

# Coronavirus (COVID-19) Infection Survey: quality and methodology information (QMI)

Quality and Methodology Information (QMI) 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 and Eleanor Fordham  
health.data@ons.gov.uk  
+44 1633 560499

**Release date:**

20 January 2023

**Next release:**

To be announced

## Notice

### 24 January 2023

Last updated: 3 August 2022. Note that this document primarily aims to reflect the design of the survey since the last update through to the end of 2022, rather than detailing all changes that have taken place since the survey began. This is to provide an updated QMI of the survey design, since our move to remote data collection in August 2022.

# Table of contents

1. [Methodology background](#)
2. [About this Quality and Methodology Information report](#)
3. [Quality summary](#)
4. [Quality characteristics of the Coronavirus \(COVID-19\) Infection Survey](#)
5. [Methods used to produce the Coronavirus \(COVID-19\) Infection Survey data](#)
6. [Other information](#)
7. [Collaboration](#)
8. [Useful links](#)

# 1 . Methodology background

- 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
- Number of participants: Enrolment for data collection via study worker home visits started on 26 April 2020 and ceased on 31 January 2022; at the time when enrolment stopped, there were 490,452 eligible individuals in 227,797 households enrolled in England, 30,146 eligible individuals in 14,346 households in Wales, 15,973 eligible individuals in 7,421 households in Northern Ireland and 48,351 eligible individuals in 23,941 households in Scotland. On 26 September 2022 after the move to remote data collection, a small number of invitations to enrol in the survey were sent to a new sample of households in Northern Ireland. The latest response rates, along with commentary, are found in our [Coronavirus \(COVID-19\) Infection Survey: technical dataset](#), Tables 2a to 2f.
- Achieved sample size: Between 1 May 2021 and 31 March 2022, an average of approximately 390,300 swab results per month (for estimating positivity rates) and 152,600 blood results per month (for estimating presence of antibodies) were analysed in the survey.
- Target sample size: Between 1 April 2022 and 31 March 2023, the sample target is to achieve a maximum of 227,300 swab tests from individuals aged 2 years and over every 28 days in England, 15,650 in Wales, 10,050 in Northern Ireland and 23,300 in Scotland (276,200 in total across the UK every 28 days); this equates to approximately 300,000 swab tests in total across the UK per month.
- The blood sample target, for the same period up to 31 March 2023, is to achieve up to 90,850 blood tests from individuals aged 8 years and over every 28 days in England, up to 6,300 in Wales, 4,150 in Northern Ireland and 9,200 in Scotland (110,500 in total across the UK every 28 days); this equates to approximately 120,000 blood tests in total across the UK per month.
- In July 2022 we moved from collecting data and samples through home visits by a study worker to a more flexible remote data collection method.

## 2 . About this Quality and Methodology Information report

In July 2022 we moved from collecting data and samples through home visits by a study worker to a more flexible remote data collection method. All questionnaires and swabs were completed remotely from 1 August 2022. This was to ensure that the survey remained as accessible and representative as possible for participants and allowed us to move to a more efficient method of data collection. Further information on these changes can be found in our [blog post](#) and our Coronavirus (COVID-19) Infection Survey: [methods article](#).

This quality and methodology report contains information up to the end of 2022 on the quality characteristics (including the [European Statistical System](#) five dimensions of quality) of the statistics produced as outputs based on Coronavirus (COVID-19) Infection Survey data collected through study worker home visits up to July 2022 and remote data collection from August 2022, 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
- reduce the risk of misusing data
- help you to decide suitable uses for the data
- understand the methods used to create the data

## 3 . Quality summary

### Important points about the Coronavirus (COVID-19) Infection Survey

In response to the COVID-19 pandemic, the Coronavirus Infection Survey was set up in April 2020 to estimate:

- how many people across England, Wales, Northern Ireland and Scotland would have tested positive for COVID-19 infection, regardless of whether they report experiencing symptoms
- the average number of new positive test cases per week
- the number of people who would have tested positive for antibodies against SARS-CoV-2 at different levels

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 communal settings are not included.

