Coronavirus (COVID-19) Infection Survey QMI

Quality and Methodology Information for the Coronavirus (COVID-19) Infection Survey (CIS), detailing the strengths and limitations of the data, methods used, and data uses and users.

Contact:
Kara Steel
infection.survey.analysis@ons.gov.uk
+44 (0)1633 651689

Release date:
16 July 2021

Next release:
To be announced

Table of contents

1 . Output information

- **Survey name:** Coronavirus (COVID-19) Infection Survey (CIS)
- **Frequency:** Continuous with weekly publications
- **How compiled:** Estimates are derived from a sample survey in which private households are followed up on a weekly basis for five weeks and then monthly thereafter
- **Geographic coverage:** UK
- **Sample size:**
  Swab sample – used for estimating positivity rates and incidence Since October 2020, approximately 150,000 people tested per fortnight in England, 15,000 in Scotland, 9,000 in Wales and 5,000 in Northern Ireland.

  Blood sample – used for estimating presence of antibodies Target is to achieve up to 125,000 people giving blood samples every month in England, and up to 7,500, 5,500 and 12,000 per month in Wales, Northern Ireland, and Scotland respectively through to April 2022. This target increased in February 2021 to allow better monitoring of the effect of vaccinations.
- **Last revised:** 16 July 2021

2 . About this Quality and Methodology Information report

This quality and methodology report contains information on the quality characteristics (including the European Statistical System five dimensions of quality) of the statistics produced based on Coronavirus (COVID-19) Infection Survey (CIS) data, as well as the methods used to create them.

The information in this report will help you to:

- understand the strengths and limitations of the statistics
- learn about existing uses and users of the data
- understand the methods used to create the data
- help you to decide suitable uses for the data
- reduce the risk of misusing data

3 . Important points

The coronavirus (COVID-19) pandemic is having a profound impact across the UK. In response to the pandemic, the Coronavirus (COVID-19) Infection Survey was set up from April 2020 to estimate:

- how many people across England, Wales, Northern Ireland, and Scotland test positive for COVID-19 infection at a given point in time, regardless of whether they report experiencing symptoms
- the average number of new positive test cases per week
- the number of people who test positive for antibodies, to indicate how many people are ever likely to have had the infection or have been vaccinated

Only private residential households and their residents aged 2 years and over are included in the survey. People in hospitals, care homes and/or other institutional settings are not included. In these settings, rates of COVID-19 infection are likely to be different. More information about rates of COVID-19 in care homes can be found in Impact of coronavirus in care homes in England: 26 May to 19 June 2020.

The Office for National Statistics (ONS) is currently working with the University of Oxford, University of Manchester, Public Health England, Wellcome Trust, IQVIA and the Lighthouse Laboratory at Glasgow to run the COVID-19 Infection Survey in the UK.

4 . Quality summary

Overview of the Coronavirus (COVID-19) Infection Survey

The Coronavirus (COVID-19) Infection Survey was launched in England on 26 April 2020 and was expanded to include Wales on 29 June 2020, Northern Ireland on 26 July 2020, and Scotland on 25 September 2020.

Nose and throat swab samples

The COVID-19 Infection Survey is based on a nationally representative survey with a random sample. We ask everyone over the age of 2 years in each household sample to have a nose and throat swab. These are tested for SARS-CoV-2 using reverse transcriptase polymerase chain reaction (RT-PCR). This is an accredited test that is part of the national testing programme. These samples are collected so we can estimate the number of people who are infected.

We need to know more about how the virus is transmitted in individuals who test positive on nose and throat swabs; whether individuals who have had the virus can be reinfected; and about the incidence of new positive tests in individuals who have not been exposed to the virus. To address these questions, we collect data over time. Every participant is swabbed once; they are then invited to have repeat tests every week for another four weeks and then monthly until April 2022.

Blood samples

To capture data about people who have had COVID-19 but have since recovered and to assess the impact of vaccinations, we aim to ask adults aged 16 years or older to also give a sample of blood, using a capillary finger prick method demonstrated by a specially trained fieldworker, which is then undertaken by the participant. The blood samples are taken at enrolment and then every month.

Initially, 20% of adults over 16 years old surveyed within our household sample were asked to provide a blood sample. However, to monitor the impact of vaccination on both immunity and infection, from February 2021, we started asking a representative sample of the other adults recruited to the study to start giving blood samples at their monthly visits.

The target is to achieve up to 125,000 people giving blood samples every month in England, and up to 7,500, 5,500 and 12,000 per month in Wales, Northern Ireland, and Scotland respectively (150,000 in total across the UK) through to April 2022.

Estimating those testing positive and antibodies

We use a number of different techniques to estimate the number of people testing positive for SARS-CoV2 (the virus that causes the coronavirus (COVID-19) disease) and the presence of antibodies, broken down by different characteristics (age, region and so on). The results are adjusted for a number of characteristics such as age, sex and region to be representative of the community population, and to help mitigate possible biases from non-consent and non-response.

Uses and users of the COVID-19 Infection Survey

The UK government, Scottish Government, Welsh Government and Northern Ireland Executive are the main users of the COVID-19 Infection Survey. They use our statistics to track the progress of the COVID-19 pandemic in the UK and to help inform decisions about COVID-19 restrictions and related policies.

