Public Health England

Protecting and improving the nation's health

Analysis of the relationship between pre-existing health conditions, ethnicity and COVID-19

Technical document

Contents

Introduction

This document accompanies the summary report on the analysis of the relationship between pre-existing health conditions, ethnicity and COVID-19. It includes details of definitions and methods used and the literature review conducted for this study.

Definitions

COVID-19 diagnoses

Defined as confirmed cases of COVID-19, identified through PHE's Second Generation Surveillance System (SGSS) as having had a positive antigen test for SARS-CoV-2. It includes all those tested through pillars 1 and 2 of the testing system in that period.[1] For the descriptive analysis, including the standardised ratios, all cases with a specimen date from 30 January 2020 to 16 August 2020 were included. For the analysis of survival, all cases with a specimen date from 20 March 2020 to 13 July 2020 were included, allowing follow-up of 28 days to establish survival.

COVID-19 deaths

Defined as people who died in all settings, identified through ONS death registration data as having COVID-19 mentioned on the death certificate. This includes some people who did not have a positive test for SARS-CoV-2 and excludes people who had a positive test but did not have COVID-19 on their death certificate. For the descriptive analysis, including the standardised ratios, all deaths which occurred from 21 March 2020 to 17 July 2020 were included. For the analysis of survival, deaths occurring from 21 March to 10 August, and registered by 31 August 2020, were included.

Ethnicity

Ethnicity was allocated to COVID-19 case records and ONS death registrations by linking to hospital records. The person's NHS number, date of birth, postcode and sex were used to match case and death records to several NHS databases: the Secondary Use Service (SUS) Live, Hospital Episode Statistics (HES) Admitted Patient Care, HES Outpatient and HES A&E datasets. A probabilistic matching process was used, which allowed cases and deaths to be linked to hospital records even if some of the data fields were missing. The best available ethnicity record was chosen, prioritising getting a valid ethnicity, cases which more closely matched on the variables and then the most recent ethnicity recorded.

[1] Pillar 1 – PCR (Swab) testing in PHE labs and NHS hospitals for those with a clinical need, and health and care workers. Pillar 2 – PCR (Swab) testing in commercial laboratories for the wider population.

Ethnicity was analysed primarily in 6 groups:

- Asian / Asian British
- Black / African / Caribbean / Black British
- Mixed / Multiple ethnic groups
- White
- Other ethnic group
- No ethnicity information

The last category captures those who matched to hospital records but for with no ethnicity recorded in any of the records.

Differences between subgroups of the Asian and Black groups were also analysed (Bangladeshi, Chinese, Indian, Pakistani and Asian Other, Black African, Black Caribbean, Black Other).

PHE analysts are reviewing their approach to the linkage and allocation of ethnicity to cases and deaths. The analyses carried out for this report may be repeated if the allocation method is changed.

Pre-existing health conditions

These were defined using Hospital Episode Statistics (HES) data, matched to case and death records as described previously for assigning ethnicity. These conditions were selected based on a review of the evidence on health conditions associated with poor outcomes for COVID-19 and other respiratory infections (such as pneumococcal disease) and expert clinical assessment. They were analysed in groups listed below. Each condition is defined as: people who were admitted to hospital in the 5-year period prior to the study period (1 March 2015 to 29 February 2020), for whom at least one of the conditions is mentioned on the admission episode record, and who hadn't died before 1 March 2020.

The pre-existing conditions examined were:

1. A long list of relevant conditions referred to as the 'long list' throughout this document (ICD-10 codes provided in the Appendix). This includes all conditions identified by the evidence review and expert clinical assessment.

2. Specific conditions of interest:

- Cardiovascular disease (ICD-10 I00 to I99)
- Hypertensive diseases (ICD-10 I10 to I15)
- Cerebrovascular disease (ICD-10 I60 to I69) or transient ischaemic attack (ICD-10 G45 to G46)
- Diabetes (ICD-10 E10 to E14)

- Chronic kidney disease (ICD-10 N18)
- Chronic lower respiratory disease (ICD-10 J40 to J47)
- Dementia / Alzheimer's disease (ICD-10 F00 to F03, G30, G31.0, G31.8)

Deprivation

Deprivation was defined using the 2019 Index of Multiple Deprivation (IMD). IMD provides deprivation scores for each lower layer super output area (LSOA) in England. These were grouped into quintiles from most deprived to least deprived. Each person was allocated a deprivation quintile based on their postcode of residence. This allowed analyses to be carried out separately for those living in the most deprived areas, least deprived areas, and so on. Individual people may not always be deprived just because their postcode is in a deprived area, but at the population level this is an effective way to analyse the effect of deprivation on outcomes.