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

### Overview of the COVID-19 Infection Survey

The 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 21 September 2020.

The survey is based on a random sample of households to provide a nationally representative survey. We ask everyone aged 2 years and over in each household sample to take 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.

The survey was designed to find out 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
- the incidence of new positive infections

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.

To monitor antibodies against SARS-CoV-2 among the population, we ask a subset of participants aged 8 years and over to also give a sample of blood. This is done 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 aged 16 years and over surveyed within our household sample were asked to provide a blood sample. 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. In November 2021, we started collecting the blood samples from children aged between 8 and 15 years. In August 2022, all existing participants were invited to move from study worker home visits to remote data collection where participants complete the survey online or by telephone, and swab and blood samples are returned through the post (or by courier for some participants).

We use many different techniques to estimate the number of people testing positive for SARS-CoV-2 (the virus that causes the coronavirus (COVID-19) disease) and the presence of antibodies above different levels, broken down by different characteristics, including age and region. Results are adjusted 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, Welsh Government, Northern Ireland Executive and Scottish Government are the main users of the COVID-19 Infection Survey. Our statistics are used to track the progress of the pandemic in the UK and to help inform decisions about coronavirus restrictions and related policies. Matters such as restrictions and policies related to the pandemic are devolved, with the Welsh Government, Northern Ireland Executive, and Scottish Government using data from the survey to inform these decisions.

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

The results of the COVID-19 Infection Survey contributed to the Scientific Advisory Group for Emergencies' (SAGE) estimates of the rate of transmission of the infection (often referred to as "R") and continues to contribute to epidemic estimates produced by the UK Health Security Agency. The survey also continues to provide important information about the socio-demographic characteristics of the people and households who have contracted COVID-19. Results are 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 pandemic, the characteristics of those testing positive (such as their occupation, work location, travel status and symptoms reported) and any possible inequalities associated with those.

The media also report widely on the COVID-19 Infection Survey data and the public are interested in the statistics produced by the 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 Organization (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 presence of antibodies in the population at different levels and how these change over time

The data cannot be used for:

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

## Strengths and limitations of the COVID-19 Infection Survey

These statistics were initially produced rapidly in response to developing world events. The Office for Statistics Regulation (OSR), 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](#).

One of the survey's main strengths is that survey subjects are a random sample of the population with a large sample size. This means that unlike other sources, such as national testing programmes, which includes primarily people reporting symptoms or their close contacts, the COVID-19 Infection Survey also identifies people not reporting symptoms. The survey presents timely estimates weekly or fortnightly on a range of domains of interest such as the presence of antibodies at different levels and symptoms reported, and social factors like return to work and travel.

The estimates presented in our weekly bulletin and monthly bulletins 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 measure. There are many sources of uncertainty including:

- uncertainty in the test results; 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 methods used to adjust for this
- uncertainty in the models used; some models borrow strength across smaller population groups and there could be possible incoherence between modelled estimates and the underlying truth
- quality of data collected in the questionnaire and the method of the test administration

Information on the main sources of uncertainty are presented in our [methods article](#) and in our blog; [Accuracy and confidence: why we trust the data from the COVID-19 Infection Survey](#).

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

## Relevance

Relevance is the degree to which statistical outputs meet current and potential user needs.

The COVID-19 Infection Survey estimates the percentage of the population who would test positive for COVID-19 and helps track the current extent of infection and transmission of COVID-19 among the population.

We use the number of people testing positive for COVID-19 with polymerase chain reaction (PCR) tests via nose and throat swabs to calculate the proportion of the population living in private households who would 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. We report the positivity rate rather than the prevalence rate, as to calculate the latter, we would need an accurate understanding of the swab test's sensitivity (true-positive rate) and specificity (true-negative rate). We also regularly present data on the characteristics of people testing positive for COVID-19. Topics to be included have varied depending on user need at different stages of the COVID-19 pandemic.