Matters such as COVID-19 restrictions and policies related to the COVID-19 pandemic are devolved and the Scottish Government, Welsh Government and Northern Ireland Executive use data from the COVID-19 Infection Survey to inform these decisions.

The Welsh Government, Department of Health (Northern Ireland Executive) and the Scottish Government use the results from the COVID-19 Infection Survey to analyse and describe trends and changes in the COVID-19 pandemic for Wales, Northern Ireland, and Scotland respectively.

The results of the COVID-19 Infection Survey contribute to the Scientific Advisory Group for Emergencies (SAGE) estimates of the rate of transmission of the infection, often referred to as "R". The survey also provides important information about the socio-demographic characteristics of the people and households who have contracted COVID-19. Results are also used to inform policy in government, providing an evidence base for decisions around changes to restrictions, helping with monitoring and surveillance, and with planning for services and vaccinations.

Other users include academics and health researchers, who conduct research and analysis of the COVID-19 pandemic, the characteristics of those testing positive (such as their occupation, work location, travel status and symptoms experienced) and any possible inequalities associated with those.

The media also report widely on the COVID-19 Infection Survey data and the general public are interested in the statistics produced by the COVID-19 Infection Survey to help understand trends in the percentage of people testing positive. The survey is also used by international audiences such as the World Health Organisation (WHO) who use the data to help measure the pandemic globally.

The data can be used for:

- estimating the number and proportion of current positive cases in the community, including cases where people do not report having any symptoms
- identifying differences in numbers of positive cases and changes in them over time between different areas and regions
- estimating the number of new cases and change over time in positive cases
- estimating the number of people testing positive for COVID-19 antibodies

The data cannot be used for:

- measuring the number of cases and infections in care homes, hospitals and/or other institutional settings
- providing information about recovery time of those infected

Strengths and limitations of the COVID-19 Infection Survey

These statistics have been produced rapidly in response to developing world events. The Office for Statistics Regulation, on behalf of the UK Statistics Authority, reviewed them in May 2020 and again in March 2021 against important aspects of the Code of Practice for Statistics and regarded them as consistent with the Code's pillars of trustworthiness, quality and value.

Sample size and selection

One of the survey's main strengths is the fact the survey subjects are a random sample of the population with a large sample size. This means that unlike other sources such as NHS Test and Trace, for example, which includes only people reporting symptoms, the COVID-19 infection survey also picks up those people not reporting symptoms too.

Timeliness

The survey presents very timely estimates either weekly or fortnightly on a range of domains of interest such as the presence of antibodies and symptoms experienced. The survey questionnaire is also adaptable as social factors change like return to work and travel.

Uncertainty

As well as low prevalence rates limiting analysis, the estimates presented in our weekly bulletin and monthly article contain uncertainty, although the statistics produced as outputs from the survey data are our best estimates, they should not be regarded as completely accurately reflecting the unknown true numbers we are trying to estimate. There are many sources of uncertainty including:

- uncertainty in the test (false negatives and false positives can exist)
- in the estimates because we have sampled only a proportion of the population
- potential non-response bias, which may not be fully mitigated by the weighting
- uncertainty in the models used (some models borrow strength and there could be possible incoherence between modelled estimates and point-in-time estimates)
- the quality of data collected in the questionnaire

Information on the main sources of uncertainty are presented in our methodology article and in our Accuracy and Confidence: why we trust the data from the COVID-19 Infection Survey blog.

5 . Quality characteristics of the Coronavirus (COVID-19) Infection Survey

Relevance

(The degree to which statistical outputs meet current and potential user needs.)

The Coronavirus (COVID-19) Infection Survey, looks to estimate the percentage of the population testing positive for COVID-19 and helps track the current extent of infection and transmission of COVID-19 among the community population as a whole.

We use the number of people testing positive for COVID-19 with PCR tests via nose and throat swabs to calculate the proportion of the community population who test positive for the infection at a given point in time (positivity rate), and the number of new infections over a given time period (incidence rate).

We calculate the positivity rate for COVID-19 in England, Wales, Northern Ireland, and Scotland as well as regions of England and, when the positivity rate allows, also for sub-regional geographies for the UK. Note that we report the positivity rate rather than the prevalence rate, as to calculate the latter would need an accurate understanding of the swab test's sensitivity (true-positive rate) and specificity (true-negative rate).

The incidence rate is a measure of only new infections in a given time period. A statistical model calculates the clearance time (the time between the first positive test and last time a participant would have tested positive). New infections are then deduced from these clearance times in England, Wales, Northern Ireland, and Scotland.

We use blood test results to identify individuals who have antibodies against SARS-CoV-2, which helps us to understand who has had COVID-19 in the past and the impact of vaccinations. These blood tests are taken from individuals aged 16 years and over from a randomly selected subsample of households. We also present data on the characteristics of people testing positive for COVID-19 in our fortnightly bulletin. Topics to be included vary depending on user need.

Statistics from the COVID-19 Infection Survey are used to aid government decision-making, providing insights on how the infection is spreading in the community. This helps the government to make informed decisions on important policies such as changes to restrictions and planning for services and vaccination rollout.

Analysis feeding into calculation of the reproduction number, R

The statistics produced from this survey contribute to modelling, which is used to calculate the reproduction number (R) of the virus.

R is the average number of secondary infections produced by one infected person. The Scientific Pandemic Influenza Group on Modelling (SPI-M), a sub-group of the Scientific Advisory Group for Emergencies (SAGE), has built a consensus on the value of R based on expert scientific advice from multiple academic groups.

