Office for National Statistics

Article

Coronavirus (COVID-19) Infection Survey technical article: Cumulative incidence of the percentage of people who have been infected with COVID-19 by variant and age, England: 9 February 2023

Analysis of the percentage of people in England who have tested positive for COVID-19 using the Coronavirus (COVID-19) Infection Survey. This survey is being delivered in partnership with the University of Oxford, the University of Manchester, the UK Health Security Agency and the Wellcome Trust.

Table of contents

1 . Main points

- The lowest estimates of cumulative incidence across all age groups, showing slow increases in the percentage of people infected over time, occurred during the periods of time before the coronavirus (COVID-19) Alpha variant and since the start of the survey (26 April to 7 December 2020), and when Alpha was most common (8 December 2020 to 17 May 2021).

- An estimated 7.0% (95% credible intervals: 6.9% to 7.2%) of people were infected with COVID-19 in the pre-Alpha period (since the start of the survey) and 8.1% (credible intervals: 7.9% to 8.2%) during the Alpha period.

- The Delta period (18 May 2021 to 13 December 2021) shows a substantially higher percentage of people infected over time at 24.2% (credible intervals: 23.9% to 24.5%) than earlier periods, especially among younger age groups.

- In the Delta period, 55.1% (credible intervals: 53.5% to 55.8%) of those aged 12 to 16 years were infected with COVID-19, in contrast with only 8.2% (credible intervals: 7.9% to 8.4%) of those aged 70 years and over.

- The three Omicron periods, BA.1 (14 December 2021 to 21 February 2022), BA.2 (22 February 2022 to 6 June 2022), and BA.4/BA.5 (6 June up to the final data-point on 11 November 2022), generally show more rapid increases in cumulative incidence than observed for previous variants.

- An estimated 33.6% (credible intervals: 33.1% to 34.0%) of people were infected with COVID-19 in the BA.1 period, 43.6% (credible intervals: 43.1% to 44.1%) in the BA.2 period, and 46.5% (credible intervals: 45.9% to 47.1%) in the BA.4/BA.5 periods.

This analysis defines periods of time by the variant of COVID-19 that was most common in England. Periods are by dominance of a variant and are not the same duration. Total number of people infected (cumulative incidence) with COVID-19 is estimated separately in each period and may include people who have been previously infected.

2 . Overview of COVID-19 variants

This technical article presents data from the Coronavirus (COVID-19) Infection Survey (CIS), from its start on 26 April 2020, until 11 November 2022. The article presents modelled estimates of the total number of people who have been infected with coronavirus (COVID-19) during different periods of time, when different COVID-19 variants were most common (cumulative incidence) as follows:

- pre-Alpha: 26 April 2020 to 7 December 2020

- Alpha: 8 December 2020 to 17 May 2021

- Delta: 18 May 2021 to 13 December 2021

- BA.1: 14 December 2021 to 21 February 2022

- BA.2: 22 February 2022 to 6 June 2022

- BA.4/5: 7 June 2022 to 11 November 2022

The period before Alpha started prior to 26 April 2020. In this analysis we do not account for infections prior to the start of the survey. Therefore we underestimate the true cumulative incidence in the pre-Alpha period.

From 11 November 2022, a range of variants have been circulating including BA.2.75 and its sub-lineages (including CH.1.1 and XBB) as well as the BA.5 sub-lineage BQ.1. Therefore we do not provide estimates for this period.

Our April 2022 analysis suggested that approximately 70.7% of people of all ages had been infected with COVID-19 between April 2020 and February 2022. This new analysis provides information about how infections are broken down by age and period (variant) across the coronavirus pandemic.

The sample includes people in England (CIS participants), who had one or more nose and throat swabs to test for COVID-19. Each participant was regularly tested during their time in the survey. The swabs were tested using polymerase chain reaction (PCR). We use COVID-19 infections to mean testing positive for SARS-CoV-2, the coronavirus causing COVID-19, using a PCR test.

The people included in the survey were aged two years and over and were living in private households. Those in hospitals, care homes or other communal establishments were not included.

We take all positive and negative tests in the survey and apply statistical modelling techniques to estimate the number of people who have had COVID-19 during each period of time. We also present these estimates by age group.