The incidence rate is a measure of the estimated number of new PCR positive cases per day, per 10,000 people at a given point in time. It is different to our positivity rate, which is an estimate of all current PCR positive cases at a point in time, regardless of whether the infection is new or existing. We use blood test results to identify individuals who have antibodies against SARS-CoV-2 at different levels. This helps us understand the impact of vaccinations and COVID-19 infections, as well as possible levels of protection or vulnerability in different population groups over time. These blood tests are taken from individuals aged 8 years and over from a randomly selected subsample of households.

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, planning for services and vaccination rollout.

## Analysis feeding into calculation of the reproduction number, R

The statistics produced from this survey contributed to the modelling used in the calculation of 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), [built a consensus on the value of R](#) based on expert scientific advice from multiple academic groups. This is now produced by the UK Health Security Agency (UKHSA).

## 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 estimate's value to the subsequent estimate's value.

## Uncertainty

The estimates presented in our weekly bulletin and monthly bulletins contain uncertainty. There are many sources of uncertainty, but the main sources in the information presented include:

- uncertainty in the test results; false-positives, false-negatives and timing of the infection
- the data are based on a sample of people rather than the whole population, so there is some statistical uncertainty in the estimates
- uncertainty on models used
- quality of data collected in the questionnaire and the method of the test administration

Results come directly from the laboratories that perform the PCR and blood tests, and no test is perfect. There will be false-positive and false-negative results from the tests; 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 [Section 5. Test sensitivity and specificity](#), in our methods article.

Any estimate based on a random sample contains some uncertainty. To quantify this uncertainty in our analyses, we use credible intervals and confidence intervals. A credible or confidence interval gives an indication of the degree of uncertainty of an estimate. A wider interval indicates more uncertainty in the estimate. Overlapping intervals indicate that there may not be a true difference between two estimates. Further information on confidence and credible intervals can be found in [Section 11. Confidence intervals and credible intervals](#) are also available in our methods article.

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 by 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 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 Office for National Statistics (ONS) surveys who agreed to further contact about other research (this now makes up a small subset of the overall sample), and the second was by writing to "the householder" at addresses selected from AddressBase (a sampling frame). For more information on the sampling process, please 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.

## 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. Where a Bayesian regression model is used for our analyses of swab and antibody positivity at different thresholds, 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 published earlier in the pandemic, confidence intervals were provided. These 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 11. 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 survey, this is important because estimates of COVID-19 positivity and antibody positivity at different levels 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 overall 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 overall sample is representative of all regions and representative of Wales, Northern Ireland and Scotland in terms of population share. At the UK level, those reporting white ethnicity and households of three or more are overrepresented. In contrast, households of one person or two people are underrepresented.

The following tables provide an example of some of the representative analysis for the sample providing swabs across the UK from the start of the UK-wide survey up to 30 June 2022. The unweighted response population is the actual number of people taking part in the survey, whereas the adjusted population has been adjusted to be representative of the target population using methods like those used for our main analysis. The calibration step ensures coherence for those variables and categories used in the adjustment. Hence the "actual proportion" and "adjusted proportion" agree for every category of age and sex because they are used in the process.

Table 1a: Actual UK population by sex

### UK

#### Relevant actual proportion

Sex	Number of people	Proportion	Cumulative proportion
<b>Female</b>	32,753,171	0.51	0.51
<b>Male</b>	31,910,622	0.49	1.00

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

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

### UK

#### Response population

Sex	Unadjusted			Adjusted		
	Number of people	Proportion (excluding missing)	Cumulative proportion	Number of people	Proportion (excluding missing)	Cumulative proportion
<b>Female</b>	283,074	0.53	0.53	32,753,172	0.51	0.51
<b>Male</b>	252,047	0.47	1.00	31,910,622	0.49	1.00
<b>Missing</b>	0			0		

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

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

**UK**

**Representativeness**

	Absolute difference to actual population	Proportional difference to actual population	Unadjusted	Adjusted
<b>Sex</b>	Unadjusted	Adjusted	Unadjusted	Adjusted
<b>Female</b>	0.02	0.00	4%	0%
<b>Male</b>	-0.02	0.00	-5%	0%