Accuracy and reliability

(The accuracy of statistical outputs is the degree of closeness between an estimate and the true value that the statistics were intended to measure. Reliability refers to the closeness of the initial estimates value to the subsequent estimates value.)

Uncertainty

The estimates presented in our weekly COVID-19 Infection Survey bulletin and fortnightly characteristics and antibody and vaccination bulletins contain uncertainty. There are many sources of uncertainty, but the main sources in the information presented include each of the following:

- uncertainty in the test (false-positives, false-negatives and timing of the infection)
- data are based on a sample of people rather than the whole population, so there is some statistical uncertainty in the estimates
- uncertainty in the model
- uncertainty in the quality of data collected in the questionnaire and in the swabbing procedure

Results come directly from the laboratory that performs the PCR test, and no test is perfect. There will be false-positive and false-negative results from the tests, and false-negatives could also come from the fact that participants in this study are self-swabbing. More information about the potential impact of false-positives and false-negatives is provided in the Test sensitivity and specificity section.

Any estimate based on a random sample contains some uncertainty. If we were to repeat the whole process many times, we would expect the true value to lie in the 95% confidence interval on 95% of occasions. A wider interval indicates more uncertainty in the estimate.

As in any survey, some data can be incorrect or missing. For example, participants and interviewers sometimes misinterpret questions, record information that is not entirely accurate, or skip them by accident. To minimise the impact of this, we clean the data, editing or removing data that are clearly incorrect. For more information, see our methodology page on statistical uncertainty.

Response rates

Participants selected and invited to take part in the Coronavirus (COVID-19) Infection Survey are not given a specified date by which to respond, and as a result reported response rates will increase as time progresses. Although most responses occur within the first few weeks after invitation letters are sent, they can continue to increase for some time after that.

We have used two approaches to selecting households for the survey: the first was to re-contact named previous respondents from other ONS surveys who agreed to further contact about other research (this now makes up a small subset of the overall sample), and the second is by writing to "the householder" at addresses selected from the AddressBase (a sampling frame).

For more information on the sampling process, see Section 2: Study design: sampling in our methods article. For up-to-date information on our response rates, please see our most recent bulletin.

Response rates for each nation are found in the dataset that accompanies this bulletin. We provide response rates separately for the different sampling phases of the study.

Communicating uncertainty

The data that are modelled are drawn from a sample and so there is uncertainty around the estimates that the model produces, which is based on a number of assumptions. Because a Bayesian regression model was used, we present estimates along with credible intervals. These 95% credible intervals can be interpreted as there being a 95% probability that such intervals will contain the true value being estimated. Again, a wider interval indicates more uncertainty in the estimate.

For our weighted estimates, confidence intervals are provided. These again are calculated so that if we were to repeat the survey many times on the same occasion and in the same conditions, in 95% of these surveys the true population value would be contained within the 95% confidence intervals. Smaller intervals suggest greater certainty in the estimate, whereas wider intervals suggest greater uncertainty in the estimate.

Further information on confidence and credible intervals can be found in Section 13: Confidence intervals and credible intervals in our methods article.

Representativeness

Ensuring a representative sample of the general population is important for producing survey-based estimates broken down by characteristics such as age, sex, region and ethnicity. In the Coronavirus (COVID-19) Infection Survey, this is important because estimates of COVID-19 positivity rates and antibody rates are required to help us understand trends in different population sub-groups and different parts of the country.

The ONS regularly produces information on the representativeness of the survey. Findings show that the swabs sample is representative of both males and females at a UK level and for all the nations of the UK. All age groups are well represented, and the swab sample is representative of all regions and representative of Wales, Scotland, and Northern Ireland in terms of population share. The white group is overrepresented at the UK level and at a UK level, households of three or more are overrepresented, while households of one person or two people are underrepresented.

The following tables provide an example of some of the representative analysis for the swabs sample for the UK for the week of the 15 May 2021. The unweighted response population is the actual number of people taking part in the survey, while the weighted population has been adjusted to be representative of the target population. The calibration step of the weighting ensures coherence for those variables and categories used in the weighting. Hence the "actual proportion" and "weighted proportion" agree for every category of age and sex because they are used in the weighting.

Table 1a: Actual UK population by sex

UK			
Sex	Relevant actual proportion		
	Number of people	Proportion	Cumulative proportion
Female	32,753,171	0.51	0.51
Male	31,190,622	0.49	1.00
Missing			

Table 1b: Response population for the COVID-19 Infection Survey by sex

UK

	Response population					
Sex	Unweighted			Weighted		
	Number of people	Proportion (excluding missing)	Cumulative proportion	Number of people	Proportion (excluding missing)	Cumulative proportion
Female	243,154	0.53	0.53	32,752,616	0.51	0.51
Male	218,754	0.47	1.00	31,910,252	0.49	1.00
Missing	0			0		

Table 1c: Representativeness of the response population compared with the actual population by sex

UK

	Representativeness			
Sex	Absolute difference to actual population		Proportional difference to actual population	
	Unweighted	Weighted	Unweighted	Weighted
Female	0.02	0.00	4%	0%
Male	-0.02	0.00	-4%	0%
Missing				

