

COVID-19 Infection Survey (Pilot): methods and further information

This methodology guide is intended to provide information on the methods used to collect the data, process it, and calculate the statistics produced from the COVID-19 Infection Survey (pilot).

Contact:
Kara Steel and Philippa
Haughton
infection.survey.analysis@ons.
gov.uk
+44 (0)203 973 4761

Release date:
21 September 2020

Next release:
To be announced

Table of contents

1. [COVID-19 Infection Survey](#)
2. [Study design: sampling](#)
3. [Study design: data we collect](#)
4. [Processing swabs and blood samples](#)
5. [Test sensitivity and specificity](#)
6. [Linking survey data, swab results and blood samples](#)
7. [Cleaning the data](#)
8. [How we analyse the data](#)
9. [Modelling](#)
10. [14-day estimates](#)
11. [Statistical testing](#)
12. [Analysis feeding into R](#)
13. [Uncertainty in the data](#)

1 . COVID-19 Infection Survey

The coronavirus (COVID-19) pandemic is having a profound impact across the UK. In response to the pandemic, the COVID-19 Infection Survey (Pilot) measures:

- how many people across England and Wales test positive for COVID-19 infection at a given point in time, regardless of whether they report to experiencing symptoms
- the average number of new infections per week over the course of the study
- the number of people who test positive for antibodies, to indicate how many people are ever likely to have had the infection

The results of the pilot 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.

The Office for National Statistics (ONS) is working with the University of Oxford, University of Manchester, Public Health England, Wellcome Trust, IQVIA and the UK Biocentre Milton Keynes to run the pilot study in England, which was launched in mid-April. We intend to expand the size of the sample over the next 12 months and look to cover people across all four UK nations.¹ This guide covers methods used in the pilot study only.

This methodology guide is intended to provide information on the methods used to collect the data, process it, and calculate the statistics produced from the COVID -19 Infection Survey (Pilot). We will continue to expand and develop methods as the study progresses, and we will publish an updated methodology guide in autumn 2020.

It can be read alongside:

- the [COVID-19 bulletin](#), which gives weekly headline statistics
- the [study protocol](#), which outlines the study design and rationale
- the [study guide](#), which explains to participants what taking part in the study entails

Notes for COVID-19 Infection Survey

1. [Protocol, p. 8.](#)

2 . Study design: sampling

At the start of the pilot study at the end of April, the sample for the survey was drawn mainly from the Annual Population Survey (APS), which consists collectively of those who successfully completed the last wave of the Labour Force Survey (LFS) or local LFS boost, and who had consented to future contact regarding research.

Around 38,000 households respond to the LFS each quarter and it is the largest regular household survey in the UK. The sampling frame for the LFS is the Postal Address File of small users, which contains approximately 26 million addresses. Only private households are included in the sample. People living in care homes, other communal establishments and hospitals are not included. Only private households in England are included in the pilot study.

At the start of the pilot study we invited about 20,000 households to take part, anticipating that this would result in approximately 21,000 individuals from approximately 10,000 households participating. Since the end of May, additional households have been invited to take part in the survey each week (roughly 5,000 a week).

At the start of the study, all respondents to the COVID-19 Infection Survey were individuals who have 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, invited households opted into the survey by contacting IQVIA, a company working on behalf of the ONS, to arrange a visit.

Since the end of July, we have further expanded the survey to invite a random sample of households from AddressBase, which is a commercially available list of addresses maintained by the Ordnance Survey.

In line with our plans to increase our overall sample size, we prioritised areas under government local restriction because of an outbreak of the coronavirus (COVID-19). We have invited 40,000 extra households from 14 local authorities within selected local authorities in Greater Manchester, Lancashire and West Yorkshire to participate in this study.

We also boosted our sample in London, inviting 50,000 extra households to increase the household involvement rates in this area.

In August, we [announced our plans](#) to further 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 until March 2021. A random sample of households from AddressBase will be invited for this expansion.

Likelihood of enrolment decreases over time, and response rate information for those initially asked to take part at the start of the survey can be considered as relatively final. We provide response rates separately for the different sampling phases of the study.

Table 4a presents response rates for those asked to take part at the start of the survey and can be considered as relatively final. Table 4b presents response rates for those invited from the end of May, and Table 4c presents response rates for those invited from the end of July, where enrolment is continuing. These response rates cannot be regarded as final since those who are invited are not given a time limit in which to respond. For up to date information on our response rates, please see [our most recent bulletin](#).