3 . Methods

In epidemiology, daily prevalence is the total number of people with an infection on a given day, while incidence is the number of people newly infected on a given day. In the survey, we estimate both the number of people in the population who would test positive for coronavirus (COVID-19) on a nose and throat swab using a PCR test (positivity) and the number of people who would be newly positive using a PCR test on a nose and throat swab each day (incidence). We do this using both positive and negative swab results.

Positivity refers to the proportion or number of people who would test positive on any given day, if we sampled the whole population. Positivity is not the true number infected on a given day, it is those testing positive on a given day. To calculate the true number of people infected on a given day (prevalence), we would need an accurate understanding of the swab test's sensitivity (true-positive rate) and specificity (true-negative rate). Positivity is therefore likely to underestimate the true prevalence.

To estimate how many people would have had a COVID-19 infection in each period, we need to first estimate the number of people who would test positive on any given day and then aggregate this over each period. We first estimate the daily proportion of the population who would test positive with their first known COVID-19 infection (if they were tested). We then apply our established incidence methodology to provide an estimate of the daily numbers testing positive for the first time in each period, and then aggregate this to estimate the number of people who have ever been test-positive in each period

Positivity (prevalence)

In order to calculate daily positivity, we use an Integrated Nested Laplace Approximation (INLA) model, estimated at regional level within England and weighted to national level. The model is fitted with an interaction between time and sex and an interaction between time and age. The models are post ranked by age and sex to reflect the population living in each region.

This is a different model to our regular [positivity estimates](). This is because our regular positivity estimates use simpler models over a short period (six weeks). It is reasonable to assume the variation across region and time is approximately constant over these shorter time periods, however this is not the case over longer periods. Figure 1 shows the positivity estimate over time split by different age groups used in this analysis.

Figure 1: Positivity by age group over time

Download the data

[.xlsx]()

Incidence

To obtain the daily incidence of new PCR-positive infection episodes, we require an estimate of how long a person with a COVID-19 infection will test positive for on a PCR test. Using survey data, we can estimate the time between a person first testing positive and when they would first test negative again. This duration varies from person to person, so we estimate and allow the duration distribution to vary over the course of the coronavirus pandemic.

We combine the estimates of positivity and duration to obtain daily incidence. In general, incidence and duration can be used to calculate positivity. The reverse process of estimating incidence is called "deconvolution."

Positivity on any given day is the sum of those first testing positive on previous days and those who still test positive on that particular day. A single linear equation relates prior (unknown) daily incidence, and corresponding (known) durations, to each day's (known) positivity. Combining multiple days of positivity gives a system of linear equations that can be solved mathematically to estimate the unknown daily incidences. The daily incidences are cumulated by period to give the estimated number of people who have tested positive over time within that period.

4 . Cumulative incidence by period (variant) and age

The cumulative incidence analysis by period produces estimates of the percentage of people who have been infected with coronavirus (COVID-19) at least once during each period. These estimates are also presented by age.

It is not appropriate to add these together to estimate total cumulative incidence, as some people will have been infected more than once, causing totals to exceed 100%. The duration of the periods is not the same but reflect the time in which particular variants were most common.

The estimates for each period for all ages are:

- 7.0% (95% credible intervals: 6.9% to 7.2%) of people infected with COVID-19 from 26 April 2020 in the pre-Alpha period

- 8.1% (credible intervals: 7.9% to 8.2%) in the Alpha period

- 24.2% (credible intervals: 23.9% to 24.5%) in the Delta period

- 33.6% (credible intervals: 33.1% to 34.0%) in the BA.1 period

- 43.6% (credible intervals: 43.1% to 44.1%) in the in BA.2 period

- 46.5% (credible intervals: 45.9% to 47.1%) in the BA.4/BA.5 period

For each age group, the highest estimates in any period are:

- 59.5% (credible intervals: 58.0% to 61.6%) of those aged 2 to 11 years old in the BA.1 period

- 55.1% (credible intervals: 53.5% to 55.8%) of those aged 12 to 16 years old in the Delta period

- 46.0% (credible intervals: 44.4% to 50.0%) of those aged 17 to 24 years old in the BA.4/BA.5 period

- 48.7% (credible intervals: 47.4% to 50.5%) of those aged 25 to 34 years old in the in BA.2 period

- 49.3% (credible intervals: 48.4% to 50.5%) of those aged 35 to 49 years old in the BA.4/BA.5 period

- 54.5% (credible intervals: 53.8% to 55.4%) of those aged 50 to 69 years old in the BA.4/BA.5 period

- 48.8% (credible intervals: 48.0% to 49.7%) of those aged 70 years and over in the BA.4/BA.5 period

There are many factors that will determine the cumulative incidence estimated by period and age. These include the biology of the variant, the levels of immunity, and mixing in the population and the levels of controls in place at any given time.

