research Centre (BRC).

The views expressed are those of the authors and not necessarily those of the NIHR, NHS England, Public Health England or the Department of Health and Social Care.
Funders had no role in the study design, collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Information governance and ethical approval

NHS England is the data controller; TPP is the data processor; and the key researchers on OpenSAFELY are acting on behalf of NHS England. This implementation of OpenSAFELY is hosted within the TPP environment which is accredited to the ISO 27001 information security standard and is NHS IG Toolkit compliant; patient data has been pseudonymised for analysis and linkage using industry standard cryptographic hashing techniques; all pseudonymised datasets transmitted for linkage onto OpenSAFELY are encrypted; access to the platform is via a virtual private network (VPN) connection, restricted to a small group of researchers; the researchers hold contracts with NHS England and only access the platform to initiate database queries and statistical models; all database activity is logged; only aggregate statistical outputs leave the platform environment following best practice for anonymisation of results such as statistical disclosure control for low cell counts. The OpenSAFELY research platform adheres to the data protection principles of the UK Data Protection Act 2018 and the EU General Data Protection Regulation (GDPR) 2016. In March 2020, the Secretary of State for Health and Social Care used powers under the UK Health Service (Control of Patient Information) Regulations 2002 (COPI) to require organisations to process confidential patient information for the purposes of protecting public health, providing healthcare services to the public and monitoring and managing the COVID-19 outbreak and incidents of exposure.[4] Taken together, these provide the legal bases to link patient datasets on the OpenSAFELY platform. GP practices, from which the primary care data are obtained, are required to share relevant health information to support the public health response to the pandemic and have been informed of the OpenSAFELY analytics platform.

This study was approved by the Health Research Authority (REC reference 20/LO/0651) and by the LSHTM Ethics Board (reference 21863).

Guarantor

RM/LS/BG are guarantors

Contributorship

Contributions are as follows:
Conceptualization: RM, CTR, KB, RME, LS, BG, BM, HJC, SJWE, KK, DH, KR
Data curation: RM, CTR, AJW, CB, JC, CM, RME, WJH, BM, SB
Formal analysis: RM, CTR
Funding acquisition: LS, BG, RME
Investigation: RM, CTR, CM, WJH
Methodology: RM, CTR, KB, RME, KK, NC, RG, DH, KR, LS, BG, BM, EW, HJC, SJWE
Codelists: RM, LT, AS, AJW, CM, BG, WJH, SB, AM
Project administration: RM, CTR, AS, AJW, CM, BG, WJH
Resources: CB JC BG BM SB AM
Software: AJW CB JC DE PI CM WJH BN SB HJC ND RC JP FH SH
Visualisation: RM RME
Writing - original draft: RM
Writing- review & editing: ALL
Information governance: CB LS BG AM

References

1 Pan D, Sze S, Minhas JS, *et al.* The impact of ethnicity on clinical outcomes in COVID-19: A systematic review. *EClinicalMedicine* 2020; **23**: 100404.

2 Williamson EJ, Walker AJ, Bhaskaran K, *et al.* Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020; **584**: 430–6.

3 Aldridge RW, Lewer D, Katikireddi SV, *et al.* Black, Asian and Minority Ethnic groups in England are at increased risk of death from COVID-19: indirect standardisation of NHS mortality data. *Wellcome Open Res* 2020; **5**: 88.

4 Public Health England. Beyond the data: Understanding the impact of COVID-19 on BAME groups. Public Health England, 2020 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/892376/COVID_stakeholder_engagement_synthesis_beyond_the_data.pdf.

5 Rentsch CT, Kidwai-Khan F, Tate JP, *et al.* Patterns of COVID-19 testing and mortality by race and ethnicity among United States veterans: A nationwide cohort study. *PLoS Med* 2020; **17**: e1003379.

6 Baqui P, Bica I, Marra V, Ercole A, van der Schaar M. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *Lancet Glob Health* 2020; **8**: e1018–26.

7 Chen J, Krieger N. Revealing the Unequal Burden of COVID-19 by Income, Race/Ethnicity, and Household Crowding: US County Versus Zip Code Analyses. *J Public Health Manag Pract* 2020. DOI:10.1097/PHH.0000000000001263.

8 Kenway P, Holden J. Accounting for the Variation in the Confirmed Covid-19 Caseload across England: An analysis of the role of multi-generation households, London and time. *New Policy Institute, London* 2020.

9 Razai MS, Kankam HKN, Majeed A, Esmail A, Williams DR. Mitigating ethnic disparities in covid-19 and beyond. *BMJ* 2021; **372**. DOI:10.1136/bmj.m4921.

10 Ethnicity sub-group of the Scientific Advisory Group for Emergencies (SAGE). Drivers of the higher COVID-19 incidence, morbidity and mortality among minority ethnic groups. SAGE, 2020 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/925135/S0778_Drivers_of_the_higher_COVID-19_incidence__morbidity_and_mortality_among_minority_ethnic_groups.pdf.

