ETHICAL: Ethnic Disparities **I**n **C**OVID-19 Admissions in east London

Statistical Analysis Plan

V Apea, R Dhairyawan, C Orkin, RM Pearse, JR Prowle, YI Wan

Short title	**ETHICAL: Eth**nic Disparities **I**n **C**OVID-19 Admissions in east London
Methodology	Observational cohort study of treatment and outcomes in Black, Asian and Minority Ethnic (BAME) and non-BAME patients with a confirmed diagnosis of COVID-19 (positive SARS-CoV-2 swab) admitted to Barts Health NHS Trust
Objectives / aims	• Assess whether there is a difference in outcomes (ITU admission/Advanced Respiratory Support/ Death) between different ethnic backgrounds of patients with confirmed diagnosis of COVID-19 (positive SARS-CoV-2 swab) admitted to Barts Health • To analyse differences in age, sex, comorbidity and measure of economic deprivation in COVID-19 positive patients of Black, Asian and other ethnicities, compared to COVID-19 positive patients of non-BAME ethnicity admitted to hospitals in Barts Health NHS Trust and to understand their association with adverse outcomes
Number of participants	Around 2000
Inclusion and exclusion criteria	**Inclusion Criteria:** • Confirmed COVID-19 by positive SARS-CoV-2 PCR • Inpatient admission at Royal London Hospital, Newham General Hospital, Whipps Cross Hospital, St Bartholomew's Hospital (+/- Nightingale Hospital) **Exclusion Criteria:** • Unknown or undisclosed ethnicity
Statistical methodology and analysis (if applicable)	• Prospectively Defined Observational study • Descriptive and comparative analyses of demographics, clinical characteristics and clinical outcomes between patients of different ethnicities using univariate and multivariate analysis.
Study duration	18 months (for full follow up and final analysis)

1. Introduction

1.1. Background

On 31st December 2019, Chinese authorities notified the World Health Organisation (WHO) of an outbreak of pneumonia in Wuhan City, which was later classified as a new disease; coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1]. It was declared a pandemic by the WHO on March 11th 2020[2] and in the United Kingdom (UK) the disease has significantly impacted all aspects of the national health and social care economy[3].

COVID-19 is a heterogeneous disease, with differential impact on individuals. The symptoms vary widely, from asymptomatic disease to pneumonia and life-threatening complications. There is a clear imperative to identify patient subpopulations with poorer prognoses to optimise their management. The presence of co-morbidity has been associated with poorer clinical outcomes[4,5]; with circulatory diseases (including hypertension, coronary heart disease and heart failure) and diabetes being key factors[4,5]. Other factors such as older age, male sex, obesity and smoking are also associated with worse outcomes[6,7].

Preliminary analyses of the UK Intensive care national audit and research centre (ICNARC) report on COVID-19 in critical care have signalled ethnicity as a predictor of poorer clinical outcomes[8]. 34.5% of patients requiring ITU admission were of Black, Asian, Mixed or Other ethnicity; this is disproportionate to Black, Asian and Minority Ethnic (BAME) only making up 13% of the UK population[9]. This over-representation extends to the number of NHS frontline staff known to have died in the pandemic so far; with 68% being of a BAME background[10]. Ethnicity has also been shown to be a pivotal driver of the US epidemic with predominantly black counties experiencing death rates 3-fold higher than predominantly white counties[11]. This emerging disparity warrants further investigation. To date, no UK hospital COVID-19 mortality data, stratified by ethnicity, has been publicly presented.

“Race" and "ethnicity" are complex and distinct parameters that are often used interchangeably. Race is a historical social construct. It refers to categories of human beings based on shared physical traits but has no genetic basis[12]. Ethnicity denotes groups that share a common identity-based ancestry, language, or culture[13]. It is currently the best marker we have for defining population subgroups as it can serve as a proxy for shared risks for people with similar, social, biological and cultural characteristics. Ethnicity is self-reported by patients at registration in clinical settings and coded based on the UK Census categorisation hierarchies[13].

The 2020 Marmot review reported that the last decade has been marked by deteriorating health and widening health inequalities[14]. Health is influenced by a wide range of factors, including biological determinants (age, gender, hereditary factors), and wider social determinants such as education, social position, income, poverty, local environment, and experiences of racism and institutional racism. Each factor shapes health individually and by intersecting with other factors to have a synergistic impact.