Table 1d: Actual UK population by age

UK

Age	Relevant actual proportion		
	Number of people	Proportion	Cumulative proportion
2 to 11	8,112,866	0.13	0.13
12 to 16	3,880,522	0.06	0.19
17 to 24	5,886,258	0.09	0.28
25 to 34	8,901,150	0.14	0.41
35 to 49	12,732,174	0.20	0.61
50 to 69	16,315,106	0.25	0.86
70+	8,835,719	0.14	1.00
Missing			

Table 1e: Response population for the COVID-19 Infection Survey by age

UK

	Response population					
Age	Unweighted			Weighted		
	Number of people	Proportion (excluding missing)	Cumulative proportion	Number of people	Proportion (excluding missing)	Cumulative proportion
2 to 11	33,131	0.07	0.07	8,084,260	0.13	0.13
12 to 16	24,812	0.05	0.13	3,890,412	0.06	0.19
17 to 24	28,964	0.06	0.19	5,889,524	0.09	0.28
25 to 34	47,478	0.10	0.29	8,895,499	0.14	0.41
35 to 49	91,121	0.20	0.49	12,729,483	0.20	0.61
50 to 69	153,167	0.33	0.82	16,324,053	0.25	0.86
70+	83,235	0.18	1.00	8,849,638	0.14	1.00
Missing	0			0		

Table 1f: Representativeness of the response population compared with the actual population by age

UK

	Representativeness			
	Absolute difference to actual population		Proportional difference to actual population	
Age	Unweighted	Weighted	Unweighted	Weighted
2 to 11	-0.05	0.00	-43%	0%
12 to 16	-0.01	0.00	-10%	0%
17 to 24	-0.03	0.00	-31%	0%
25 to 34	-0.03	0.00	-25%	0%
35 to 49	0.00	0.00	0%	0%
50 to 69	0.08	0.00	31%	0%
70+	0.04	0.00	32%	0%
Missing				

Table 1g: : Actual UK population by ethnicity

UK

Ethnicity	Relevant actual proportion		
	Number of people	Proportion	Cumulative proportion
White	55,513,974	0.86	0.86
Asian/Asian British	4,791,530	0.07	0.93
Black/African/Caribbean/Black British	2,150,975	0.03	0.97
Mixed/Multiple ethnic groups	1,064,744	0.02	0.98
Other ethnic groups	1,142,570	0.02	1.00
Missing			

Table 1h: Response population for the COVID-19 Infection Survey by ethnicity

UK

Ethnicity	Response population					
	Unweighted			Weighted		
	Number of people	Proportion (excluding missing)	Cumulative proportion	Number of people	Proportion (excluding missing)	Cumulative proportion
White	426,150	0.92	0.92	59,341,180	0.92	0.92
Asian/Asian British	18,425	0.04	0.96	2,649,675	0.04	0.96
Black/African/ Caribbean/ Black British	4,497	0.01	0.97	647,962	0.01	0.97
Mixed/Multiple ethnic groups	8,109	0.02	0.99	1,338,596	0.02	0.99
Other ethnic groups	4,509	0.01	1.00	659,863	0.01	1.00
Missing	218			25,591		

Table 1i: Representativeness of the response population compared with the actual population by ethnicity

UK

Ethnicity	Representativeness			
	Absolute difference to actual population		Proportional difference to actual population	
	Unweighted	Weighted	Unweighted	Weighted
White	0.06	0.06	8%	7%
Asian/Asian British	-0.03	-0.03	-46%	-45%
Black/African/Caribbean/Black British	-0.02	-0.02	-71%	-70%
Mixed/Multiple ethnic groups	0.00	0.00	7%	26%
Other ethnic groups	-0.01	-0.01	-45%	-42%
Missing				

To address the fact that some individuals in the survey will have dropped out and others will not respond to the initial invite and to reduce potential bias, the regression models used to produce our estimates adjust the survey results to be more representative of the overall population in terms of age, sex, region (for England) and ethnicity (for England, Scotland and Wales). For more information see our methods article.

We are also looking at further ways we can improve the representativeness of individuals taking part in the survey. For example, we have a programme of work to look at increasing the representativeness of ethnicity in the sample. This includes strategies such as sending out reminders and using community engagement officers to go into communities with underrepresented ethnic groups to explain why taking part in the survey is important.

Coherence and comparability

(Coherence is the degree to which data that are derived from different sources or methods, but refer to the same topic, are similar. Comparability is the degree to which data can be compared over time and between geographic areas.)

The ONS and its academic partners carry out extensive quality assurance in producing these statistics, from checking data received is in the expected format and statistics produced look plausible, to triangulation with other COVID-19 data sources. These are detailed further in this section along with information on why estimates between the different data sources may differ.

NHS Test and Trace

Each nation of the UK (England, Wales, Northern Ireland and Scotland) has a Test and Trace system. These ensure that anyone who develops symptoms of COVID-19 can quickly be tested to find out if they have the virus. Some nations also include targeted asymptomatic testing of NHS and social care staff and care home residents. Additionally, it helps trace close recent contacts of anyone who tests positive for COVID-19 and, if necessary, notify them that they must self-isolate. We have published an article that compares the methods used in the COVID-19 Infection Survey and NHS Test and Trace in England.

In comparison with Public Health data and Test and Trace data, the statistics presented in our weekly bulletin take a representative sample of the community population (those in private residential households), including people who are not otherwise prioritised for testing. This means that we can estimate the number of people in the community population with COVID-19 who do not report any evidence of symptoms, which is one of the unique features of the Coronavirus (COVID-19) Infection Survey.