Additionally, 10% of adults over 16 years old surveyed within our household sample were asked to provide a blood sample, which is used to test for the presence of antibodies to the coronavirus (COVID-19). These adults are sampled from people who are currently participating in the [ONS Opinions COVID-19 Survey](#) for practical reasons and to enable future data linkage to more detailed data on [reported self-isolation behaviours](#).

More information about how participants are sampled can be found in the [study protocol](#).

We also need to know more about how the virus is transmitted and immunity in individuals who are positive; whether individuals who have had the virus can be re-infected symptomatically or asymptotically; and about incidence of new infection in individuals who have not been exposed to the virus before.

To address these questions, we collect data over time. Every participant is swabbed once; participants are also invited to have repeat tests every week for the first five weeks as well as monthly for a period of 12 months in total.

3 . Study design: data we collect

Nose and throat swab

We ask everyone aged 2 years or older in each household to have a nose and throat swab. Those aged 12 years and older 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 old. 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 the coronavirus (COVID-19) so we can measure the number of people who are infected. To do this, [laboratories use real-time reverse transcriptase polymerase chain reaction \(RT-PCR\) \(PDF, 731KB\)](#). Diagnostic RT-PCR¹ 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](#).

However, RT-PCR from nose and throat swabs may be [falsely negative](#)², because of their quality or the timing of collection. The virus in nose and throat secretions peak in the first week of symptoms but may decline below the limit of detection in patients who present with symptoms beyond this time frame. For people who have been infected and then recovered, the RT-PCR technique provides no information about prior exposure or immunity. To address this, we also collect blood samples to test for antibodies (see following section).

Blood sample

To capture data about people who have had COVID-19 but have since recovered, we aim to ask adults aged 16 years or older from around 10% of enrolled households to also give a sample of blood, which will be taken by a trained nurse, phlebotomist or healthcare assistant. The antibody sample is a follow-up after four swab samples are collected.

Blood samples are tested for antibodies using an assay for IgG immunoglobins³, 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.

We do not take blood from anyone in a household where someone has symptoms compatible with COVID -19 infection, or is currently self-isolating or shielding, to make sure that study staff always stay at least two metres away from them.

Survey data

We collect information from each participant, including those under 16 years of age, 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, and whether the participant has come into contact with a suspected carrier of COVID-19.

Notes for Study design: data we collect

1. Konrad R, Eberle U, Dangel A, and others. Rapid establishment of laboratory diagnostics for the novel coronavirus SARS-CoV-2 in Bavaria, Germany, February 2020. *Euro Surveill* 2020; 25(9).
2. To KK, Tsang OT, Leung WS, and others. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect Dis* 2020.
3. Li Z, Yi Y, Luo X, and others. Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J Med Virol* 2020.
4. Zhou P, Yang XL, Wang XG, and others. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; 579(7798): 270-3
5. National COVID Testing Scientific Advisory Panel. Antibody testing for SARS-CoV-2 using ELISA and lateral flow immunoassays. [MedRxiv 2020](#).
6. The National SARS-CoV-2 Serology Assay Evaluation Group. [Head-to-head benchmark evaluation of the sensitivity and specificity of five immunoassays for SARS-CoV-2 serology on more than 1,500 samples](#).

4 . Processing swabs and blood samples

The nose and throat swabs are sent to the National Biosample Centre at Milton Keynes. Here, they 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. Swabs are discarded after testing.

Blood tubes are kept in a cool bag during the day, and then sent to the University of Oxford overnight. Blood is tested for antibodies using a novel ELISA for immunoglobulins IgG, based on tagged and purified recombinant SARS-CoV-2 trimeric spike protein. Antibody binding to the S protein is detected with ALP-conjugated anti-human IgG. Residual blood samples will be stored by the University of Oxford after testing.

More information about swab and blood sample procedure and analysis can be found in the [study protocol](#).

5 . Test sensitivity and specificity

Understanding false-positives and false-negative results

The estimates provided in Section 2 of the [Coronavirus \(COVID-19\) Infection Survey pilot bulletin](#) are for the percentage of the private-residential population testing positive for the coronavirus (COVID-19), otherwise known as the positivity rate. We do not report the prevalence rate. To calculate the prevalence rate, we would need an accurate understanding of the swab test's sensitivity (true-positive rate) and specificity (true-negative rate).