The signal of ethnic inequality being highlighted in COVID-19 is not new. Health inequalities within and between ethnic minority groups are known to be transmitted across generations[13]. Ethnic minorities tend to have poorer health than majority ethnic groups. This has been clearly demonstrated in a range of conditions such as cardiovascular disease, diabetes, stroke and hypertension[13]. The relationship between ethnicity and health is complex, interacting in ways we are only beginning to understand. These inequalities have led to unjust and preventable inequalities in health[13,15]. In the context of COVID-19, a number of potential causes have been postulated.

BAME people are twice as likely to be living in poverty and are more likely to be employed in a key worker role with an increased risk of COVID-19 acquisition or experience housing deprivation[16]. This exposure is compounded by a background increased risk of heart disease and diabetes; both shown to be risk factors of severe COVID-19 infection. Public concern has been raised about BAME patients presenting late to hospital with advanced symptoms due to language barriers or fear of reprisal for employers, being unable to advocate for adequate personal protective equipment in certain keyworker roles[17] and receiving different approaches to management and escalation of care. Atchison et al have emphasised that those most economically disadvantaged in society are less able to comply with certain governmental advice[18]. Although willingness to self-isolate was high overall in their cross-sectional survey, those from more disadvantaged backgrounds were less likely to be able to work from home or self-isolate if needed, suggesting the existence of structural barriers to adopting preventive behaviours in these groups.

1.1. Rationale

The Barts Health footprint encompasses diverse populations; some challenged by some of the worst socio-economic deprivation and highest neighbourhood Index of Multiple Deprivation (IMD) scores. As such there is local incentive to identify and swiftly intervene as necessary. Inequalities will not be eliminated overnight. However, understanding where they emerge is a crucial first step towards tackling them. Failing to respond to the emerging disparities in this pandemic significantly risks not only reinforcing pre-existing structures of racial inequality in society, but further embedding them.

Our study is responding to national call to generate robust scientific data to inform the debate of the impact of COVID-19 on BAME communities. This will be the first UK study to assess whether there is a difference in hospital outcomes in patients of different ethnicities diagnosed with COVID-19. It will test the hypothesis that patients of Black, Asian and other ethnicities have poorer in-patient outcomes; in terms of need for ITU admission and mortality. Data generated has the potential to influence and shape relevant policy and practice to improve health outcomes in BAME communities in this devastating pandemic.

2. Study objectives

2.1. Primary objective

This study aims to assess whether there is an age-adjusted difference in outcomes in different ethnicities of patients with confirmed COVID 19 admitted to Barts Health.

The study will test the hypothesis that there is an association between health outcomes; in terms of need for ITU admission and mortality; and ethnicity in COVID-19 positive patients admitted to Barts Health; with patients of Black, Asian and other ethnicities having poorer outcomes as compared to patients of white ethnicity.

2.2. Secondary objective

We will examine the strength and nature of relationship between ethnic background and outcome in multivariable age and co-morbidity adjusted multivariable survival analysis.

We hypothesise that BAME background patients may be disproportionately younger, have greater baseline co-morbidity and have greater economic deprivation than non-BAME counterparts and we will document these differences and explore their effects on outcome in the analysis.

2.3. Primary endpoint

The primary endpoint will be 30-day Mortality from time of first hospital admission with COVID 19 diagnosis.

2.4. Secondary endpoints

The secondary endpoints will be:

- ITU admission
- Mortality (in hospital)
    - Death in critical care
    - Death on ward
- Discharged from hospital alive
    - Discharge destination
- Length of stay in hospital
- Length of stay in ITU
- Organ support received at any time
    - Yes
    - No
- Duration of organ support
- Respiratory support
    - Yes
        - Basic
        - Advanced
    - No
- Acute kidney injury
- Need for Renal support
- HLH score >169
- Highest HLH

3. Study population

3.1. Inclusion criteria

- Confirmed SARS COV19 diagnosis by PCR
- Inpatient admission at Royal London Hospital, Newham General Hospital, Whipps Cross Hospital, St Bartholomew's Hospital (+/- Nightingale Hospital)

3.2. Exclusion criteria

- Unknown or undisclosed ethnicity status

4. Study design

This is a prospectively defined observational cohort study utilising data from the Barts Health Electronic Patient Record (EPR) system. It will involve analysis of data of patients admitted to hospitals within Barts Health hospitals.

