Outputs by ethnic group in PHE's COVID-19 Health Inequalities Monitoring for England (CHIME) tool.

Assigning ethnicity to records of deaths, hospital admissions, and confirmed COVID-19 cases

Introduction

Early in the first wave of the COVID-19 pandemic in England, there was great demand for data on COVID-19 cases and deaths by ethnic group. However, ethnicity is not collected at death registration and there were gaps in the ethnicity data available for confirmed COVID-19 cases. A process was therefore needed to assign ethnicity from NHS Digital Hospital Episodes Statistics (HES).

Since different ethnicities may be recorded in different treatment episodes, the method selected a single ethnic group from a patient's HES records and linked it with the COVID-19 case or death record. This method was used to provide data by ethnic group in PHE's report on [Disparities in the risks and outcomes of COVID-19](#).

During the pandemic, it became evident that this original method of assigning ethnicity had some limitations, in particular that it overestimated the number of people in the 'Other' ethnic group.

Alternative methods of assigning ethnicity from HES were therefore investigated and were discussed with stakeholders in PHE, as well as external stakeholders from the Office for National Statistics, the Race Disparity Unit, NHS Digital, The King's Fund and the Institute of Health Equity.

A new method has now been agreed and has been used for the data provided by ethnic group in PHE's [COVID-19 Health Inequalities Monitoring for England (CHIME) tool](#). This document presents the new method for assigning ethnicity used within the CHIME tool. It also notes the source of the population estimates used for calculating rates for hospital admissions, deaths and confirmed COVID-19 cases by ethnic group in the tool.

New PHE method for assigning ethnicity from Hospital Episode Statistics (for deaths and hospital admissions indicators)

The original method used by PHE assigned the **most recent** usable ethnic code for an individual available in HES data.

The new method, is based on the NHS Digital HES ethnicity index with a few modifications.

This method:

- Uses the **most frequent** ethnicity recorded across the three HES data sets (APC from 2003/04 onwards, AE from 2007/08 onwards and OP from 2003/04 onwards) excluding any unknown values. OP data were not used from 2006/07 through to 2009/10 as, due to a technical issue, no ethnic code entries were recorded in those years. APC data is restricted to 2003/04 onwards, as the quality and completeness of admitted patient care data was lower before then.
- If there are multiple ethnicities in the data sets with the same frequency, the most recent is chosen.

- If there are multiple ethnicities with the same frequency and latest date, precedence is given to the most recent value from the APC data set as it is considered more robust, followed by the AE data set, followed by the OP data set. Checks completed by NHS Digital indicate completeness in the AE data set is better than the OP data set.
- If there are multiple ethnicities with the same frequency, latest date and source of data we select the ethnicity that occurs more frequently in the general population of England and Wales, according to the 2011 Census (See Appendix A). Incidences of this are very small, and this step was introduced in order to automate the process and to receive the exact same result each time the analysis is completed.
- A value of ethnicity unknown will only be present if there are no known ethnicities in any of the HES data sets.
- To take into the account the overrepresentation of the Other ethnic group, if the most common ethnic group assigned by the method above is Other:
  - The second most common usable ethnic group is assigned instead
  - If there are no other usable ethnic groups, the person is assigned to the Other ethnic group. A person will only be assigned to the Other ethnic group if there are no other usable ethnic groups.

To note, it is perfectly valid for patients to decide to not state their ethnicity when this information is collected in hospital data. People may also decide to state their ethnicity on some occasions but not others. The original and new methods used for assigning ethnicity do not select 'Not Stated' records if there are alternative ethnic codes available. Only those who do not have a usable ethnic code and have repeatedly not stated their ethnicity will have the ethnicity 'Not Stated' recorded.

Impact of the change in method (for deaths and hospital admissions indicators)

The biggest impact of the change in method has been on the Other ethnic group. In PHE's report on [Disparities in the risks and outcomes of COVID-19](#), the highest mortality rates for deaths involving COVID-19 in the first wave of the pandemic were, by some margin, in the Other group (see Appendix B).

That is not the case for the mortality rates presented in the CHIME tool. Across the pandemic period to date, the cumulative mortality rates (and hospital admission rates) using the new method of ethnicity assignment were highest for the Black and Asian groups.