While we do not know the true sensitivity and specificity of the test because COVID-19 is a new virus, our data and related studies provide an indication of what these are likely to be. To understand the potential impact, we have estimated what prevalence would be in two scenarios using different possible test sensitivity and specificity rates.

Test sensitivity

Test sensitivity measures how often the test correctly identifies those who have the virus, so a test with high sensitivity will not have many false-negative results. Studies suggest that [sensitivity may be somewhere between 85% and 98%](#).

Our study involves participants self-swabbing under the supervision of a study healthcare worker. It is possible that some participants may take the swab incorrectly, which could lead to more false-negative results. However, research suggests that [self-swabbing under supervision is likely to be as accurate as swabs collected directly by healthcare workers](#).

Test specificity

Test specificity measures how often the test correctly identifies those who do not have the virus, so a test with high specificity will not have many false-positive results.

We know the specificity of our test must be very close to 100% as the low number of positive tests in our study means that specificity would be very high even if all positives were false. For example, in the most recent six-week period (31 July to 10 September), 159 of the 208,730 total samples tested positive. Even if all these positives were false, specificity would still be 99.92%.

We know that the virus is still circulating, so it is extremely unlikely that all these positives are false. However, it is important to consider whether many of the small number of positive tests we do have might be false. There are a couple of main reasons we do not think that is the case.

Symptoms are an indication that someone has the virus; therefore, if there are many false-positives, we would expect to see more false-positives occurring among those not reporting symptoms. If that were the case, then risk factors such as working in health care would be more strongly associated with symptomatic infections than with asymptomatic infections. However, in our data the risk factors for testing positive are equally strong for both symptomatic and asymptomatic infections.

The percentage of individuals reporting no symptoms among those testing positive has remained stable over time despite substantial declines in the overall number of positives. If false-positives were high, the percentage of individuals not reporting symptoms among those testing positive would increase when the true prevalence is declining.

More information on sensitivity and specificity is included in [Community prevalence of SARS-CoV-2 in England: Results from the ONS Coronavirus Infection Survey Pilot](#) by the Office for National Statistics' academic partners.

The impact on our estimates

We have used Bayesian analysis to calculate what prevalence would be in two different scenarios, one with medium sensitivity and the other with low sensitivity. Table 1 shows these results alongside the weighted estimate of the percentage testing positive in the period from 28 August to 10 September.

Scenario 1 (medium sensitivity, high specificity)

Table 1: The effects of test sensitivity on estimates

Reference period: 28 August to 10 September	95% credible interval	
	Lower	Upper
Estimated average percentage of the population who had COVID-19 (weighted)	0.13 0.10	0.17
Prevalence rate in Scenario 1 (medium sensitivity, high specificity)	0.13 0.08	0.18
Prevalence rate in Scenario 2 (low sensitivity, high specificity)	0.20 0.12	0.30

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

Based on similar studies, the sensitivity of the test used is plausibly between 85% and 95% (with around 95% probability) and, as noted earlier, the specificity of the test is above 99.9%.

Scenario 2 (low sensitivity, high specificity)

To allow for the fact that individuals are self-swabbing, Scenario 2 assumes a lower overall sensitivity rate of 60% (or between 45% and 75% with 95% probability), incorporating the performance of both the test itself and the self-swabbing. This is lower than we expect the true value to be for overall performance but provides a lower bound.

The results show that when these estimated sensitivity and specificity rates are taken into account, the prevalence rate would be slightly higher but still very close to the main estimate presented in Section 2 of the [Coronavirus \(COVID-19\) Infection Survey pilot bulletin](#). This is the case even in Scenario 2, where we use a sensitivity estimate that is lower than we expect the true value to be. For this reason, we do not produce prevalence estimates for every analysis, but we will continue to monitor the impacts of sensitivity and specificity in future.

[Evaluation](#) of the test sensitivity and specificity of five immunoassays for SARS-CoV-2 serology, including the ELISA assay used in our study, has shown that this assay has sensitivity and specificity (95% confidence interval) of 99.1% (97.8 to 99.7%) and 99.0% (98.1 to 99.5%) respectively; compared with 98.1% (96.6 to 99.1%) and 99.9% (99.4 to 100%) respectively for the best performing commercial assay. We periodically provide estimates of positivity rates, which adjust for imperfect test sensitivity and specificity.

6 . Linking survey data, swab results and blood samples

Each participant in a household that agrees to participate are provided with an individual identifier (for example; A, B, C and so on). This allows for the differentiation of data collected between each household member.