Age

For the descriptive analysis, age was defined as at the specimen date or death date. For the denominator populations used for the standardised ratios (the people with pre-existing health conditions as defined above) age was taken as at 30 June 2020. This does mean that there is inconsistency between the age definitions for the numerators and for the denominators for the standardised ratio analyses, but the discrepancy is consistent throughout the analysis, so should not affect comparisons between groups. For the survival analysis age at specimen date was used.

All analyses exclude children aged under 5, those with no age or sex, people with no LSOA (no valid home address) or an LSOA outside England and cases or deaths that could not be linked to HES.

Methods

Descriptive outputs

A range of breakdowns of cases and deaths were produced prior to the standardised ratios, and are provided in the data pack which accompanies this report.

Standardised diagnosis ratios

The aim of the descriptive analysis was to compare the number of COVID-19 diagnoses by ethnic group among people with similar pre-existing conditions. As population registers for people with similar pre-existing health conditions or access to primary care data for the whole population were not available, the population with pre-existing conditions was estimated by identifying all people in the general population who had had a hospital admission in the previous five years with any of the pre-existing conditions mentioned.

Age standardised COVID-19 diagnosis ratios among people with pre-existing health conditions were calculated to take account of the age distributions of different ethnic groups when comparing outcomes. The ratios identify differences between ethnic groups in the risk of getting COVID-19 among people in the general population with specified pre-existing conditions. They were calculated separately for the population with each pre-existing condition, for example the population with diabetes, as follows:

- Age-specific COVID-19 diagnosis rates (in 5-year age bands from 5 to 9 up to 90+) were calculated for the 'reference population' (all people with the relevant health condition for all ethnic groups combined). The numerators were the numbers of diagnoses (confirmed cases of COVID-19) and the denominators were the numbers of people in the population group with the pre-existing condition as defined above.
- These 'reference rates' were applied to the 'study population' (people with the relevant health condition in the specific ethnic group), by age, to calculate the number of expected COVID-19 cases in that ethnic group.
- The observed number of COVID-19 cases in the study population (ethnic group with the pre-existing condition) was then divided by the expected number to give the standardised diagnosis ratio.
- Standardised ratios for diagnoses were calculated for males, females and all persons
- Confidence intervals for the standardised ratios were calculated using Byar's method.[2]

A ratio greater than 1 indicates that there were more cases in the study population (in a particular ethnic group with a particular pre-existing condition) than would be expected if

[2] https://fingertips.phe.org.uk/documents/PHDS Guidance - Confidence Intervals.pdf

they had the same age-specific COVID-19 diagnosis rates as the reference population (all people with that pre-existing condition).

In this study, the ratios were used to compare each ethnic group in turn with the reference (all ethnic group) population. This was done for the long list of all relevant health conditions, and then for subgroups with specific conditions recorded.

For example, amongst all people with cardiovascular disease (that is, having had an admission for cardiovascular disease), each ethnic group in turn was compared with the all ethnic group reference population to assess whether they experienced higher or lower numbers of cases than expected. Each analysis is separate, so comparisons should not be made between different specific ethnic groups or between different health conditions.

Standardised mortality ratios

Age standardised mortality ratios (SMRs) were calculated in exactly the same way as the standardised diagnosis ratios. The same population with a pre-existing condition was used (people in the general population who had had an admission in the previous five years with any of the pre-existing conditions mentioned). The numerators were the numbers of deaths amongst the defined population with a pre-existing condition instead of the numbers of COVID-19 cases.

SMRs were calculated for each of the following breakdowns:

- COVID deaths (COVID-19 mentioned on the death certificate), non-COVID deaths (where COVID-19 was not mentioned on the death certicate) and all deaths
- Males, females and all persons

Baseline period comparison

For SMRs, rates were calculated for 2020 as set out above, and a comparable analysis carried out for a baseline period of 2015 to 2019. This was to establish whether ethnic inequalities were specific to COVID-19 deaths or present also in deaths from other causes in previous years.

For each of the baseline years (2015, 2016, 2017, 2018 and 2019), the populations with health conditions and calculations were derived in exactly the same way as for 2020 (for example, the 2015 analysis uses admissions from 1 March 2010 to 28 February 2015, and so on), and then the populations and deaths were each summed to create aggregates across the whole baseline period.

Analysis of survival rates

Two analytical approaches were investigated:

1. Kaplan Meier curves and Cox proportional hazards models
2. logistic regression models

The first method focused on the length of survival from the date of diagnosis to the date of death whereas the second defined death/survival in binary terms. Given the short timescales covered by the study, the binary outcome of death vs survival is more relevant than the question of whether someone survives 2 days after diagnosis, 10 days, 30 days or longer, within the context of this pandemic. Both methods are limited by the lack of follow-up time after diagnosis, but this is potentially a more significant factor in the logistic regression approach; a Cox proportional hazards model can make use of more of the information available.