Laboratory confirmed cases in the UK

Public Health England (PHE) presents data on the total number of laboratory-confirmed cases in the UK, which capture the cumulative number of people in the UK who have tested positive for COVID-19. These statistics present all known cases of COVID-19, both current and historical, for the UK, and by nation, by regions of England, and because of the large sample size, by local authority. Further information can be found on the Coronavirus Dashboard. A summary for England, Wales, Scotland and Northern Ireland is also available.

Other studies

This study is one of a number of studies that look to provide information around the coronavirus pandemic within the UK.

COVID Symptom Study (ZOE app and King's College London), UK

The COVID Symptom Study app allows users to log their health each day, including whether or not they have symptoms of COVID-19. The study aims to predict which combination of symptoms indicate that someone is likely to test positive for COVID-19. The app was developed by the health science company ZOE with data analysis conducted by King's College London. Anyone over the age of 18 years can download the app and take part in the study. Respondents can report symptoms of children.

The study estimates the total number of people with symptomatic COVID-19 and the daily number of new cases of COVID-19 based on app data and swab tests taken in conjunction with the Department of Health and Social Care (DHSC). The study investigates the "predictive power of symptoms", and so the data do not capture people who are infected with COVID-19 but who do not display symptoms.

Unlike the data presented in the COVID-19 Infection Survey bulletins, the COVID Symptom Study may not be a representative sample of the population. It is reliant on app users and so captures only some cases in hospitals, care homes and other communities where fewer people use the app. To account for this, the model adjusts for age and deprivation when producing UK estimates. The larger sample size allows for detailed geographic breakdown.

Real-time Assessment of Community Transmission-1 and -2 (REACT-1 and -2), England

Like our study, the Real-time Assessment of Community Transmission-1 (REACT-1) survey, led by Imperial College London, involves taking swab samples to test for COVID-19 antigens to estimate the prevalence and transmission of the virus that causes COVID-19 in the community. Each round of the study currently involves around 160,000 participants aged five years and over, selected from a random cross-section sample of the general public from GP registration data. It is also possible to look at trends in infection rates by different characteristics, such as age, sex, ethnicity, symptoms and key worker status through the study.

One of the main differences from our COVID-19 Infection Survey is that the REACT surveys do not require follow-up visits, as the study is interested primarily in prevalence at a given time point.

Public Health England surveillance

Public Health England (PHE) also publishes an estimate of the prevalence of antibodies in the blood in England using blood samples from healthy adult blood donors. PHE provides estimates by region and currently do not scale up to England.

Estimates in our bulletins and those published by PHE are based on different tests; PHE estimates are based on testing using the Euroimmun assay method, while blood samples in our survey are tested for antibodies by research staff at the University of Oxford using a novel ELISA.

For more information about our antibody tests, see the COVID-19 Infection Survey protocol.

Insights

The ONS's latest insights tool provides an overview of the coronavirus (COVID-19) pandemic in the UK bringing together data from across the ONS and other data sources to explore the latest data and trends.

Accessibility and clarity

(Accessibility is the ease with which users are able to access the data, also reflecting the format in which the data are available and the availability of supporting information. Clarity refers to the quality and sufficiency of the release details, illustrations and accompanying advice.)

Our recommended format for accessible content is a combination of HTML webpages for narrative, charts and graphs, with data being provided in usable formats such as Excel spreadsheets. Our website also offers users the option to download the narrative in PDF format. Our outputs conform to the ONS Web accessibility policy in terms of formats and font sizes and the presentation of tables and charts.

More details on related releases can be found on the release calendar on GOV.UK. If there are any changes to the pre-announced release schedule, public attention will be drawn to the change and the reasons for the change will be explained fully.

Early management information from the Coronavirus (COVID-19) Infection Survey is made available to government decision-makers to inform their response to COVID-19. Occasionally, we may publish figures early if it is considered in the public interest. We will ensure that we pre-announce any ad hoc or early publications as soon as possible. These will include supporting information where possible to aid user understanding. This is consistent with guidance from the Office for Statistics Regulation.

In addition to this Quality and Methodology Information Report, quality and methods information is included in our weekly bulletin.

COVID-19 Infection Survey data are available in our Secure Research Service (SRS); this provides access to microdata and disclosive data, which have the potential to identify individuals. Access to such data requires Approved Researcher accreditation.

Timeliness and punctuality

(Timeliness describes the length of time between data availability and the event they describe. Punctuality is the time lag between the actual delivery of data and the target date on which they were scheduled for release as announced in an official release calendar.)

Survey fieldwork for the Coronavirus (COVID-19) Infection Survey began in England on 26 April 2020 and was expanded to cover Wales, Northern Ireland and Scotland across summer and autumn 2020. Headline figures were provided for each country as soon as sample sizes were sufficiently large to allow for good quality estimates to be produced.