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

Table 1d: Actual UK population by age

**UK**

**Relevant actual proportion**

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

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

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

**UK**

**Response population**

Age	unadjusted			adjusted		
	Number of people	Proportion (excluding missing)	Cumulative proportion	Number of people	Proportion (excluding missing)	Cumulative proportion
<b>2 to 11</b>	33,507	0.06	0.06	7,844,965	0.12	0.12
<b>12 to 16</b>	29,147	0.05	0.12	4,034,610	0.06	0.18
<b>17 to 24</b>	35,403	0.07	0.18	5,936,426	0.09	0.28
<b>25 to 34</b>	52,359	0.10	0.28	8,845,897	0.14	0.41
<b>35 to 49</b>	103,068	0.19	0.47	12,731,831	0.20	0.61
<b>50 to 69</b>	176,814	0.33	0.80	16,363,920	0.25	0.86
<b>70+</b>	104,823	0.20	1.00	8,906,145	0.14	1.00
<b>Missing</b>	0			0		

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

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

**UK**

**Representativeness**

<b>Age</b>	Absolute difference to actual population		Proportional difference to actual population	
	Unadjusted	Adjusted	Unadjusted	Adjusted
<b>2 to 11</b>	-0.06	0	-50%	-3%
<b>12 to 16</b>	-0.01	0	-9%	4%
<b>17 to 24</b>	-0.02	0	-27%	1%
<b>25 to 34</b>	-0.04	0	-29%	-1%
<b>35 to 49</b>	0.00	0	-2%	0%
<b>50 to 69</b>	0.08	0	31%	0%
<b>70+</b>	0.06	0	43%	1%

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

Table 1g: Actual UK population by ethnicity

**UK**

**Relevant actual proportion**

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

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

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

## UK

## Response population

Ethnicity	Unadjusted			Adjusted		
	Number of people	Proportion (excluding missing)	Cumulative proportion	Number of people	Proportion (excluding missing)	Cumulative proportion
<b>White</b>	493,026	0.92	0.92	59,113,687	0.91	0.91
<b>Asian/Asian British</b>	22,128	0.04	0.96	2,799,307	0.04	0.96
<b>Black/African/Caribbean/Black British</b>	5,476	0.01	0.97	701,434	0.01	0.97
<b>Mixed/Multiple ethnic groups</b>	9,598	0.02	0.99	1,422,681	0.02	0.99
<b>Other ethnic groups</b>	4,804	0.01	1.00	617,125	0.01	1.00
<b>Missing</b>	89			9,560		

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

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

## UK

## Representativeness

Ethnicity	Absolute difference to actual population		Proportional difference to actual population	
	Unadjusted	Adjusted	Unadjusted	Adjusted
<b>White</b>	0.06	0.06	7%	7%
<b>Asian/Asian British</b>	-0.03	-0.03	-44%	-42%
<b>Black/African/Caribbean/Black British</b>	-0.02	-0.02	-69%	-67%
<b>Mixed/Multiple ethnic groups</b>	0.00	0.01	9%	34%
<b>Other ethnic groups</b>	-0.01	-0.01	-49%	-46%

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

To address the fact that some individuals in the survey will have stopped participating (for example, if they have moved house) 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, and region (for England). 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 those reporting different ethnicities in the sample. This includes strategies such as sending out reminders to 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 as well as triangulation with other coronavirus (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 national testing programmes

Each UK nation ([England](#), [Wales](#), [Northern Ireland](#) and [Scotland](#)) has made use of a national testing system. These ensured that anyone who developed symptoms of COVID-19, or was in close contact of someone with symptoms, could quickly be tested to find out if they had the virus. Some nations included targeted asymptomatic testing of NHS and social care staff, and care home residents. Additionally, national testing systems could be used to identify close recent contacts of anyone who tested positive for COVID-19, who were then advised to follow the relevant government guidelines at the time. We published an [article that compares the methods used in the COVID-19 Infection Survey and NHS Test and Trace in England](#). Information on changes to COVID-19 testing programmes for [England](#), [Wales](#), [Northern Ireland](#) and [Scotland](#) has been published by the relevant official sources.

