Article

Coronavirus (COVID-19) Infection Survey technical article: Cumulative incidence of the number of people who have tested positive for COVID-19, UK: 22 April 2022

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

Contact:
Charlotte Gaughan, Leanne Massie, David Braunholtz, Teri Howells, Eleanor Fordham
infection.survey.analysis@ons.gov.uk
+44 1633 560499

Table of contents

1 . Main points

- In England, an estimated 38.5 million people had coronavirus (COVID-19) between 27 April 2020 and 11 February 2022 (90% credible intervals: 36.0 million to 41.2 million), equating to 70.7% of the population (90% credible intervals: 66.0% to 75.6%).

- In Wales, an estimated 1.7 million people had COVID-19 between 30 June 2020 and 11 February 2022 (90% credible intervals: 1.3 million to 2.1 million), equating to 56.0% of the population (90% credible intervals: 44.3% to 69.4%).

- In Northern Ireland, an estimated 1.3 million people had COVID-19 between 27 July 2020 and 11 February 2022 (90% credible intervals: 1.0 million to 1.7 million), equating to 72.2% of the population (90% credible intervals: 56.0% to 90.9%).

- In Scotland, an estimated 2.7 million people had COVID-19 between 22 September 2020 and 11 February 2022 (90% credible intervals: 2.1 million to 3.3 million), equating to 51.5% of the population (90% credible intervals: 40.5% to 63.6%).

2 . Overview

About this article

This technical article presents modelled estimates of the number of people who have had at least one episode of coronavirus (COVID-19) since the start of the UK Coronavirus Infection Survey (CIS) on 27 April 2020 until 11 February 2022.

The sample includes 535,116 people in the UK (CIS participants), who had one or more nose and throat swabs to test for COVID-19. Each participant was regularly tested during the duration of their time in the study. 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 in the UK. The people included in the survey were aged two years and over and were living in private households; those in hospitals, care homes and/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 in the population, in each of the four UK nations for the duration of the survey.

3 . Methods

This analysis provides an estimate of the number of people who have ever had a coronavirus (COVID-19) infection during the time periods covered by our survey. The starts dates for this analysis relate to when the survey started, which is different for each country, and the analysis goes up to 11 February 2022. Therefore, the time periods covered are:

- 27 April 2020 to 11 February 2022 for England

- 30 June 2020 to 11 February 2022 for Wales

- 27 July 2020 to 11 February 2022 for Northern Ireland

- 22 September 2020 to 11 February 2022 for Scotland

In epidemiology, daily prevalence is the 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 on a nose and throat swab (positivity) and the number of people who would be newly positive 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)](#).

To estimate how many people have had at least one COVID-19 infection we need to first estimate the number of people who would test positive for the first time on any given day and then aggregate this over time. 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 becoming test-positive for the first time, and then aggregate this to estimate the number of people who have ever been test-positive.

Positivity (prevalence) from "first infection" episodes

To differentiate between subsequent infections in the same person, and to estimate the length of time a person would test positive for, we need to define an episode of infection. In this analysis, a new episode of infection is defined by:

- a new positive test which occurs 120 days or more after an individual's first positive test in the survey and their most recent prior test result was negative

- or, if 120 days has not passed since their first positive test in the survey, the individual's last positive test has been followed by four consecutive negative tests

Any other positive tests are counted as being within the same infection episode. In this analysis, to estimate the daily proportion of the population who would test positive with their first COVID-19 infection, we reclassify any positive tests after the first infection episode as "negative". Reclassifying positive tests in observed subsequent infection episodes allows us to include only first infections to estimate positivity with a first COVID-19 infection, while retaining in the population people who have had COVID-19 before.

We then apply a general additive mixed model (GAMM) to obtain a smoothed time series of positivity and post-stratify using age and region to obtain a nationally representative estimate. For England, because of computing constraints, two models were run from 27 April 2020 to 11 May 2021 and 12 April 2021 to 11 February 2022. These two models overlap by 30 days from 12 April to 11 May 2021 to allow for the resulting incidence estimates series to be smoothly spliced together.

Incidence methodology

To obtain the daily incidence of new positive infection episodes, we require an estimate of how long a person with a COVID-19 infection will test positive for. Using data from people who have tested positive in the survey 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 and so we estimate and allow for the duration distribution to vary over the course of the coronavirus pandemic.

We combine the estimates of positivity and duration to obtain daily incidence. That is, we transform the "first infection episode" positivity series into daily incidence of "first episodes" of being test-positive. In general, incidence and duration can be used straightforwardly to give prevalence. The reverse process of estimating incidence is called "deconvolution".

Specifically in our case, having reclassified the data so only positive tests in first episodes are counted positive, positivity on any given day is the sum of those first testing positive on previous days who are 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 which can be solved mathematically to give the unknown daily incidences. The daily incidences are cumulated to give the estimated number of people who have ever been test-positive over time.