The main aim of the COVID-19 Infection Survey is to provide data on the spread of infection to inform the public and organisations involved in decision-making. The survey also provides valuable information on characteristics of people testing positive (such as symptoms or amount of contact with others) and estimates of the population who would test positive for antibodies. Therefore, these data need to be collected, processed and published within a short time frame. Our typical publications are as follows:

- on a weekly basis, we publish results from the survey in our COVID-19 Infection Survey bulletin, which includes headline positivity estimates for England, Wales, Northern Ireland and Scotland, as well as an estimate of the incidence rate for England when possible and a breakdown by age
- on a fortnightly basis, we publish analysis on the characteristics of those testing positive for COVID-19
- on a fortnightly basis, we publish analysis on the number of people testing positive for COVID-19 antibodies

Other products such as blogs and technical articles are also published on an ad hoc basis. For more details on related releases, the GOV.UK release calendar is available online and provides advance notice of release dates.

Reference dates

We aim to provide the estimates of positivity rates and incidence that are most timely and most representative of each week. We decide the most recent week we can report on is based on the availability of test results for visits that have already happened, accounting for the fact that swabs have to be couriered to the labs, tested and results returned. Typically, the cut-off date for data that are published on the Friday will be the previous Saturday. For example, our bulletin published on Friday 16 July 2021 included data related to 4 to 10 July 2021.

Within the most recent week, we provide an official estimate for positivity rate and incidence based on a reference point from the modelled trends. For positivity rates, we can include all swab test results, even from the most recent visits. Therefore, although we are still expecting further swab test results from the labs, there is sufficient data for the official estimate for infection to be based on a reference point after the start of the reference week. To improve stability in our modelling while maintaining relative timeliness of our estimates, we report our official estimates based on the midpoint of the reference week.

The calculation of incidence uses time between two tests; so, for example, a participant who was last seen two weeks ago and is not due their next visit for another two weeks only contributes to the model up to two weeks ago. Our official estimates of incidence are therefore based on the first day of the reference week.

Why you can trust our data

The Office for National Statistics (ONS) is the UK's largest independent producer of statistics and its National Statistical Institute. The Data Policies and Information Charter details how data are collected, secured and used in the publication of statistics. We treat the data that we hold with respect, keeping it secure and confidential, and we use statistical methods that are professional, ethical and transparent. View more information about our data policies.

The COVID-19 Infection Survey has been carefully designed and tested and is being delivered in partnership with University of Oxford, University of Manchester, Public Health England and Wellcome Trust.

Output quality trade-offs

(Trade-offs are the extent to which different dimensions of quality are balanced against each other.)

Provisional estimates and revisions

The general principle applied to the Coronavirus (COVID-19) Infection Survey will be that when data are found to be in error, both the data and any associated analysis that has been published by the Office for National Statistics (ONS) will be revised in line with our revisions and corrections policy.

There are a number of reasons why we may wish to revise Coronavirus (COVID-19) Infection Survey estimates once they have been published and/or the datasets disseminated, including:

- errors are discovered in raw, or derived variables
- initial estimates are released with the expectation that these may be revised and updated as further data become available; for example, the use of models where later data points will affect the modelled estimate at earlier time periods
- a significant methods change is made

Revisions made because of errors discovered in raw, or derived variables

While every effort is made to thoroughly check the data before they are either published or released for dissemination, errors do on occasion occur. This can include errors with the analysis produced, such as categories not including the correct people, or errors made when data are inputted into spreadsheets. When errors occur, corrections are made in a timely manner, announced and clearly explained to users in line with the ONS guide to statistical revisions. Work is also undertaken to mitigate the same error happening again, for example, by reviewing and improving code.

Revisions made when more recent estimates become available

Modelling is used to produce times series estimates of positivity. Without modelling, changes in the point estimates of positivity over time could be quite erratic, caused by statistical uncertainty in the data (small sample sizes and low prevalence rates). This could provide time series that would not be considered a credible description of real-world changes, which would be much smoother. However, the use of modelling means that the estimate for any specified time point will be subject to revision as more time points are added to the model.

Therefore, estimates presented in our weekly bulletin are provisional results and subject to revision. Modelled estimates include all swab results that are available at the time the official estimates are produced. This is done to provide timely estimates to government decision-makers.

Official estimates should be used to understand the positivity rate for a single point in time. This is based on the modelled estimate for the latest week and is our best and most stable estimate. Additional swab tests that become available after this are included in subsequent models, meaning that modelled estimates can change slightly as additional data are included.

A new model for the most recent six-week period available is produced for each weekly bulletin, meaning that estimates for days within that six-week period that were covered in previous bulletins are revised. The modelled estimate is more suited to understand the recent trend, given it is regularly updated to include new test results and smooths the trend over time. In line with the ONS guide to statistical revisions, it is made clear in the bulletin that figures are initial estimates and subject to revision later.

Revisions made due to significant method changes

The COVID-19 Infection Survey was rapidly set up in response to the COVID-19 pandemic and launched on 26 April 2020. Because the survey is relatively new and there is an ongoing need for analysis to be responsive to the changing nature of the pandemic, methodological changes are inevitable.

In line with the ONS guide to statistical revisions and the Code of Practice for Statistics (Quality 2.5), when possible users are consulted and provided with advance notice about changes to methods, explaining why the changes are being made. When a methods change is made, a consistent time series is produced, with back series provided where possible. Users are made aware of the nature and extent of the change within the publications.

Concepts and definitions

(Concepts and definitions describe the legislation governing the output as well as harmonisation principles and classifications used in the output.)

Community

The Coronavirus (COVID-19) Infection Survey presents estimates for the number of current COVID-19 infections within the community population; community in this instance refers to private residential households and it excludes those in hospitals, care homes and/or other institutional and communal establishment settings.