Kaplan Meier curves and Cox proportional hazards models were investigated initially. The proportional hazards assumption (that the survival ratios are independent of the length of time between diagnosis and death) was tested using Schoenfeld residuals and this assumption was found to be violated for ethnicity, region, pillar, days since the pandemic began and diabetes. The complexity of the time varying component of these covariates meant that it was not possible to pursue this approach further at this time.

The logistic regression approach was also investigated, with a binary outcome of death within 28 days of specimen date. To check the robustness of using a cut-off of 28 days for death, and identify how the effect of ethnicity varied on different lengths of follow up, sensitivity analysis was carried out by varying the number of days used to follow up.

The dataset was structured for the analysis of death within 28 days of specimen date. Specimen dates were restricted to those taken between 20 March 2020, when a reliable data feed for mortality data became available, and 13 July 2020, to allow for a 28 day follow up. Death records were used to identify if a death had occurred with COVID-19 mentioned on the death certificate within 28 days of the specimen date for the positive test. The final analysis dataset consisted of of 221,276 cases and 33,493 deaths.

Logistic regression models were used to model the odds of death within 28 days of specimen date among positive cases of COVID-19 by age, sex, ethnic group, region, deprivation, pillar of testing, days since the pandemic began (this reflects changes in testing levels and potentially other policy-related factors). To determine the extent to which the relationship between ethnic group and mortality could be explained by pre-existing health conditions, these models were then extended to include the conditions using a forward stepwise regression with likelihood ratio tests. The final models included all sociodemographic variables and flags for cardiovascular disease (CVD), diabetes, chronic kidney disease (CKD), chronic respiratory disease (CRD) and dementia.

Hypertension and cerebrovascular flags were excluded based on collinearity (they are subsets of CVD). Pre-existing health conditions were each included separately in preference to including the single long list variable.

Interactions between age and ethnic group were assessed (adjusting for all other variables in the model). Age modified the effect of ethnic group on mortality, that is relative survival rates between the ethnic groups were not the same in all age groups. A likelihood ratio test on the interaction showed that the interaction was extremely significant.

The interaction between sex and ethnic group was also tested, but was not significant. The odds of death for men were consistently around 1.75 times those for women, across all ethnic groups.

There was a very clear relationship between deprivation (IMD quintile) and odds of death: adjusting for all other variables in the model, compared with the most deprived areas, those in the least deprived quintile had 0.87 times the odds of death, with a gradient between the 2 extremes.

To check the extent to which deprivation confounds the relationship between ethnicity and survival, both the a priori and final models were repeated with deprivation removed. For all the ethnic minority groups, the odds ratios were slightly higher when deprivation is removed, demonstrating that a small amount of the lower survival in the Black and Asian ethnic groups was due to higher levels of deprivation, but only a small amount: in the final model, removing deprivation increases the odds ratios from 1.21 to 1.23 for the Asian group and from 1.12 to 1.15 for the Black group.

Sensitivity analysis was carried out to determine the effect of varying length of follow up, in order to check whether the 28 day cut off is affecting different ethnic groups in different ways. These included death within 10, 20 days and 60 days of specimen date. The 28 day analysis was repeated using the same subset of data available for 60 day follow-up, to enable comparisons to be made.

Data for the 2 pillars of testing were analysed together, with testing pillar being included in the model. Although the 2 pillars constitute generally quite distinct groups of COVID-19 patients, there are 2 reasons why the analysis of survival could not be presented separately:

- The data are becoming less easy to disaggregate – the lines between the pillars are becoming blurred.
- Excluding pillar 2 from the analysis has barely any effect on the results, and the data for pillar 2 did not support separate analysis: only 1,149 of the deaths came from pillar 2 cases (compared with 32,344 for pillar 1). This does not give enough power for the model to take into account the range of factors that were identified as

contributing to variations in survival. Although the major effects of age and sex are clear, the model has insufficient power to detect differences between ethnic groups, a gradient with deprivation or effects of other pre-existing health conditions (with the exception of dementia). Variation between regions in pillar 2 survival would be difficult to interpret, as it may be affected by inconsistency in testing rates.

Literature Review

A structured literature review was completed to address the question 'What is the relationship between pre-existing health conditions, ethnicity and risk of diagnosis and/or mortality from COVID-19?'

The review was limited to UK publications in 2020 only. Searches were carried out in Embase and OVID on 1 October 2020.