Method for assigning ethnicity (for confirmed COVID-19 cases)

Indicators looking at confirmed COVID-19 cases within the CHIME tool primarily use the ethnicity recorded during pillar 2 testing (swab testing for the wider population as part of the UK government testing programme). If the ethnicity collected is a usable ethnic code, this will be the ethnicity used for the CHIME indicators.

However, for any confirmed cases that do not have a usable ethnic code (i.e. 'Prefer not to say' or 'NULL') and for all pillar 1 cases (swab testing in PHE labs and NHS hospitals for those with a clinical need, and health and care workers), the ethnicity used will be determined for the individual through the use of the new PHE method for assigning ethnicity from Hospital Episode Statistics, detailed above.

As with deaths and hospital admissions, the biggest impact of adopting this method of assigning ethnicity to confirmed cases, has been a reduction in the percentage of cases assigned to the Other ethnic group.

Appendix C shows how ethnicity coding in pillar 2 cases has improved throughout the pandemic.

Broad ethnic groups

Because of small populations for some ethnic groups in some regions, rates for hospital admissions, deaths and confirmed cases are only presented for detailed ethnic groups within England as a whole in the CHIME tool.

Rates are also presented for broad ethnic groups within England and within regions (but the rates for regions are cumulative and are not presented by month).

The broad Black / Black British ethnic group is made up of the detailed groups: Black African, Black Caribbean, and any other Black background.

The broad Asian / Asian British ethnic group is made up of the detailed groups: Bangladeshi, Chinese, Indian, Pakistani, and any other Asian background.

Population estimates for ethnic groups

In PHE's report on Disparities in the risks and outcomes of COVID-19 the populations for ethnic groups were from population estimates from the Office for National Statistics.

These estimates were based on population data from 2011 Census which were 'aged-on' to create annual estimates for years up to 2018.

These estimates, however, had some limitations, including the fact that they did not take into account the effect of international migration since 2011 on the ethnic distribution of the population. ONS noted that this was likely to lead to some underestimate of the population in some ethnic groups, particularly, the Asian and Other groups, which would increase over time.

In the CHIME tool, an alternative set of populations have therefore been used for the rates of hospital admissions, deaths and confirmed cases by ethnic group.

These are estimates from the ETHPOP projections which were produced for PHE by Philip Rees and Paul Norman in the School of Geography, University of Leeds.[1]

The estimates used were for 2019. Estimates for 2020 and 2021 are planned, and it will be possible to revise the rates presented in the CHIME tool once they are available. The 2019 estimates themselves are also expected to be modified, but these changes are unlikely to significantly alter the rates presented.