Positivity rate

The positivity rate is the percentage of people who test positive for COVID-19 at a given point in time. We use current COVID-19 infections to mean testing positive for SARS-CoV-2, with or without having symptoms, on a swab taken from the nose and throat. This is different to the incidence rate, which is a measure of only the new infections in a given time period.

Incidence rate

The estimates of incidence of polymerase chain reaction (PCR)-positive cases use a new method based on our positivity estimate. This gives the rate at which new positives occur, and subsequently become detectable, within the population. The new incidence method uses an estimate of the length of time for which an individual will test positive, based on modelling the time from first positive to first subsequent negative test in the survey. This estimate is used alongside the positivity model to produce an incidence estimate. For more information on this method of incidence please see our methods article.

Characteristics

Participants are asked to provide their ethnicity and occupation (among other things) in the participant questionnaire to allow analysis of the characteristics of those testing positive for COVID-19.

The options provided on the questionnaire for ethnicity are harmonised to allow for consistency and comparability of statistical outputs from different sources across the UK. The participant's occupation is provided in a free text box and responses are coded using the Standard Occupation Classification, again to allow for consistency and comparability of statistical outputs from different sources across the UK.

Geographic coverage

Survey fieldwork for the pilot study began in England on 26 April 2020. Survey fieldwork in Wales began on 29 June 2020, and since 7 August 2020 we have reported headline figures for Wales. Survey fieldwork began in Northern Ireland on 26 July 2020 and since 25 September 2020 we have reported headline figures for Northern Ireland. Survey fieldwork in Scotland began on 21 September 2020, and we have reported headline figures for Scotland since 23 October 2020.

Sub-regional analysis

Where possible, we present modelled estimates for the most recent week of data at the sub-regional level. This analysis was first presented in our weekly COVID-19 Infection Survey bulletin on 20 November 2020. To balance the granularity with the statistical power, we have grouped together local authorities into COVID-19 Infection Survey sub-regions. The geographies are a rules-based composition of local authorities, and local authorities with a population over 200,000 have been retained separately where possible.

The boundaries for these COVID-19 Infection Survey sub-regions can be found on the Open Geography Portal.

6 . Methods used to produce the Coronavirus (COVID-19) Infection Survey data

The Coronavirus (COVID-19) Infection Survey collects data to allow estimation of the number of people testing positive for COVID-19 and the number of new infections. For more information on the sampling method, data processing and analysis, quality assurance and dissemination of our survey results, view our methods article.

How we collect the data

Sampling method

At the start of the study, all respondents to the Coronavirus (COVID-19) Infection Survey were individuals who had previously participated in an Office for National Statistics (ONS) social survey, which means the number of ineligible addresses in the sample is substantially reduced. To take part, selected households were invited to opt into the survey by contacting IQVIA, a company working on behalf of the ONS, to arrange a visit.

In August 2020, we announced our plans to expand the study with the aim of increasing from 28,000 people tested per fortnight in England to 150,000 people tested per fortnight by October 2020. Random samples of households from AddressBase have been drawn to enable this expansion.

The sample is stratified geographically, ensuring we sample people from all local areas of the UK, and we adjust sample sizes to account for differing response rates to help maintain the representativeness of the sample. The survey is longitudinal. New panels are selected most weeks and each panel is surveyed initially for five weeks and thereafter monthly.

Coverage of the study was extended to include Wales, Northern Ireland and Scotland, with survey fieldwork beginning on 29 June 2020 in Wales, 26 July 2020 in Northern Ireland and 21 September 2020 in Scotland.

The likelihood that a sampled household will enrol for CIS decreases as time passes from receiving the initial contact letter, and response rate information for those initially asked to take part at the start of the survey can be considered relatively final. We provide response rates separately for the different sampling phases of the study. For up-to-date information on our response rates, please see our most recent bulletin.

We include children over the age of 2 years, adolescents and adults in the survey. This allows for estimation of clustering effects and analysis of within-household transmission, as well as reducing survey costs. We include everyone within the community population, not just those with symptoms.

More information on the sampling design for the COVID-19 Infection Survey, the data we collect and how the data are processed can be found in our methods article.

Data we collect

Nose and throat swab

We ask everyone aged 2 years or over in each household to have a nose and throat swab. Those aged 12 years and over take their own swabs using self-swabbing kits, and parents or carers use the same type of kits to take swabs from their children aged between 2 and 11 years. This is to reduce the risk to the study health workers and to respondents themselves. We take swabs from all households, whether anyone is reporting symptoms or not.

We take swabs to detect microbes of the infection caused by COVID-19 so we can estimate the number of people who are infected. To do this, laboratories use real-time reverse transcriptase polymerase chain reaction (RT-PCR). Diagnostic RT-PCR usually targets the viral ribonucleic acid (RNA)-dependent RNA polymerase (RdRp) or nucleocapsid (N) genes using swabs collected from the nose and throat.

More information on how the swabs are analysed can be found in the study protocol.

Blood sample

To capture data about people who have been vaccinated or have had COVID-19 in the past, we ask a sub-sample of adults aged 16 years or over to also give a sample of blood. The blood samples are taken at monthly intervals. Initially, 10% to 20% of adults aged 16 years or over surveyed within our household sample were asked to provide a blood sample.