Participants are also uniquely identified by their month and year of birth (not personally identifiable). This information is needed in order to match survey responses with results from the swabs and blood samples. Swabs are labelled with the unique household study code, date of sampling, the participant suffix and their month and year of birth. The blood sample is also given a barcode.

7 . Cleaning the data

As in any survey, some data can be incorrect or missing. For example, participants and interviewers sometimes misinterpret questions or skip them by accident. It is important to run a pilot before running a full survey, so that the survey instrument can be improved. To minimise the impact of incorrect or missing data, we clean the data, by editing or removing things that are clearly incorrect.

In the pilot survey data, we identified some specific quality issues with the healthcare and social care worker question responses and have therefore applied some data editing (cleaning) to improve the quality. Cleaning will continue to take place to further improve the quality of the data on healthcare and social care workers, which may lead to small revisions in future releases.

8 . How we analyse the data

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

The analysis of the data is a collaboration between the Office for National Statistics (ONS) and researchers from the University of Oxford and University of Manchester, Public Health England and Wellcome Trust. Our academic collaborators aim to publish an extended account of the modelling methodology outside the ONS bulletin publication in peer-reviewed articles in due course.

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 this analysis. Estimates may therefore be revised as more test results are included.

This is a pilot study where the analysis is developed at pace and these quality enhancements may lead to minor changes in estimates, for example, the positive test counts across the study period.

9 . Modelling

Modelling to estimate the headline rate of people testing positive for SARS-CoV-2, the virus that causes the coronavirus (COVID-19) disease, and incidence rate is conducted by our research partners at the University of Oxford.

The study is based on a nationally representative survey sample; however, some individuals in the original Office for National Statistics (ONS) survey samples will have dropped out and others will not have responded to the pilot. To address this and reduce potential bias, the regression models adjust the survey results to be more representative of the overall population in terms of age, sex and region. They do not adjust for household tenure or household size. This technique is known as dynamic Bayesian multi-level regression post-stratification (MRP) and is used by organisations such as the Centers for Disease Control and Prevention (CDC) to provide prevalence of diseases at both a national and subnational level in the United States.

As the number of people testing positive (known as the positivity rate) and incidence rates are unlikely to follow a linear trend, time measured in days is included in the model using a non-linear function (thin-plate spline). Time trends are allowed to vary between regions by including an interaction between region and time. Given the low number of positive cases, the effect of time is not allowed to vary by other factors.

The modelled estimates use a non-linear function for time to better fit to the most recent data and to allow for likely departures from a linear trend in the future. To be able to do this, time is measured in days.

The model for the positivity rate uses all available swab data from participants within time periods to estimate the number of people who currently have SARS-CoV-2. A Bayesian multi-level generalised additive model with a complementary log-log link was used.

The generalised additive Poisson model for incidence considers every day that each participant is in the study from the date of their first negative test to the earlier of their latest negative test or first positive test, which are called days at risk (for a new infection). Each new infection is counted as a positive at the mid-point between the day of the test and the previous negative swab or at seven days before the sample, whichever is closest to the most recent positive swab. This is because we do not know the exact point when the infection occurred and infections only last so long. We exclude from this analysis everyone whose first swab test in the study was positive, so this analysis just looks at new infections acquired during the study. Note, only data from 11 May onwards are included in the incidence rate model, as to be included in the incidence analysis at least two repeated swab test visits are required.

The data that are modelled are drawn from a sample, and so there is uncertainty around the estimates that the model produces. Because a Bayesian regression model was used, we present estimates along with credible intervals. These 95% credible intervals can be interpreted as - provided that the model is correct - there being a 95% probability that the true value being estimated lies within the credible interval. Again, a wider interval indicates more uncertainty in the estimate.

Our research partners have published a [pre-print article](#) with more detail about the methodology behind the modelling.

Previous methodology used for calculating the incidence rate

Estimates for the incidence rates in the bulletins on 25 June and 9 July were calculated using a different method, which is explained in the following.

The model for incidence considered every day that each participant is in the study from the date of their first negative test to the earlier of their latest negative test or first positive test, which are called days at risk (for a new infection). Each new infection is counted as a positive on the day of the test.

We used a Poisson model to estimate the rate of new infections by the number of days at risk. Because of few new infections, counting the positive as happening at a different time, for example halfway between the last negative and first positive test, does not change the results. The modelled estimates use a non-linear function for time called a natural cubic spline to better fit to the most recent data and to allow for likely departures from a linear trend in the future.