11 Warden J. NHS to come under race relations act. *BMJ* 1999; **318**: 625.

12 Race Relations (Amendment) Act 2000. 2000 https://www.legislation.gov.uk/ukpga/2000/34/contents (accessed Sept 9, 2020).

13 Mathur R, Grundy E, Smeeth L. Availability and use of UK based ethnicity data for health research. National Centre for Research Methods Working Paper Series. 2013; : 1–30.

14 Lee C. 'Race' and 'ethnicity' in biomedical research: how do scientists construct and explain differences in health? *Soc Sci Med* 2009; **68**: 1183–90.

15 Bhopal RS. Terminology and classifications for migrant, ethnic, and racial groups: the centrality of census and population registers. In: Migration, Ethnicity, Race, and Health in Multicultural Societies. Oxford University Press, 2013: 33–61.

16 Kabarriti R, Brodin N, Maron M, *et al.* Association of Race and Ethnicity With Comorbidities and Survival Among Patients With COVID-19 at an Urban Medical Center in New York. *JAMA netw open* 2020; **3**: e2019795.

17 Yehia B, Winegar A, Fogel R, *et al.* Association of Race With Mortality Among Patients Hospitalized With Coronavirus Disease 2019 (COVID-19) at 92 US Hospitals. *JAMA netw open* 2020; **3**: e2018039.

18 Munoz-Price L, Nattinger A, Rivera F, *et al.* Racial Disparities in Incidence and Outcomes Among Patients With COVID-19. *JAMA netw open* 2020; **3**: e2021892.

19 Harrison EM, Docherty AB, Barr B, *et al.* Ethnicity and Outcomes from COVID-19: The ISARIC CCP-UK Prospective Observational Cohort Study of Hospitalised Patients. 2020; published online May 31. DOI:10.2139/ssrn.3618215.

20 Sapey E, Gallier S, Mainey C, *et al.* Ethnicity and risk of death in patients hospitalised for COVID-19 infection in the UK: an observational cohort study in an urban catchment area. *BMJ Open Respir Res* 2020; **7**. DOI:10.1136/bmjresp-2020-000644.

21 Apea VJ, Wan YI, Dhairyawan R, *et al.* Ethnicity and outcomes in patients hospitalised with COVID-19 infection in East London: an observational cohort study. *medRxiv* 2020. https://www.medrxiv.org/content/10.1101/2020.06.10.20127621v1.abstract.

22 Lassale C, Gaye B, Hamer M, Gale CR, Batty GD. Ethnic Disparities in Hospitalization for COVID-19: a Community-Based Cohort Study in the UK. *medRxiv* 2020; published online May 26. DOI:10.1101/2020.05.19.20106344.

23 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* 2020; **42**: 451–60.

24 OpenSAFELY. https://opensafely.org/ (accessed Oct 20, 2020).

25 Smith CW. 2011 Census analysis: What does the 2011 Census tell us about concealed families living in multi-family households in England and Wales? 2014; published online Feb 6.

https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/articles/2011censusanalysis/2014-02-06 (accessed Sept 15, 2020).

26 Covid-19 test results. https://docs.opensafely.org/en/latest/dataset-sgsscovid/ (accessed Jan 13, 2021).

27 Coronavirus (COVID-19): scaling up our testing programmes. https://www.gov.uk/government/publications/coronavirus-covid-19-scaling-up-testing-programmes/coronavirus-covid-19-scaling-up-our-testing-programmes (accessed Jan 13, 2021).

28 Index of Multiple Deprivation. English indices of deprivation 2015. *Ministry of Housing, Communities & Local Government* 2015.

29 Chisholm J. The Read clinical classification. *BMJ* 1990; **300**: 1092.

30 OpenSAFELY Codelists. https://codelists.opensafely.org/ (accessed Jan 24, 2021).

31 Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol* 2009; **170**: 244–56.

32 Griffith G, Morris TT, Tudball M, *et al.* Collider bias undermines our understanding of COVID-19 disease risk and severity. Epidemiology. 2020; published online May 8. DOI:10.1101/2020.05.04.20090506.

33 Ortiz-Fernández L, Sawalha AH. Genetic variability in the expression of the SARS-CoV-2 host cell entry factors across populations. *Genes Immun* 2020; **21**: 269–72.

34 Maiti AK. The African-American population with a low allele frequency of SNP rs1990760 (T allele) in IFIH1 predicts less IFN-beta expression and potential vulnerability to COVID-19 infection. *Immunogenetics* 2020; **72**: 387–91.

35 Hawkins D, Ms S. Differential occupational risk for COVID-19 and other infection exposure according to race and ethnicity. *Am J Ind Med* 2020. DOI:10.1002/ajim.23145.

36 Laster Pirtle W. Racial Capitalism: A Fundamental Cause of Novel Coronavirus (COVID-19) Pandemic Inequities in the United States. *Health Educ Behav* 2020; **47**: 504–8.