Owing to computing constraints because of the large amount of data, it was not possible to run one positivity model for England across the whole time period. The positivity series and deconvolution were therefore run twice, and the resulting two incidence series spliced together. The two series were weighted and averaged on each day of the 30 day overlap proportionally to their position in the overlap period. For example, the first day of the overlap period gave the early model 100% weight and late model 0% weight, linearly swapping over time. The overlap of 30 days gives a reliable, smooth join, and the join is positioned at a period of very low positivity to minimise the impact of the splicing.

A smaller number of participants from each of the devolved administrations meant it was possible to run a single model for Wales, Northern Ireland and Scotland separately. As such, splicing was not necessary.

The figures from the different devolved administrations are not directly comparable with each other or the England estimate as they refer to different time periods.

4 . Estimates of cumulative incidence by country

The cumulative incidence analysis produces an estimate of the number of people who have been infected with coronavirus (COVID-19) since the start of the Coronavirus Infection Survey (CIS) to 11 February 2022 for each of the four UK countries. The start dates for this analysis relate to when the survey started, which is different for each country.

Across all four UK countries, the percentage of the population that have had COVID-19 since the start of the survey has increased at varying rates up to February 2022. An estimated:

- 70.7% of the population in England (90% credible intervals: 66.0% to 75.6%) had COVID-19 between 27 April 2020 and 11 February 2022

- 56.0% of the population in Wales (90% credible intervals: 44.3% to 69.4%) had COVID-19 between 30 June 2020 and 11 February 2022

- 72.2% of the population in Northern Ireland (90% credible intervals: 56.0% to 90.9%) had COVID-19 between 27 July 2020 and 11 February 2022

- 51.5% of the population in Scotland (90% credible intervals: 40.5% to 63.6%) had COVID-19 between 22 September 2020 and 11 February 2022

Figure 1: Across all four UK countries, the percentage of the population that have had coronavirus (COVID-19) since the start of the survey has increased at varying rates up to February 2022

Estimated cumulative percentage of the population who have tested positive for COVID-19 during the survey period by country, UK, 27 April 2020 to 11 February 2022

Notes:

1. All results are provisional and subject to revision.

2. These statistics refer to infections occurring in private households, and exclude those in hospitals, care homes and/or other communal establishments.

3. The starts dates for this analysis relate to when the survey started in each country. Therefore, estimates start from 27 April 2020 for England, 30 June 2020 for Wales, 27 July 2020 for Northern Ireland and 22 September 2020 for Scotland.

Download the data

.xlsx

5 . Comparisons with other sources

Our estimates are based on the Coronavirus (COVID-19) Infection Survey (CIS), a nationally representative survey that tests a large sample of people living in private households each month. Because we retest the same people, regardless of whether they have symptoms, we can identify both infections and re-infections and our data includes asymptomatic cases. This allows us to estimate the number of people who have had COVID-19 since the survey began in April 2020.

Our estimates of cumulative incidence are broadly consistent with those modelled by the Medical Research Council (MRC) Biostatistics Unit at Cambridge University, which are based on Office for National Statistics (ONS) data. The Cambridge University modelling estimates that 40.2 million people had COVID-19 in England from 20 February 2020 to 10 February 2022. Their approach is different in how re-infections are treated. Currently (April 2022), they estimate that 30% of all infections are re-infections.

Other data sources provide different totals of people who test positive for COVID-19 over periods of time. However, because they are not based on repeat testing, it is not possible to establish whether a positive test is an initial infection or a re-infection.

The UK Health Security Agency (UKHSA) figures compile the number of people who have tested positive in the last seven days and the total number of people who tested positive during the coronavirus pandemic (from 30 January 2020). The dashboard is based on clinical testing data, including NHS Test and Trace data (using polymerase chain reaction (PCR) test results and results of registered lateral flow devices), which have a number of limitations.

NHS Test and Trace data are primarily a clinical testing service for people in England with symptoms or who have been in contact with known COVID-19 cases. Testing capacity has varied over the course of the pandemic. This means that the number of people testing positive for COVID-19 through NHS Test and Trace data will not be nationally representative and are likely to underrepresent people that were infected but did not experience symptoms. The data were not intended for use in monitoring prevalence or incidence, and we do not know how many of these positive tests are re-infections.

Additionally, clinical testing data are collected differently across England, Wales, Scotland, and Northern Ireland, meaning that total estimates are not directly comparable between countries. The methodology used to produce estimates from the CIS is consistent for all UK countries.

The Real-time Assessment of Community Transmission (REACT) study also collects data about the prevalence of COVID-19 in the general population, but is limited in scope to England. Each month, 150,000 different participants are tested to assess community transmission of COVID-19 in real time. However, since participants are different each month, it is not possible to calculate incidence or numbers of re-infections.

6 . Coronavirus (COVID-19) Infection Survey data

Coronavirus (COVID-19) Infection Survey: Cumulative incidence of the number of people who have tested positive for COVID-19 infections, UK
Dataset | Released 22 April 2022
Estimated percentage of the population in England, Wales, Northern Ireland and Scotland who have tested positive for COVID-19 during the survey period from the Coronavirus (COVID-19) Infection Survey.