In comparison with data from national testing programmes, the statistics presented in our weekly bulletin take a representative sample of the population living 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 population with COVID-19 who do not report any evidence of symptoms, which is one of the unique features of the survey.

## Laboratory confirmed cases in the UK

The UK Health Security Agency (UKHSA) 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. They are presented 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 each nation - [England](#), [Wales](#), [Northern Ireland](#) and [Scotland](#) - is also available.

## Other studies

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

## 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, involved taking swab samples to PCR test for COVID-19 to estimate the prevalence of the virus that causes COVID-19 in the community. Each round of the study involved around 160,000 participants aged 5 years and over, selected from a random cross-section sample of the public from GP registration data. It was 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 did not require follow-up visits, as the study was interested primarily in prevalence at a given time point. The REACT-1 study has now ended, and [REACT-1 study of coronavirus transmission: March 2022 final results](#) report presents the last findings from this survey.

## 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 aimed initially 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 aged 18 years and over 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. This is 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 ZOE 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 a [detailed geographic breakdown](#).

### Insights

The ONS' [latest insights](#) provides an overview of the 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 can 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 web pages 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 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 (OSR).

In addition to this [Quality and Methodology Information report](#), quality and methods information is included 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.

The COVID-19 Infection Survey provides data on current rates of coronavirus 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 reinfections) and estimates of the population who would test positive for antibodies against SARS-CoV-2 among the population, at different levels. These data are collected, processed, and published within a short time frame. Our typical publications include:

- the weekly bulletin, which contains results from swab tests taken in the survey, and headline swab positivity estimates for England, Wales, Northern Ireland and Scotland, as well as an estimate of incidence rates when possible, and breakdowns by age and sometimes different variants
- the characteristics of those testing positive for COVID-19 every fortnight
- the number of people testing positive for COVID-19 antibodies at different levels every fortnight
- additional products, such as blogs and technical articles, which are published on an ad-hoc basis
- the [latest insights](#) about the pandemic, updated as and when data become available

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 swab positivity rates and incidence that are most timely and most representative of each week. 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 must be couriered to the labs, tested, and results returned. Typically, the cut-off date for data will be 10 days prior to the date of publication. For example, the [bulletin published on Friday 20 January 2023](#) included data related to 4 to 10 January 2023.

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.

We also calculate incidence rates where possible, and this measures new PCR positive cases per day per 10,000 people, in each time period. Data are based only on new confirmed positive coronavirus test results and differ to positivity rates. Positivity rates are an estimate of all current PCR positive cases at a point in time, regardless of whether the infection is new or existing. Our official estimates of incidence are based on the first day of the reference week.

## Why you can trust our data

The ONS is the UK's largest independent producer of statistics and its National Statistical Institute. The [Data Policies and Information Charter](#), available on the ONS website, details how data are collected, secured and used in the publication of statistics. We treat the data that we hold with respect, keeping the data secure and confidential. We use statistical methods that are professional, ethical and transparent. More information about our data policies is [available](#) in the About Us section of our website.

The COVID-19 Infection Survey has been carefully designed and tested and is being delivered in partnership with the University of Oxford, University of Manchester, UK Health Security Agency (UKHSA) and the 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 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 ONS will be revised in line with our [revisions and corrections policy](#).

There are several reasons why we may wish to revise the 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
- 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 the data are published or released for dissemination, errors do occasionally occur. This can include errors 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 to existing estimates after more recent estimates become available

Modelling is used to produce estimates of positivity over time. Without modelling, changes in the point estimates of positivity over time could be erratic, caused by statistical uncertainty in the data (small sample sizes and low prevalence rates), and hence could provide time series that would not be considered a credible description of real-world changes. 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.

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. This means 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, so 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 because of significant method changes

The COVID-19 Infection Survey was rapidly set up in response to the pandemic and launched on 26 April 2020. Because the survey was relatively new, and there remains 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 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 communal establishment settings.

## Positivity rate

The positivity rate is the percentage of people who would 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 new infections in a given time period.