37 Bhala N, Curry G, Martineau AR, Agyemang C, Bhopal R. Sharpening the global focus on ethnicity and race in the time of COVID-19. *Lancet* 2020; **395**: 1673–6.

38 Asimakopoulou K, Hoorens V, Speed E, *et al.* Comparative optimism about infection and recovery from COVID-19; Implications for adherence with lockdown advice. *Health Expect* 2020. DOI:10.1111/hex.13134.

39 Aleta A, Martín-Corral D, Pastore Y Piontti A, *et al.* Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav* 2020; published online Aug 5. DOI:10.1038/s41562-020-0931-9.

40 COVID-19 Racial Disparities in Testing, Infection, Hospitalization, and Death: Analysis of Epic Patient Data. 2020; published online Sept 16. https://www.kff.org/coronavirus-covid-19/issue-brief/covid-19-racial-disparities-testing-infection-hospitalization-death-analysis-epic-patient-data/ (accessed Sept 20, 2020).

41 English language skills. 2018; published online Feb 7. https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/english-language-skills/latest (accessed Sept 8, 2020).

42 Tapper J. Low-paid shun Covid tests because the cost of self-isolating is too high. The Guardian. 2021; published online Jan 16. http://www.theguardian.com/society/2021/jan/16/low-paid-shun-covid-tests-cost-of-self-isolating-too-high (accessed Jan 24, 2021).

43 Dr Ben Windsor-Shellard And. Coronavirus (COVID-19) related deaths by occupation, England and Wales - Office for National Statistics. 2020; published online June 25. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/bulletins/coronaviruscovid19relateddeathsbyoccupationenglandandwales/deathsregisteredbetween9marchand25may2020 (accessed Sept 3, 2020).

44 Mahase E. Covid-19: Ethnic minority doctors feel more pressured and less protected than white colleagues, survey finds. *BMJ* 2020; **369**. DOI:10.1136/bmj.m2506.

45 BME workers have been asked to “shoulder more risk” during pandemic, says TUC. 2020; published online Sept 14. https://www.tuc.org.uk/news/bme-workers-have-been-asked-shoulder-more-risk-during-pandemic-says-tuc (accessed Jan 24, 2021).

46 Iacobucci G. Covid-19: PHE review has failed ethnic minorities, leaders tell BMJ. *BMJ* 2020; **369**: m2264.

47 Ward H, Atchison CJ, Whitaker M, *et al.* Antibody prevalence for SARS-CoV-2 in England following first peak of the pandemic: REACT2 study in 100,000 adults. *medRxiv* 2020; : 2020.08.12.20173690.

48 Nafilyan V, Islam N, Ayoubkhani D, *et al.* Ethnicity, household composition and COVID-19 mortality: A national linked data study. bioRxiv. 2020; published online Dec 2. DOI:10.1101/2020.11.27.20238147.

49 Overcrowded households. 2020; published online Sept 9. https://www.ethnicity-facts-figures.service.gov.uk/housing/housing-conditions/overcrowded-households/latest (accessed Sept 15, 2020).

50 Bear L, James D, Simpson N, *et al.* A right to care: the social foundations of recovery from Covid-19. 2020. http://eprints.lse.ac.uk/107060/ (accessed Nov 4, 2020).

51 SPI-B. SPI-B/EMG: MHCLG Housing Impacts Paper. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/923605/s0744-4a-mhclg-housing-impacts-summary-paper.pdf.

52 SAGE advisory group. Factors influencing COVID-vaccine uptake among minority ethnic groups. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/952716/s0979-factors-influencing-vaccine-uptake-minority-ethnic-groups.pdf.

53 Bhopal RS, Gruer L, Cezard G, *et al.* Mortality, ethnicity, and country of birth on a national scale, 2001-2013: A retrospective cohort (Scottish Health and Ethnicity Linkage Study). *PLoS Med* 2018; **15**: e1002515.

54 Scott AP, Timæus IM. Mortality differentials 1991-2005 by self-reported ethnicity: findings from the ONS Longitudinal Study. *J Epidemiol Community Health* 2013; **67**: 743–50.

55 Khunti K, Routen A, Banerjee A, Pareek M. The need for improved collection and coding of ethnicity in health research. *J Public Health (Oxf)* 2020; published online Dec 7. DOI:10.1093/pubmed/fdaa198.

56 Stevenson O, Morris S. Coronavirus: record ethnicity on all death certificates to start building a clearer picture. *The Conversation* 2020; published online May 11. http://theconversation.com/coronavirus-record-ethnicity-on-all-death-certificates-to-start-building-a-clearer-picture-137860 (accessed Aug 24, 2020).

57 Christie B. Scotland introduces record of ethnicity on death certificates. *BMJ* 2012; **344**: e475.

58 London Health Obsrevatory. Missing Record: The Case for Recording Ethnicity at Birth and Death Registrations. 2003 https://www.kent.ac.uk/chss/docs/Births_Deaths_Reg1.pdf.