Figure 2: The percentage of people who have had coronavirus (COVID-19) across all variants and age groups

Download the data

.xlsx

5 . Coronavirus (COVID-19) Infection Survey: cumulative incidence data

Coronavirus (COVID-19) Infection Survey: Cumulative incidence of the percentage of people who have tested positive for COVID-19 infections, England
Dataset | Released 9 February 2023
Estimated percentage of the population in England who have tested positive for COVID-19 during the survey period from the Coronavirus (COVID-19) Infection Survey.

Coronavirus (COVID-19) Infection Survey: Positivity of COVID-19 overtime for COVID-19 infections, England
Dataset | Released 09 February 2023
Estimated percentage of the population in England who have tested positive for COVID-19 during the survey period from the Coronavirus (COVID-19) Infection Survey.

6 . Glossary

Credible interval

A credible interval gives an indication of the uncertainty of an estimate from data analysis. The 95% credible intervals are calculated so that there is a 95% probability of the true value lying in the interval.

Cumulative incidence

The percentage of individuals experiencing the outcome of interest over a specific time period. In this case, the percentage of individuals testing positive using a PCR test for coronavirus (COVID-19) over a specific time period.

Uncertainty in the test (false-positives, false-negatives)

These results are directly from the test, and no test is perfect. There will be false positives and false negatives from the test, and false negatives could also come from the fact that participants in this study are self-swabbing. More information about the potential effect of false positives and false negatives is provided in Section 5 of our Coronavirus (COVID-19) Infection Survey: methods and further information methodology.

Period of time dominated by each specific variant of COVID-19

The periods of time reflect the time in which particular variants were most common. In those tests with high viral load (CT<30) we look for 50% of the tests to change from showing S-gene positive to S-gene negative or vice versa

7 . Data sources and quality

Limitations

The cumulative incidence estimates the number of PCR-positive infections during the period, expressed as a percentage of the respective population. Because an individual may have been infected more than once, even within a period, these estimates will tend to be higher than the percentage of people who have ever had coronavirus (COVID-19) in the period.

Re-infections with different variants occur significantly more commonly than reinfections with the same variant. This means adding up the percentages across the different periods (COVID-19 variants) would give a misleading estimate of total percentage of the people who have ever been infected. To accurately estimate the percentages of people who have ever had COVID-19 across all periods, we need an accurate estimate of the number of re-infections over time. This is difficult to measure and estimate.

This analysis is based on periods of time dominated by each specific variant of COVID-19. However, we are not able to sequence every PCR-positive infection. This is a particular problem at the beginning of each period, as we do not know whether a person has been infected with the previous dominant variant, or the new variant. There are also some infections by variants that did not go on to become dominant in any period. This means that we cannot use the terms period and variant interchangeably.

The uncertainty measures (credible intervals) reflect the uncertainty of the positivity estimates. We have not incorporated additional from the calculation of the duration of positivity. This means that the uncertainty in the cumulative incidence is likely to underrepresent the true uncertainty.

8 . Collaboration

This Coronavirus (COVID-19) Infection Survey (CIS) analysis was produced by the Office for National Statistics (ONS) in collaboration with our research partners at the University of Oxford. Of particular note are:

- Sarah Walker – University of Oxford, Nuffield Department for Medicine: Professor of Medical Statistics and Epidemiology and Study Chief Investigator

- Koen Pouwels – University of Oxford, Health Economics Research Centre, Nuffield Department of Population Health: Senior Researcher in Biostatistics and Health Economics

9 . Related links

Coronavirus (COVID-19) Infection Survey, UK
Bulletin | Released weekly
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.

10 . Cite this article

Office for National Statistics (ONS), released 9 February 2023, ONS website, article, Coronavirus (COVID-19) Infection Survey technical article: Cumulative incidence of the number of people who have been infected with COVID-19 by variant and age, England: 09 February 2023