Statistical bulletin

Sociodemographic differences in the risk of testing positive for coronavirus (COVID-19), England: 12 September 2020 to 13 December 2021

Comparison of sociodemographic differences in the risk of testing positive for coronavirus (COVID-19) using the COVID-19 Infection Survey (CIS) and routinely collected data from national testing programmes.

Table of contents

1 . Main points

- Different data sources produce different patterns in the estimated relative risk of testing positive for coronavirus (COVID-19) across sociodemographic groups and periods of the pandemic; this may partly be driven by people in different groups being more or less likely to engage with national testing programmes.

- During the pre-Alpha variant period of the coronavirus pandemic (12 September 2020 to 7 December 2020) and the Alpha variant period (8 December 2020 to 17 May 2021), the relative risk of testing positive for COVID-19 among ethnic minority groups, compared with those who identified with the "White: British" group, was largely similar when estimated from the COVID-19 Infection Survey (CIS) and administrative national testing data (NHS Test and Trace); this was after restricting both data sources to registered NHS patients who could be linked to the 2011 Census, and adjusting for age, sex and geographic variables.

- During the Delta variant period (18 May 2021 to 13 December 2021), there were differences in the estimated relative risks across data sources: from the CIS, there was no evidence of differences in the risk of testing positive between ethnic groups; however, when using administrative data, the risk was lower for all ethnic minority groups compared with the "White: British" group.

- In the pre-Alpha variant period, the risk of testing positive was higher for people in more deprived areas than less deprived areas when using the CIS, but there were no clear differences in risk when using administrative data; conversely, in the Delta variant period, there were no clear differences in risk using the CIS, while the risk of testing positive was higher for people in less deprived areas than more deprived areas when using administrative data.

- The risk of testing positive was consistently higher among females than males using administrative data; however, this was not the case in the CIS, where females had lower risk than males in the Delta period.

2 . Differences in the risk of testing positive for COVID-19 by ethnic group using survey and administrative data

This analysis compares estimated relative risks of testing positive for coronavirus (COVID-19) by ethnic group based on the COVID-19 Infection Survey (CIS) and administrative national testing data (NHS Test and Trace). The data source for this analysis is the Office for National Statistics (ONS) Public Health Data Asset (PHDA): 2011 Census data, which was linked to primary care records and national COVID-19 testing data (for more information, see Section 4: Measuring the data). For this analysis, we only included CIS respondents who could be linked to the PHDA.

During the pre-Alpha variant period of the coronavirus pandemic (12 September 2020 to 7 December 2020) and the Alpha variant period (8 December 2020 to 17 May 2021), for the majority of ethnic groups (including "Bangladeshi", "Chinese", "Indian", "Mixed", "Other" and "Pakistani"), the estimated relative risk of testing positive (compared with the "White: British" group) was similar, and similarly significantly different from the "White: British" group, when estimated from the linked CIS-PHDA and PHDA data sources. However, there were differences in the estimates from the two data sources for the "White: Other" and "Black African" groups.

During the Delta variant period (18 May 2021 to 13 December 2021), there was almost no consistency in the estimated relative risks across the data sources. From the linked CIS-PHDA data, the risk of testing positive was not found to be significantly different to the "White: British" group for any of the ethnic groups. Conversely, the risk of testing positive was found to be significantly lower for all ethnic groups compared with the "White: British" group using the PHDA.

This analysis shows that different data sources produce different estimates of patterns in the estimated relative risk of testing positive for COVID-19 across ethnic groups and periods of the coronavirus pandemic, which may partly be driven by people in different groups being more or less likely to engage with national testing programmes. This work aligns with good practice of reviewing the quality of official statistics, which can then inform how future analyses are interpreted.

Results showing the relative risk of testing positive for COVID-19 by sex and area deprivation can be found in our Sociodemographic differences in the risk of testing positive for coronavirus (COVID-19) dataset.

Figure 1: Different data sources produce different estimates of patterns in the relative risk of testing positive for COVID-19 across ethnic groups and periods of the coronavirus pandemic

Adjusted rate ratios for testing positive for COVID-19 by ethnic group, data source and time period, England, 12 September 2020 to 13 December 2021

Notes:

1. "White: British" is the reference group.

2. Estimates are adjusted for age, sex, area deprivation quintile, region, and rural-urban classification.

3. Error bars are 95% confidence intervals.

Download the data

.xlsx

3 . Sociodemographic differences in the risk of testing positive for coronavirus (COVID-19) data

Sociodemographic differences in the risk of testing positive for coronavirus (COVID-19), England
Dataset | Released 31 August 2023
Sociodemographic differences in the risk of testing positive for coronavirus (COVID-19) using the COVID-19
Infection Survey (CIS) and routinely collected data from national testing programmes.

4 . Measuring the data

Data sources

The Office for National Statistics (ONS) Public Health Data Asset (PHDA) is a dataset combining the 2011 Census, death registrations, the General Practice Extraction Service (GPES) Data for Pandemic Planning and Research (GDPPR), and Hospital Episode Statistics (HES), linked via the NHS number. This provides a study population of people in England who were counted in the 2011 Census and present in the GDPPR, which shows that they were registered with a General Practitioner (GP) after November 2019 and had interacted with their GP for a relevant health condition from the year 2000 onwards. At the start of the study period (12 September 2020), this included 40,350,265 people.

The Pillar 2 (NHS Test and Trace) dataset includes results of swab tests for the wider population, as set out in the UK Health Security Agency's (UKHSA) NHS Test and Trace statistics (England) methodology. It excludes tests conducted in UKHSA labs and NHS hospitals for those with a clinical need, and health and care workers. We restricted the Pillar 2 datasets to only positive polymerase chain reaction (PCR) tests. We did not consider positive tests that were less than 120 days following a previous positive test; this removed positive tests that were potentially part of a previous infection episode rather than indicating a new episode.