We started recruiting participants on 26 April and started repeating tests on 1 May. Our analysis therefore starts at 11 May, as before this too few participants had repeated test results to produce robust estimates of incidence over time. Because of small numbers of new infections, the model is not weighted. We exclude from this analysis everyone whose first swab test in the study was positive, so this analysis just looks at new infections acquired during the study period.

The household level incidence analysis tries to estimate the rate of new infections being imported into a household and to remove the effect of people passing the infection on within a household. For this, we still count new infections in individual people in each household, but we only include the first new infection within a household. We also exclude households where anyone in the household was positive at their first test in the study.

10 . 14-day estimates

In addition to our headline modelled estimates, we provide weighted estimates of the positivity rate and unweighted estimates of the incident rate in 14-day non-overlapping time periods.

Weighting

The 14-day estimates of the number of people who have the coronavirus (COVID-19) are based on weighted data to ensure the estimates are representative of the target population in England. The pilot is based on a nationally representative survey sample; however, some individuals in the original Office for National Statistics (ONS) survey samples will have dropped out and others will not have responded to the pilot. To address this and reduce potential bias, we apply weighting to ensure the responding sample is representative of the population in terms of age (grouped), sex, region, housing tenure and household size. This is different from the modelled estimates, which adjust for potential non-representativeness of the sample through multi-level regression post-stratification (described in [Section 9: Modelling](#)).

The 14-day incidence rate is based on unweighted data because the sample size of positive cases is small.

Confidence intervals for 14-day estimates

The statistics are based on a sample, and so there is uncertainty around the estimate. [Confidence intervals](#) 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.

Confidence intervals for weighted estimates are calculated using the Korn-Graubard method to take into account the expected small number of positive cases and the complex survey design. For unweighted estimates, we use the Clopper-Pearson method as the Korn-Graubard method is not appropriate for unweighted analysis.

11 . Statistical testing

Where we have done analysis of the characteristics of people who have ever tested positive for the coronavirus (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. Fisher's exact test was used to determine whether the differences were compatible with chance given the numbers sampled.

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 if there truly is no difference between the groups in the population. We used 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 and 0.05 are considered to provide relatively strong and moderate evidence of difference between the groups being compared respectively.

Any estimate based on a random sample rather than an entire population contains some uncertainty. Given this, it is inevitable that sample-based estimates will occasionally suggest some evidence of difference when there is in fact no systematic difference between the corresponding values in the population as a whole. Such findings are known as "false discoveries". If we were able to repeatedly draw different samples from the population, then, for a single comparison, we would expect 5% of findings with a p-value below a threshold of 0.05 to be false discoveries. However, if multiple comparisons are conducted, as is the case in the analysis conducted within the Infection Survey, then the probability of making at least one false discovery will be greater than 5%. Multiplicity can occur at different levels. For example, in the Infection Survey we have:

- two primary outcomes of interest – positivity for current infection based on a swab test and positivity for previous infection based on a blood test
- several different exposures of interest (for example, age and sex)
- several exposures with multiple different categories (for example, working location)
- repeated analysis over calendar time.

Consequently, the p-values use in our analysis have not been adjusted to control either the familywise error rate (FWER, the probability of making at least one false discovery) or the false discovery rate (FDR, the expected proportion of discoveries that are false) at a particular level. Instead, we focus on presenting the data and interpreting results in the light of the strength of evidence that supports them.

12 . Analysis feeding into R

The statistics produced by analysis of this survey contribute to modelling, which predicts 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.

13 . Uncertainty in the data

The estimates presented in this bulletin 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)

These results are directly from the test, and no test is perfect. There will be false-positives and false-negatives from the test, and false-negatives could also come from the fact that participants in this study are self-swabbing. More information about the potential impact of false-positives and false-negatives is provided in 'Sensitivity and Specificity analysis'

The data are based on a sample of people, so there is some uncertainty in the estimates

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.

Quality of data collected in the questionnaire

As in any survey, some data can be incorrect or missing. For example, participants and interviewers sometimes misinterpret questions or skip them by accident. To minimise the impact of this, we clean the data, editing or removing things that are clearly incorrect. In these initial data, we identified some specific quality issues with the healthcare and social care worker question responses and have therefore applied some data editing (cleaning) to improve the quality. Cleaning will continue to take place to further improve the quality of the data on healthcare and social care workers, which may lead to small revisions in future releases.