

COVID-19 Schools Infection Survey: 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 Schools Infection Survey (SIS).

Contact:

schools.infection.survey@ons.gov.uk
+44 (0)20 8039 0326

Release date:

11 August 2021

Next release:

To be announced

Table of contents

1. [COVID-19 Schools Infection Survey](#)
2. [Study design: the sample](#)
3. [Study design: data we collect](#)
4. [Timing](#)
5. [Participation rates](#)
6. [Weighting](#)
7. [Seroconversion](#)
8. [Linking survey data and biological samples](#)
9. [Linking COVID-19 Schools Infection Survey data to vaccination data](#)
10. [Test sensitivity and specificity](#)
11. [Uncertainty in the data](#)

1 . COVID-19 Schools Infection Survey

The “COVID-19 Schools Infection Survey” is jointly led by the Office for National Statistics (ONS), London School of Hygiene and Tropical Medicine (LSHTM) and Public Health England (PHE).

The COVID-19 Schools Infection Survey aims to investigate the prevalence of current coronavirus (COVID-19) infection and presence of antibodies to COVID-19 among pupils and staff in sampled primary and secondary schools in England.

Repeated surveys are being carried out to collect risk factor information together with virus and antibody samples in a cohort of pupils and staff. Antibody conversion and viral prevalence at points in the academic year are key outcome measures.

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 Schools Infection Survey. We will continue to expand and develop methods as the study progresses.

This methodology guide can be read alongside:

- the [COVID-19 School Infection Survey](#), which gives headline statistics
- the [study guide](#), which explains to participants what taking part in the study entails

2 . Study design: the sample

The COVID-19 Schools Infection Survey (SIS) has a stratified, multi-staged sample design. Strata are formed by a cross-classification of prevalence (each local authority being classified as either “high” or “low” prevalence according to coronavirus (COVID-19) rates at the week 2 September to 8 September 2020) and school type (primary or secondary). The first stage of sampling was the selection of local authorities, and it is from each selected local authority that samples of schools (primary and secondary) have been drawn.

The following schools were excluded from the sampling frame:

- special schools, independent schools, pupil referral units and further education colleges
- schools taking part in other school-based COVID-19 studies

Sampling of upper-tier local authority areas in England

The study oversampled schools in “high prevalence” areas of the country. High prevalence upper-tier local authorities (N=30) were defined as local authorities in the top 20% when ranked by the rate of confirmed cases of COVID-19 infection (100,000 population from Pillar 2 testing in the week 2 September to 8 September 2020). Low prevalence upper-tier local authorities (N=119) were defined as local authorities in the bottom 80% when ranked by the rate of confirmed cases of COVID-19 infection (100,000 population from Pillar 2 testing in the week 2 September to 8 September 2020).

Ten upper-tier local authorities were randomly sampled from the “high prevalence” group and five from the remaining “low prevalence” upper-tier local authorities.

High prevalence

Bradford, Gateshead, Knowsley, Lancashire, Leicester, Liverpool, Manchester, Salford, Sunderland and Warrington.

Low prevalence

Barking and Dagenham, Bournemouth, Christchurch and Poole, Norfolk, Reading, and Redcar and Cleveland.

Sampling of schools

The aim of the study was to recruit 100 secondary schools and 50 primary schools across the 15 selected upper-tier local authorities, with approximately 70% (70 secondary and 35 primary) schools in high-prevalence areas and 30% (30 secondary and 15 primary) in low-prevalence areas.

Within each of the four strata (High-prevalence Primary, High-prevalence Secondary, Low-prevalence Primary, Low-prevalence Secondary), a further stratification by upper-tier local authority (of the ones selected in the first stage of sampling) has been applied. The aim has been to achieve an equal number of schools in each upper-tier local authority within the prevalence-by-type stratum.

To compensate for non-response and refusal we selected 250 schools to be approached.

Practical modifications

Several practical modifications were made to the sample selection procedure because of practical constraints on the selection of schools:

- the number of academy trusts (which manage some schools, the rest being managed by the local authorities themselves) that could be selected in each upper-tier local authority was limited to a maximum of four across primary and secondary school types combined; this was because limited capacity to approach academy trusts for enrolment
- once the cap on academy trusts was reached, no schools from other academy trusts could be included in the sample; this amendment has not been explicitly factored into the calculation of inclusion probabilities or design weights, but the later calibration applied should help mitigate the effects of this
- changes were also made with respect to one of the selected local authorities; we stopped recruiting schools run by that local authority as the local authority decided to withdraw their managed schools from the study; an attempt to mitigate this was made by sampling more academic trust-run schools within the upper-tier local authority
- lower than anticipated response rates have seen the school samples expanded; additional schools were selected in November, and a further sample within particular upper-tier local authorities was drawn in February
- we have attempted to ensure there are at least three responding schools of each type (primary or secondary) in each upper-tier local authority to increase the reliability of individual local authority comparisons at the data publication stage; this criterion was not met in five upper-tier local authorities (Norfolk, Lancashire, Bournemouth, Leicester and Sunderland) in the achieved sample for Rounds 1 and 2