The Coronavirus (COVID-19) Infection Survey (CIS) is a large longitudinal survey of households randomly sampled from the UK population. The study was commissioned to estimate the number of people in private households infected with COVID-19, with or without symptoms, by taking nose and throat swabs for PCR testing. People living in communal establishments such as care homes, hospitals, prisons and halls of residence were not included. As with the Pillar 2 testing dataset, positive tests less than 120 days following a previous positive test were excluded.

The PHDA study population

For the PHDA study population, sociodemographic variables were defined by information from the PHDA, and the outcome (testing positive for COVID-19) was defined by information from Pillar 2 testing data. Ethnicity and sex information was self-reported on the 2011 Census, and region, rural-urban classification and area deprivation (Index of Multiple Deprivation (IMD) quintile) were derived from Lower layer Super Output Areas (LSOA) in the GDPPR dataset. In instances where LSOA was not available in the GDPPR dataset, the LSOA recorded on the 2011 Census was used instead.

The CIS study population

The CIS sample was restricted to participants who had recorded a test within the study period, which gave a sample size of 302,979 at the beginning of the study period (12 September 2020). For this population, the sociodemographic and outcome variables were all defined by information from the CIS. Ethnicity, sex, area deprivation, region, and rural-urban classification were derived using the same methods as for the PHDA, though ethnicity, sex and LSOA information recorded at the time of survey enrolment were used.

The CIS-PHDA study population

The joined CIS-PHDA study population involved the combination of the restrictions for the PHDA and CIS populations outlined above. Therefore, it was restricted to people who were usual residents and counted in the 2011 Census, could be linked to an NHS Number, and were present in the GDPPR dataset. It was also restricted to CIS participants who recorded a test within the study period and could be linked to the PHDA using the NHS Number.

This resulted in a population of 259,618 people at the start of the study period (12 September 2020). For this population, the sociodemographic characteristics were defined by information from the PHDA, and the outcome was defined by information from the CIS.

We used this study population instead of the unlinked CIS population because it allows for better comparability with the results from the PHDA study population. Differences in results between the unlinked CIS population and the PHDA population could be because of differences in how sociodemographic characteristics were recorded – based on enrolment responses for the CIS population and the 2011 Census for the PHDA population – and sociodemographic differences in the rates of linkage of 2011 Census respondents to NHS numbers. Results based on the unlinked CIS population can be found in our Sociodemographic differences in the risk of testing positive for coronavirus (COVID-19) dataset.

In all three study populations outlined above, the study sample was restricted to people in England who were alive and aged between 10 and 100 years at the beginning of each study period. People who were not living in a private household were not included in the scope of the CIS, so they were also removed from the PHDA and CIS-PHDA populations.

Time periods

For each of the three study populations, the analysis was stratified into three time periods, defined by the main COVID-19 variant in circulation at the time:

- pre-Alpha: 12 September 2020 to 7 December 2020

- Alpha: 8 December 2020 to 17 May 2021

- Delta: 18 May 2021 to 13 December 2021

In previous ONS publications, the pre-Alpha variant period has typically referred to the period from March 2020 to December 2020; however, in this analysis the period was restricted to when national testing was widely adopted.

Statistical methods

We estimated sociodemographic differences in rates of testing positive for COVID-19 (rate ratios) for the CIS, PHDA and CIS-PHDA study populations using Poisson regression models, including the natural logarithm of time at risk as an offset.Time at risk started on the first day of each time period for the PHDA study population, and the first recorded test in each time period for the CIS and CIS-PHDA study populations. For the PHDA study population, time at risk ended at the time of the earliest event, either a first positive test, death, or the end of each time period. For the CIS and CIS-PHDA study populations, time at risk ended at the time of the earliest event, either a first positive test or the last recorded test in each time period.

The models used to estimate the rate ratios included age, sex, ethnicity, area deprivation and rural-urban classification, therefore:

- the rate ratios presented by ethnic group were adjusted for age, sex, region, rural-urban classification and area deprivation

- the rate ratios presented by sex were adjusted for age, ethnicity, region, rural-urban classification and area deprivation

- the rate ratios presented by area deprivation were adjusted for age, sex, ethnicity, region and rural-urban classification

The adjusted rate ratios show the relative risk of testing positive for COVID-19 for:

- different ethnic minority groups compared with the "White: British" group

- different deprivation quintiles compared with the most deprived quintile

- females compared with males

Information that was only available in one data source (such as educational qualifications, which were collected on the 2011 Census but not the CIS) were not considered in this analysis. We did not consider individual-level socioeconomic variables (such as occupation or household composition) or vaccination status because, although these are likely to be related to the risk of testing positive for COVID-19, they may lie on the causal pathway originating from each of the exposures of interest (that is, these variables are partly driven by ethnicity, sex and geography, and in turn they partly determine infection risk).

5 . Related links

Coronavirus (COVID-19) case rates by socio-demographic characteristics, England: 1 September 2020 to 10 December 2021
Bulletin | 14 February 2022
Analysis of age-standardised case rates for coronavirus (COVID-19) in England by socio-demographic characteristics between 1 September 2020 and 10 December 2021.

Coronavirus (COVID-19) Infection Survey, UK: 24 March 2023
Bulletin | 24 March 2023
Percentage of people testing positive for coronavirus (COVID-19) in private residential households in England, Wales, Northern Ireland and Scotland, including regional and age breakdowns.

6 . Cite this statistical bulletin

Office for National Statistics (ONS), released 31 August 2023, ONS website, statistical bulletin, [Sociodemographic differences in the risk of testing positive for coronavirus (COVID-19), England: 12 September 2020 to 13 December 2021](#)