7 . Collaboration

This Coronavirus (COVID-19) Infection Survey 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

- Anna Seale - University of Warwick, Warwick Medical School: Professor of Public Health; UK Health Security Agency, Data, Analytics and Surveillance: Scientific Advisor

8 . Glossary

Credible interval

A credible interval gives an indication of the uncertainty of an estimate from data analysis. The 90% credible intervals are calculated so that there is a 90% 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 for coronavirus (COVID-19) over a specific time period.

9 . Data sources and quality

Our Coronavirus (COVID-19) Infection Survey methodology article provides further information around the survey design and how we process data.

More information on the strengths and limitations of the data, data uses and users is available in the Coronavirus Infection Survey QMI and the Coronavirus Infection Survey statistical bulletin.

10 . Limitations

There is no source of data that provides a precise count of the number of people who have been infected with coronavirus (COVID-19). This analysis is based on a method that provides our best current estimate of the number of people who have been infected since the survey started. Our method contains several sources of potential bias, which may have affected our estimates.

The method assumes first infection episodes identified in the Coronavirus Infection Survey (CIS) data are truly the first time a person had COVID-19. While this will have been (nearly) true at the start of the pandemic, as more people have been infected and then re-infected with COVID-19, an increasing proportion of "first infection episodes" will be re-infections where the first infection was not identified in CIS. This means our estimates of first-infection positivity will be inflated (by inclusion of some second or later infections which we were unable to identify as such). Therefore, our current estimates of how many people have had COVID-19 since the start of the survey could be biased upwards. The credible intervals do not allow for this uncertainty.

Our definition of an infection episode assumes that individuals can test positive for up to 120 days within the same infection. Since the Omicron variants became dominant, we and others have observed a high number of re-infections, many shortly after Delta infections, suggesting that this is less appropriate. As we use the definition of an infection episode to estimate the time people will test positive for, changing the definition to allow for re-infections happening much more quickly after a previous infection will reduce the estimated number of days an individual will test positive for. This will increase the estimates of incidence, meaning our current estimates would be biased downwards.

Infections before the CIS survey started in the respective UK countries are not included. This means our current estimates will be biased downwards from the true figure, particularly in Wales, Northern Ireland and Scotland where data collection started later. However, this will not change the shape of the curves noticeably. Estimates of antibody positivity from CIS suggest that around 6% of the England population had COVID-19 before the start of the survey.

The beginning and ends of the time series are subject to less information and instability and subject to greater relative uncertainty which may not be fully captured in the credible intervals. We have removed the last 14 days of the series to attempt to mitigate against some of this. The net effect on the bias is unclear.

We do not know the exact specificity (percentage of true negatives that test negative) and sensitivity (percentage of true positives that test positive) of the polymerase chain reaction (PCR) test. The credible intervals do not adjust for this uncertainty. As this uncertainty is primarily driven by the uncertainty in the sensitivity of the tests, and given the high specificity of the test, true positivity and duration of positivity are both likely to be biased slightly downwards. However, the deconvolution means that the net bias could go in either direction.

More information about test sensitivity and specificity is available in our methodology article.

11 . Future developments

To develop the analysis presented in this article further, we plan to apply our updated definition of an episode of infection to the current method. Our updated definition of an episode has been applied to our re-infections analysis , and reflects the shorter time between re-infections that have occurred during the Omicron variants period, compared with earlier variants.

We are also developing models to estimate the percentage of people who have had coronavirus (COVID-19) during the study period by age group.

12 . Related links

Coronavirus (COVID-19) Infection Survey, UK
Bulletin | Updated weekly
Estimates for England, Wales, Northern Ireland and Scotland. This survey is being delivered in partnership with the University of Oxford, University of Manchester, UK Health Security Agency and Wellcome Trust.

Coronavirus (COVID-19) Infection Survey: characteristics of people testing positive for COVID-19, UK
Bulletin | Updated fortnightly
The characteristics of people testing positive for coronavirus (COVID-19) from the COVID-19 Infection Survey. This survey is being delivered in partnership with the University of Oxford, the University of Manchester, UK Health Security Agency and Wellcome Trust.

COVID-19 Infection Survey: methods and further information
Methodology article | Updated 7 February 2022
Information on the methods used to collect the data, process it, and calculate the statistics produced from the Coronavirus (COVID-19) Infection Survey.

Coronavirus (COVID-19) Infection Survey QMI
Methodology | Released 16 July 2021
Quality and Methodology Information for the Coronavirus (COVID-19) Infection Survey (CIS), detailing the strengths and imitations of the data, methods used, and data and users.

Coronavirus (COVID-19) Infection Survey: antibody data for the UK
Bulletin | Updated fortnightly
Antibody data by UK country and English regions from the Coronavirus (COVID-19) Infection Survey. This survey is being delivered in partnership with the University of Oxford, University of Manchester, UK Health Security Agency and Wellcome Trust.