## Incidence rate

The estimates of incidence of polymerase chain reaction (PCR)-positive cases use a method based on our positivity estimate. This gives the rate at which new positive cases occur, and subsequently become detectable, within the population. The method uses an estimate of the length of time for which an individual will test positive, based on modelling the time from first positive test 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 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 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 survey sub-regions can be found on the [Open Geography Portal](#).

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

The survey collects data to allow estimation of the number of people who would test positive for COVID-19 and the number of new test positive cases. For more information on the sampling method, data processing and analysis, quality assurance and dissemination of our survey results, see our [methods article](#).

## How we collect the data

### Sampling method

At the start of the study, all those invited to join COVID-19 Infection Survey were individuals who had previously participated in an Office for National Statistics (ONS) social survey, to increase the speed at which sufficient participants could be recruited to provide estimates to government. This means the number of ineligible addresses in the sample was 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 were drawn to enable this expansion.

The sample was stratified geographically, ensuring people were sampled from all local areas of the UK. We adjusted sample sizes to account for differing response rates to help maintain the representativeness of the sample. The survey is longitudinal. New panels were 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.

In December 2021, we stopped sending out new invites to people to achieve a reduced sample from 1 April 2022 onwards. However, on 26 September 2022, after the move to remote data collection, a small number of invitations to enrol in the survey (fewer than 50) were sent to a new sample of households in Northern Ireland.

All children aged 2 years and over, adolescents and adults are included 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 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 and 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. 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 risk to both study health workers and participants. We take swabs from all households, whether anyone is reporting symptoms or not.

We take swabs to detect the virus that causes COVID-19 infection so we can estimate the number of people who are infected. To do this, our laboratories use [real-time reverse transcriptase polymerase chain reaction \(RT-PCR\)](#). The RT-PCR test looks for three genes present in coronavirus: N (nucleocapsid) protein, S (spike) protein and ORF1ab. Each swab can have one, two or all three genes detected.

More information on how the swabs are analysed can be found in the [study protocol](#).

#### Blood samples

To capture data about levels of immunity, including from vaccinations or having had COVID-19 in the past, we ask a sub-sample of participants aged 8 years and over to also give a sample of blood. The blood samples are taken at monthly intervals. Blood samples are tested for antibodies using an assay for spike 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 under 16 years, 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 vaccination, whether they have come into contact with someone with COVID-19, and their contacts with others inside and outside their home.

## How we analyse the data

The primary objective of the study is to estimate the number of people in the population who would 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 and UK Health Security Agency (UKHSA).

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.

## 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 those who have previously had a COVID-19 infection and those who have not.

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 COVID-19 Infection Survey 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 [monthly bulletin detailing the characteristics and behaviours of those testing positive for COVID-19](#) as well as a [monthly antibody bulletin](#).

Our [latest insights](#) provides an overview of the COVID-19 pandemic in the UK and brings together data from across the ONS and other data sources to explore the latest data and trends.

For more detailed information on the survey's design, how we process data and how data are analysed, see our [methods article](#).

# 6 . 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 key 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. This is to ensure 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 vaccination people have received) to be added to the questionnaire in a timely way.

We receive feedback on our analysis from the UK Government. 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: [health.data@ons.gov.uk](mailto:health.data@ons.gov.uk)

## 7 . 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, UK Health Security Agency (UKHSA) 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 University of Oxford, working with UKHSA and Lighthouse Laboratories for the collection and testing of test samples.

## 8 . Useful links

Our [methods 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 bulletin](#) provides headline estimates for the percentage of people testing positive for coronavirus (COVID-19) in England, Wales, Northern Ireland, and Scotland, as well as an estimate of the incidence rate for England.

We also publish a [monthly bulletin on the characteristics of people testing positive for COVID-19](#) and a separate [monthly bulletin 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 Coronavirus Infection Survey that describe COVID-19 infections in Northern Ireland.

The [Scottish Government publishes results](#) from the Coronavirus Infection Survey that describe COVID-19 infections in Scotland.