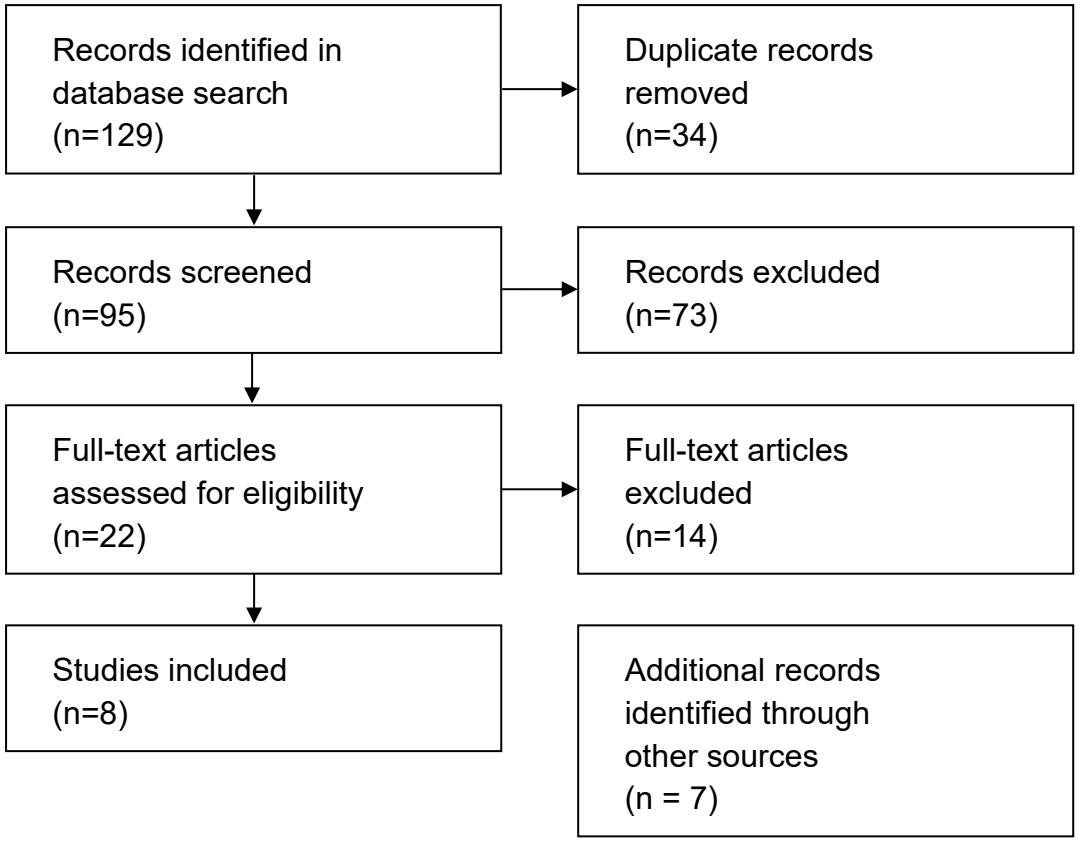
Summary of resources searched and results

Source	Number of results before deduplication	Number of results after deduplication
Embase 1996 to 2020 Week 40	67	67
Ovid MEDLINE(R) and Epub Ahead of Print, In-Process and Other Non-Indexed Citations, Daily and Versions(R) 1946 to 1 October 2020	62	28
TOTAL	129	95

In addition to the search, 6 additional papers were identified by topic experts, some of which were published after 1 October 2020.

The search yielded 95 papers, which were screened by one reviewer by title and abstract. The reviewer then accessed 22 papers in full text and excluded a further 14. No assessment of quality was made, though they are all from recognised academic journals.

Chart 1: Results of literature search



The 8 papers from the literature review are detailed below:

Chadeau-Hyam, M., et al. (2020). "Risk factors for positive and negative COVID-19 tests: a cautious and in-depth analysis of UK biobank data." International Journal of Epidemiology

de Lusignan, S., et al. (2020). "Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study." The Lancet Infectious Diseases 20(9):1034-1042

Hippisley-Cox, J., et al. (2020). "Risk of severe COVID-19 disease with ACE inhibitors and angiotensin receptor blockers: cohort study including 8.3 million people." Heart (British Cardiac Society) 106(19):1503-1511

McQueenie, R., et al. (2020). "Multimorbidity, polypharmacy, and COVID-19 infection within the UK Biobank cohort." PloS one 15(8):e0238091

Patel, A. P., et al. (2020). "Race, socioeconomic deprivation, and hospitalization for COVID-19 in English participants of a national biobank." International Journal for Equity in Health 19(1):115