Data analysis will include, but not limited to:

- A comparison of clinical characteristics between inpatients with confirmed COVID-19 of different ethnicities.
- An assessment of clinical outcomes of inpatients with confirmed COVID-19.
- An assessment of ethnicity as a predictor of outcome using univariate and multivariate analysis. This will include adjustment for potential confounders such as co-morbidity and postcode of residence/Index of Multiple Deprivation (IMD).
- An assessment of other potential predictors of outcome using univariate and multivariate analysis.

5. Study procedures

Consent
This study involves the aggregated analysis of an anonymised dataset collated by members of the direct care team and should not require REC review or consent. See ethics section below.

Data collection
Hospital Cerner Millennium Data Warehouse records will then be used to obtain clinical, demographic and blood results by members of the clinical team participating in this study.

The existing ICNARC (Intensive Care National Audit & Research Centre) hospital database will be screened for ICU admissions fitting the inclusion criteria. National and local ICNARC data analysis are Care Quality Commission National Information Governance Committee approved processes.

No patient notes will be required or examined. This data will be recorded in a Microsoft Excel Database maintained under password protection on the secure Trust internal network. After results for each patient have been collated by a member of the direct clinical care team. Once collation of the database is complete an anonymised database (lacking all patient identifiable details - (name, hospital number, date of birth and admission no.) will be analysed by the investigators. For analysis see statistical details below.

6. Statistical considerations

The Chief Investigator takes responsibility for the statistics and statistical oversight for this study.

6.1. Sample size

The cohort to be analysed in this study comprises patients admitted during the study period. This includes all inpatients with COVID-19 admitted to Barts Health NHS Trust between 1st January 2020 and 31st July 2020.

Due to the urgent need for data examining outcome in this healthcare emergency an Initial analysis based on admissions to 27th April 2020 will be performed

This will be a retrospective review of routinely collected data and we anticipate that over 2000 patients will meet our inclusion criteria, but the final sample size will be dictated by the data available.

6.2. Method of analysis

Software
Statistical analyses will be performed using R: A language and environment for statistical computing (R Core Team (2020). R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/).

Statistical Analysis

Definitions

Ethnicity
The ethnic categories to be used will be the 16+1 ethnic data categories defined in the 2001 census which form national mandatory standard for the collection and analysis of ethnicity in the NHS data dictionary.

These are:

White

- A British
- B Irish
- C Any other White background

Mixed

- D White and Black Caribbean
- E White and Black African
- F White and Asian
- G Any other mixed background

Asian or Asian British

H Indian

J Pakistani

K Bangladeshi

L Any other Asian background

Black or Black British

M Caribbean

N African

P Any other Black background

Other Ethnic Groups

R Chinese

S Any other ethnic group

Z Not stated (excluded from this analysis)

In order to preserve statistical power to detect differences between groups primary analysis will be performed between ethnicity as defined by the 5-high level groups White, Mixed, Asian or Asian British, Black or Black British and Other.

Death will be defined by the presence of a date of death or discharge destination “patient died” in hospital electronic records (these are synchronised with NHS spine to capture out of hospital deaths).

Secondary haemophagocytic lymphohistiocytosis
Secondary haemophagocytic lymphohistiocytosis (sHLH) risk score will be calculated from peak values using the following table (see Mehta P, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet. 2020 Mar 28;395(10229):1033-1034.)

	Number of points
Temperature	
<38·4°C	0
38·4–39·4°C	33
>39·4°C	49
Organomegaly	
None	0
Hepatomegaly or splenomegaly	23
Hepatomegaly and splenomegaly	38
Number of cytopenias*	
One lineage	0
Two lineages	24
Three lineages	34
Triglycerides (mmol/L)	
<1·5 mmol/L	0
1·5–4·0 mmol/L	44
>4·0 mmol/L	64
Fibrinogen (g/L)	
>2·5 g/L	0
≤2·5 g/L	30
Ferritin ng/ml	
<2000 ng/ml	0
2000–6000 ng/ml	35
>6000 ng/ml	50
Serum aspartate aminotransferase	
<30 IU/L	0
≥30 IU/L	19
Haemophagocytosis on bone marrow aspirate	
No	0
Yes	35
Known immunosuppression†	
No	0
Yes	18