From February 2021, a randomly selected sample of already enrolled households who are currently providing swab but not blood samples, and where at least one person has provided consent for 12 months' follow-up, have been invited to additionally provide blood samples at their monthly visits and extend their follow-up through to April 2022. This will be done to assess the impact of vaccine rollout on population level immunity.

The target is to achieve up to 125,000 people giving blood samples every month in England, and up to 7,500 in Wales, 5,500 in Northern Ireland and 12,000 in Scotland per month (150,000 in total across the UK) through to April 2022.

Blood samples are tested for antibodies using an assay for IgG immunoglobins, which are produced to fight the virus, irrespective of symptoms. More information on the methods around this antibody assay can be found in a study comparing its performance with four other assays.

Survey data

We collect information from each participant, including those aged 16 years and under, about:

- their socio-demographic characteristics
- any symptoms that they are experiencing
- whether they are self-isolating or shielding
- their occupation
- how often they work from home
- if they have received a vaccine
- whether the participant has come into contact with a suspected carrier of COVID-19

How we analyse the data

The primary objective of the study is to estimate the number of people in the population who test positive for COVID-19, with and without symptoms.

The analysis of the data is a collaboration between the ONS and researchers from the University of Oxford and University of Manchester, Public Health England and Wellcome Trust.

All estimates presented in our bulletins are provisional results. As swabs are not necessarily analysed in date order by the laboratory, we have not yet received test results for all swabs taken on the dates included in our bulletins. Estimates may therefore be revised as more test results are included.

Our headline estimates of the percentage of people testing positive in England, Wales, Northern Ireland and Scotland are the latest official estimates. Official estimates should be used to understand the positivity rate for a single point in time. This is based on the modelled estimate for the latest week and is our best and most stable estimate, used in all previous outputs. The modelled estimate is more suited to understand the recent trend. This is because the model is regularly updated to include new test results and smooths the trend over time.

The estimates for non-overlapping 14-day periods are produced using a different method of weighting to the model and are available for people who wish to compare infection levels over time in this way. Information about how the modelled and 14-day non-overlapping estimates are calculated can be found in our methods article.

Statistical testing

Where we have analysed the characteristics of people who have ever tested positive for COVID-19, we have used pairwise statistical testing to determine whether there was a significant difference in infection rates between pairs of groups for each characteristic. For instance, we used statistical testing to identify any evidence of differences in infection rates between those aged 2 to 11 years and those aged 12 to 19 years.

The test produces p-values, which provide the probability of observing a difference at least as extreme as the one that was estimated from the sample by chance. We use the conventional threshold of 0.05 to indicate evidence of differences not compatible with chance, although the threshold of 0.05 is still relatively weak evidence. P-values of less than 0.001 and 0.01 are considered to provide relatively strong and moderate evidence of difference between the groups being compared respectively.

How we disseminate the data

The Coronavirus (COVID-19) Infection Survey bulletin for the UK is published weekly on the ONS website. The release describes the main patterns and trends in the datasets for each of the four UK countries. We also release a fortnightly article detailing the characteristics and behaviours of those testing positive for COVID-19 as well as a fortnightly antibody and vaccinations article.

The ONS's latest insights tool provides an overview of the coronavirus (COVID-19) pandemic in the UK, bringing together data from across the ONS and other data sources to explore the latest data and trends.

For more detailed information on the COVID-19 Infection Survey design, how we process data and how data are analysed, view our methodology article.

7 . Other information

Assessment of user needs and perceptions

(The processes for finding out about uses and users, and their views on the statistical products.)

We hold regular weekly meetings with departments across government ensuring we keep up to date with changing user needs. We have a clear process for reviewing, prioritising and responding to user requests ensuring we balance the public good of the request with the resource required to meet it. In addition, the questionnaire is regularly reviewed which allows new information and questions (for example, which type of vaccine people have received) to be added in a timely way.

We receive feedback on our analysis from the UK government and the Scientific Advisory Group for Emergencies (SAGE). We welcome feedback and encourage users to provide feedback in our releases by including the following text:

We are continuously refining and looking to improve our modelling and presentations. We would welcome any feedback via email: infection.survey.analysis@ons.gov.uk.

Collaboration

The 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, the University of Manchester, Public Health England (PHE) and Wellcome Trust. 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
- Thomas House - University of Manchester, Department of Mathematics: Reader in Mathematical Statistics

This study is jointly led by the ONS and the Department for Health and Social Care (DHSC) working with the University of Oxford and Lighthouse laboratories to collect and test samples.

Useful links

Our methodology article provides further information around the survey design, how we process data and how data are analysed. The study protocol specifies the research for the study.

Our weekly COVID-19 Infection Survey bulletin provides headline estimates for the percentage of people testing positive for COVID-19 in England, Wales, Northern Ireland and Scotland, as well as an estimate of the incidence rate for England.

We also publish a fortnightly article on the characteristics of people testing positive for COVID-19 and a separate fortnightly article including antibody and vaccination data by UK country and region for England.

The Welsh Government publishes results from the COVID-19 Infection Survey that describe COVID-19 infections in Wales in English and in Welsh.

The Department of Health (Northern Ireland) publishes results from the COVID-19 Infection Survey that describe COVID-19 infections in Northern Ireland.

The Scottish Government publishes results from the COVID-19 Infection Survey that describe COVID-19 infections in Scotland.