Perez-Guzman, P. N., et al. (2020). "Clinical characteristics and predictors of outcomes of hospitalized patients with COVID-19 in a multi-ethnic London NHS Trust: a retrospective cohort study." Clinical infectious diseases: an official publication of the Infectious Diseases Society of America

Raisi-Estabragh, Z., et al. (2020). "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." Journal of public health (Oxford, England) 42(3):451-460

Sapey, E., et al. (2020). "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 Respiratory Research 7(1):000644

The following 7 papers were identified by experts in the field:

Clift, A.K., Coupland, C.A.C., Keogh, R.H., Diaz-Ordaz, K., Williamson, E., Harrison, E.M., et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. BMJ. 2020;371:m3731.

Harrison, Ewen M., et al. Ethnicity and Outcomes from COVID-19: The ISARIC CCP-UK Prospective Observational Cohort Study of Hospitalised Patients (5/31/2020) https://ssrn.com/abstract=3618215 or http://dx.doi.org/10.2139/ssrn.3618215

Mathur, R., et al. 2020. Ethnic differences in COVID-19 infection, hospitalisation, and mortality: an OpenSAFELY analysis of 17 million adults in England. DOI. 10.1101/2020.09.22.20198754 https://www.medrxiv.org/content/medrxiv/early/2020/09/23/2020.09.22.20198754.full.pdf

Niedzwiedz C.L., O'Donnell C.A., Jani B.D., Demou E., Ho F.K., Celis-Morales C., et al. 2020. Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. BMC Med. 2020 29;18(1):160 https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-01640-8

Office for National Statistics. 2020. Updating ethnic contrasts in deaths involving the coronavirus (COVID-19), England and Wales: deaths occurring 2 March to 28 July 2020 https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/updatingethniccontrastsindeathsinvolvingthecoronaviruscovid19englandandwales/deathsoccurring2marchto28july2020

Williamson, E., Walker, A.J., Bhaskaran, K.J., Bacon, S., Bates, C., Morton, C.E., et al. 2020. OpenSAFELY: factors associated with COVID-19-related hospital death in the

linked electronic health records of 17 million adult NHS patients. DOI: 10.1101/2020.05.06.20092999 (pre-print; not peer-reviewed)

Williamson, E.J., Walker, A.J., Bhaskaran, K. et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 584, 430–436 (2020) https://doi.org/10.1038/s41586-020-2521-4

Appendix: ICD-10 codes included in list of all relevant conditions

The conditions are listed here by chapter or section of ICD-10. Not all conditions in each chapter are included – the codes give the precise specification.

Human immunodeficiency virus [HIV] disease	B20 to B24
Neoplasms	C00 to C85, C88 to D48
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	D56.1, D57, D595, D61, D70 to D73, D76, D80 to D84, D86, D89
Diabetes	E10 to E14
Obesity	E66
Cystic fibrosis	E84
Dementia	F00 to F03
Diseases of the nervous system	G10, G12.2, G20, G30, G31.0, G31.8, G35, G40, G45, G46, G80, G59.0, G63.2, G73.0, G99.0
Diseases of the circulatory system	I00 to I99
Diseases of the respiratory system	J31, J40 to J47, J61, J62.0, J63.0, J63.1, J63.3 to J63.5, J66, J67.0, J67.1, J67.8, J68.4, J68.8, J69.8, J70.1, J70.3, J80 to J84, J98.3, J99.1
Diseases of the liver and intestinal malabsorption	K70 to K77, K90
Diseases of the musculoskeletal system and connective tissue	M05, M06.0, M06.2, M06.3, M06.8, M06.9, M31.2, M31.3
Diseases of the genitourinary system	N00 to N05, N07, N08, N11, N12, N14 to N16, N18, N19, N25
Diabetes mellitus in pregnancy	O24
Certain conditions originating in the perinatal period	P70.0 to P70.2, P78.8
Congenital malformations, deformations and chromosomal abnormalities	Q20 to Q30, Q32 to Q37, Q44, Q60, Q61
Failure and rejection of transplanted organs and tissues	T86

Radiotherapy session, chemotherapy session for neoplasm or transplanted organ and tissue status	Z51.0, Z51.1, Z94

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. We do this through world-class science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. We are an executive agency of the Department of Health, and are a distinct delivery organisation with operational autonomy to advise and support government, local authorities and the NHS in a professionally independent manner.

Public Health England
Wellington House
133-155 Waterloo Road
London SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: @PHE_uk
Facebook: www.facebook.com/PublicHealthEngland

Published: December 2020
PHE gateway number: GW-1740

PHE supports the UN
Sustainable Development Goals