The HScore[11] generates a probability for the presence of secondary HLH. HScores greater than 169 are 93% sensitive and 86% specific for HLH. Note that bone marrow haemophagocytosis is not mandatory for a diagnosis of HLH. HScores can be calculated using an online HScore calculator.[11] HLH=haemophagocytic lymphohistiocytosis. *Defined as either haemoglobin concentration of 9·2 g/dL or less (≤5·71 mmol/L), a white blood cell count of 5000 white blood cells per mm³ or less, or platelet count of 110 000 platelets per mm³ or less, or all of these criteria combined. †HIV positive or receiving long-term immunosuppressive therapy (ie, glucocorticoids, cyclosporine, azathioprine).

Table: **HScore for secondary HLH, by clinical parameter**

Hospital Frailty risk score

Hospital Frailty risk score will be calculated from prior ICD-10 coding of hospital attendances using mapped described in Lancet 2018; 391: 1775–82.

Acute Kidney injury

Will be defined using the KDIGO 2012 creatinine criteria. Baseline Creatinine will be the median value in the 7 to 365 days before hospitalisation. Absent baseline creatinine will be imputed based on and eGFR of 75ml/min/1.72m^2 or the admission value whichever is lower.

Comorbidity

Charlson Comorbidity Index and co-morbidities will be calculated from prior ICD-10 codes based on the mapping of Quan H et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005 Nov;43(11):1130-9.

Diagnosis of Hypertension will be based on mapping ICD-10 codes to the Elixhauser comorbidity index. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998;36:8-27

End stage Renal disease will be defined based on prior ICD-10 coding following: Crellin E, et al (2017). Clinical Code List - ICD-10 - End-Stage Renal Disease. [Data Collection]. London School of Hygiene & Tropical Medicine. https://doi.org/10.17037/DATA.241.

History of tobacco use will be defined by presence of the ICD codes F17.1-F17.2, Z72.0, Z87.8, Z71.6 and T65.2.

Index of Multiple Deprivation
Index of Multiple Deprivation will be defined from patient home address postcode using UK government statistics https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019. IMD will be presented as raw score and quintiles within the greater London area

Analysis
Clinical outcomes for the whole cohort will be analysed and also stratified by ethnicity. Univariate and multivariate analyses will subsequently be used to investigate predictors of outcome for each ethnicity.

All continuous data will be expressed as median and interquartile range (IQRs) and categorical data as proportion. For continuous data, we will compare the mean difference between groups by ANOVA or Kruskal-Wallis test. For dichotomous data, we will use Pearson Chi-square test or Fisher exact test (if expected number less than 5). The log-rank and Wilcoxon test will be used for survival time comparisons between groups.

Sample Table 1

	Black	**Asian**	**Mixed**	**Other**	**White**	P value
Age						
Sex						
IMD						
BMI						
Smoking						
Hospital admission						
CCI						
Hospital Frailty Risk Score						
IMD						
DM						
DMc						
HT						
IHD						
COPD						
CKD						
Baseline eGFR						
ESRD						
AKI						
ICU?						
Basic Resp						
Advanced Resp						
Circulatory						
Renal						

Creatinine						
CRP						
D-dimer						
Peak HLH score						
Peak HLH >169?						

Primary and Secondary Endpoint Analysis

Primary endpoint will be 30-day mortality between ethic groupings adjusted for patient age and sex.

Mortality comparison will be analysed Cox-proportional Hazard analysis censored at 30 days or time of maximal follow-up.

Age and sex adjusted incidence of secondary endpoints will be assessed using logistic regression on Cox hazard models as appropriate.