---

[1] Philip Rees and Paul Norman (2021) Proposals for Improving Local Population Estimates by Ethnicity in England: Report to Public Health England. April. School of Geography, University of Leeds, Leeds, LS2 9JT, United Kingdom.
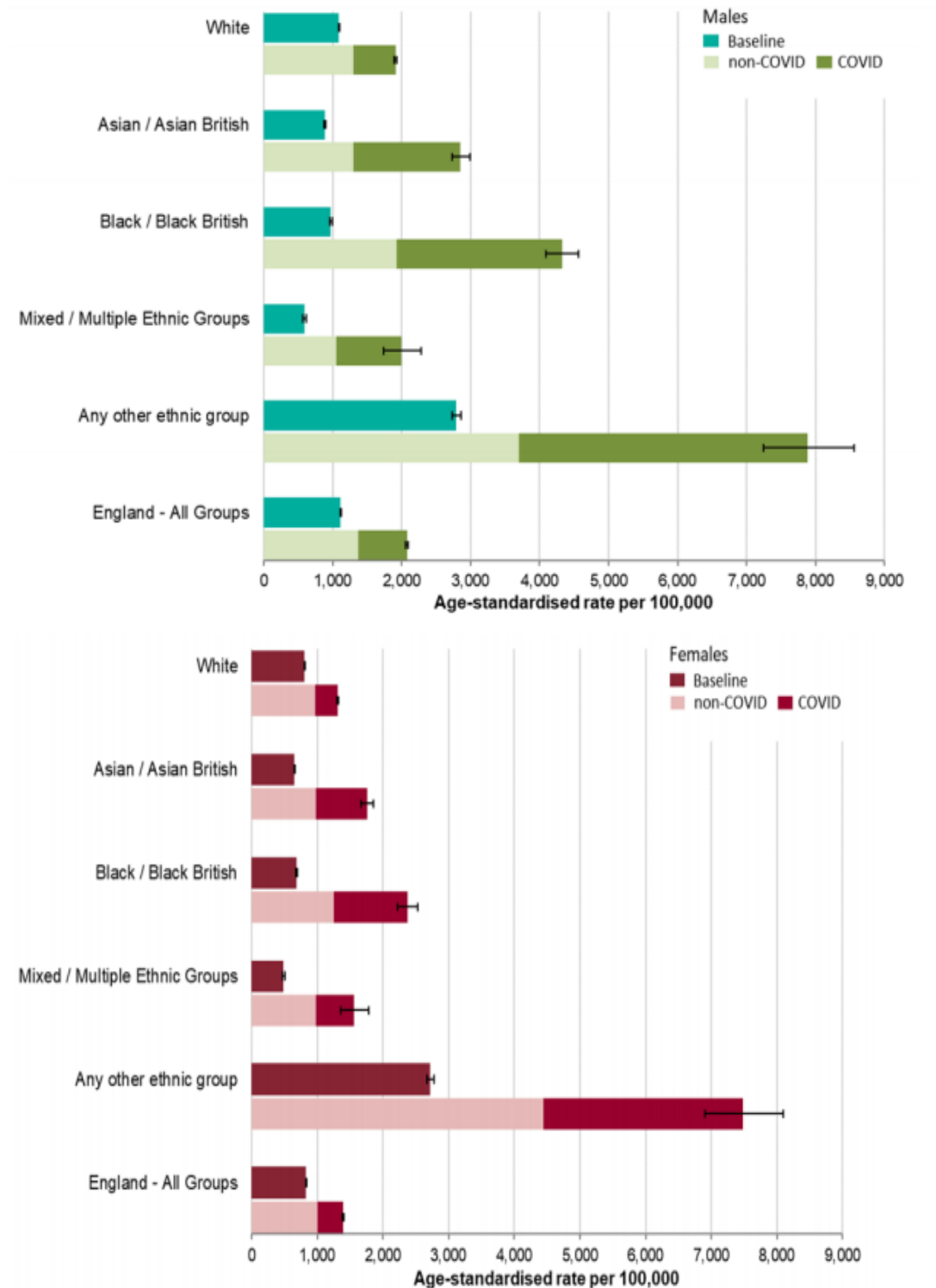
Available from the authors. Contact: p.h.rees@leeds.ac.uk, p.d.norman@leeds.ac.uk.

These populations are modelled rather than being counts of individuals. They are thus presented in the CHIME tool at 1 decimal place, rather than as whole numbers.

Appendix A – Proportions of population in England and Wales, split by ethnic group, Census 2011

Ethnicity	Ethnic code	Percentage	Order
White British	A	80.5%	1
White Other (including Gypsy/Traveller)	C	4.5%	2
Indian	H	2.5%	3
Pakistani	J	2.0%	4
Black African	N	1.8%	5
Asian Other	L	1.5%	6
Black Caribbean	M	1.1%	7
White Irish	B	0.9%	8
Bangladeshi	K	0.8%	9
Mixed White/Black Caribbean	D	0.8%	10
Chinese	R	0.7%	11
Mixed White/Asian	F	0.6%	12
Mixed Other	G	0.5%	13
Black Other	P	0.5%	14
Mixed White/Black African	E	0.3%	15
Other	S	1.0%	16

Appendix B - Age-standardised mortality rates for all cause deaths and deaths mentioning COVID-19, 21 March to 1 May 2020, compared with baseline mortality rates (2014 to 2018), by ethnicity and sex, England

Source: Disparities in the risks and outcomes of COVID-19

Appendix C - Proportion of pillar 2 cases where no ethnicity information was recorded (either 'Prefer not to say' or 'NULL') by year and month

Year	Month	% of cases with no ethnicity information
2020	Mar	97.1%
2020	Apr	100.0%
2020	May	88.4%
2020	Jun	16.1%
2020	Jul	8.5%
2020	Aug	8.9%
2020	Sept	5.9%
2020	Oct	4.7%
2020	Nov	5.7%
2020	Dec	6.3%
2021	Jan	7.3%
2021	Feb	7.1%
2021	Mar	6.6%
2021	Apr	5.6%