Multi-variable analyses

If available, the following variables will be collected and considered in multi-variable analysis for primary and secondary endpoints using logistic regression on Cox hazard models as appropriate

- Age
- Gender
- BMI
- Smoking status
- Clinical frailty Scale at admission if recorded
- Co-morbidities based on ICD-10 coding for current and previous admission based on established mapped to Charlson Co-morbidty index and Elixhauser Index
- Specific comorbidities for consideration
    - Hypertension with and without complications
    - Diabetes with and without complications
- Prior end stage renal disease
- Index of Multiple deprivation

Sensitivity analyses

Multiple imputation will be undertaken to account for missing data for prespecified variables included in the above models. Results from imputed analysis will be presented only for the primary exposure variable of ethnicity. Multi-variable analysis will be repeated using aggregate measures of frailty in place of specific co-morbidities listed above. Survival will also be assessed using 90-day mortality.

7. References

1. Guan WJ, Ni ZY, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China [published online ahead of print, 2020 Feb 28]. N Engl J Med. 2020;NEJMoa2002032. doi:10.1056/NEJMoa2002032
2. Eurosurveillance Editorial Team. Updated rapid risk assessment from ECDC on the novel coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU/EEA and the UK. Euro Surveill., 25 (10) (2020), 10.2807/1560-7917.ES.2020.25.10.2003121
3. Department of Health and Social Care. Coronavirus (COVID-19) action plan, 3 March 2020:

https://www.gov.uk/government/publications/coronavirus-action-plan/coronavirus-action-plan-a-guide-to-what-you-can-expect-across-the-uk
4. Guan WJ, Liang WH, Zhao Y, et al. Comorbidity and its impact on 1590 patients with Covid-19 in China: A Nationwide Analysis [published online ahead of print, 2020 Mar 26]. Eur Respir J. 2020;2000547. doi:10.1183/13993003.00547-2020
5. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China [published correction appears in Lancet. 2020 Jan 30;:]. Lancet. 2020;395(10223):497–506. doi:10.1016/S0140-6736(20)30183-5
6. Xu XW, Wu XX, Jiang XG, et al. . Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case studies. BMJ 2020; 368: m606. doi:10.1136/bmj.m606
7. Zhao, Q., Meng, M., Kumar, R., Wu, Y., Huang, J., Lian, N., Deng, Y. and Lin, S. The impact of COPD and smoking history on the severity of Covid-19: A systemic review and meta-analysis. J Med Virol. 2020. Accepted Author Manuscript. doi:10.1002/jmv.25889
8. Intensive Care National Audit and Research Centre. ICNARC report on COVID-19 in critical care. 17 April 2020.
9. Office for National Statistics ; National Records of Scotland ; Northern Ireland Statistics and Research Agency. 2011 Census aggregate data. UK Data Service (Edition: June 2016). DOI: http://dx.doi.org/10.5257/census/aggregate-2011-1
10. Barr C and Siddique H. Failure to publish data on BAME deaths could put more lives at risk, MPs warn. The Guardian. https://www.theguardian.com/world/2020/apr/16/data-on-bame-deaths-from-covid-19-must-be-published-politicians-warn. Published 2020. Accessed April 2020.
11. Yancy CW. COVID-19 and African Americans. ***JAMA.*** Published online April 15, 2020. doi:10.1001/jama.2020.6548
12. Rosenberg NA ***et al***. Genetic structure of human populations. Science 2002; 298, 2381.
13. Nazroo JY, Mason D, (ed.). Patterns of and explanations for ethnic inequalities in health'. In Explaining Ethnic Differences: Changing patterns of disadvantage in Britain. Policy Press. 2003
14. Marmot M et al. Health Equity in England: The Marmot Review ten years on. 2020. London: Institute of Health Equity
15. Nazroo JY. The structuring of ethnic inequalities in health: economic position, racial discrimination, and racism. Am J Public Health. 2003;93(2):277–284. doi:10.2105/ajph.93.2.277
16. www.irr.org.uk. Inequality, housing and employment statistics. 2014. Available at: http://www.irr.org.uk/research/statistics/poverty/. Accessed April 2020.
17. Bécares L and and Nazroo J. Racism is the root cause of ethnic inequities in covid19. 2020. Available at www.discoversociety.org/2020/04/17/racism-is-the-root-cause-of-ethnic-inequities-in-covid19. Accessed April 2020.
18. Atchison CJ et al. Perceptions and behavioural responses of the general public during the COVID-19 pandemic: A cross-sectional survey of UK Adults. medRxiv 2020.04.01.20050039; doi: https://doi.org/10.1101/2020.04.01.20050039