Sampling of individuals: staff and pupils

Within the selected schools, primary and secondary, all staff were eligible and invited to participate in the study. Within primary schools, all pupils were eligible to participate, however, because of the larger number of pupils in secondary schools, eligibility was restricted to two consecutive year groups in each secondary school. Year groups in secondary schools were chosen at random and in equal proportions across the schools and local authorities. Low response however, in Rounds 1 and 2, has led to a decision to widen eligibility to pupils in all year groups (except Year 11) within secondary schools from Round 4 onwards. In Round 6, 63 out of the 80 participating secondary schools had extended participation to other year groups.

Pupils from Year 11 are not eligible for enrolment. It was deemed that this study would be too disruptive for these pupils during their final year of secondary school.

3 . Study design: data we collect

In each school that agreed to participate, head teachers were asked to register and complete a short questionnaire. Head teachers were also provided with information about the survey to forward to staff, parents of pupils aged under 16 years old, and pupils aged 16 years old or over. After completing a consent form, participants (or their parent if the pupil was under 16) will be asked to complete a short online “enrolment” questionnaire. A questionnaire collecting further and more-detailed information will be delivered to participants following each round of current coronavirus (COVID-19) infection and SARS-CoV-2 (COVID-19) antibody tests.

A study team visited each school to collect the biological samples for testing from the staff and pupils who had enrolled in the study. Tests for pupils involved a nose swab for current coronavirus (COVID-19) infection, and an oral fluid (saliva) sample for SARS-CoV-2 (COVID-19) antibodies against the virus. Tests for staff involved a nose swab for current COVID-19 infection and a finger prick blood test for COVID-19 antibodies against the virus. Everyone enrolled was offered testing regardless of whether they were experiencing COVID-19 symptoms, although people experiencing COVID-19-like symptoms should not be attending school.

For each subsequent round of testing, participants receive advance notification of the date of the sample collection day, with a short follow-up questionnaire.

4 . Timing

The first round of testing took place between 3 November and 20 November 2020. The second round of testing took place between 30 November and 11 December 2020, the fourth round of testing took place between 15 March and 31 March 2021, the fifth round of testing took place between 5 May and 21 May 2021, and the sixth round of testing took place between 14 June and 6 July 2021. Only those who had enrolled in the study and were in school on the day of testing were tested. This means those with coronavirus (COVID-19) symptoms and those instructed to self-isolate would not be present in the school building to be tested on the assigned test day. Those absent from school on the day of testing received a testing kit, sent to their house so that they could take part in that round of the study.

The closure of schools during the lockdown from 5 January 2021 meant that the third round of the COVID-19 Schools Infection Survey was cancelled. However, anyone who had enrolled in the study but had not had an antibody test was offered a home testing kit.

5 . Participation rates

Recruitment of schools to the study began on 12 October 2020. At the time of the June 2021 testing period, the sample included 55 primary schools, 84 secondary schools and 2 all-through schools (which were included with secondary schools for the purpose of the sample selection) across the 15 sampled local authorities. Estimated participation rates for each round of testing, by participant type, are given in Tables 1 to 5. In total, across the different rounds of testing, 23,867 (8,038 staff and 15,829 pupils) participated in at least one COVID-19 current infection or antibody test.

Table 1: Response rates for participants in Round 1 by participant and school type

School type	Participant type	Estimates number eligible* round 1	Number of participants in round 1*	% of eligible
Primary				
	Staff	2,200	1,166	53%
	Pupils	13,700	2,248	16%
Secondary				
	Staff	7,200	3,171	44%
	Pupils	22,300	3,147	14%
Totals				
	Staff	9,400	4,337	46%
	Pupils	36,000	5,395	15%

Source: Office for National Statistics - COVID-19 Schools Infection Survey

Table 2: Response rates for participants in Round 2 by participant and school type

School type	Participant type	Estimates number eligible* round 2	Number of participants in round 2*	% of eligible
Primary				
	Staff	2,200	1,181	54%
	Pupils	13,500	2,511	19%
Secondary				
	Staff	9,500	3,933	41%
	Pupils	28,000	4,579	16%
Totals				
	Staff	11,700	5,114	44%
	Pupils	41,500	7,089	17%

Source: Office for National Statistics - COVID-19 Schools Infection Survey

Table 3: Response rates for participants in Round 4 by participant and school type

School type	Participant type	Estimates number eligible* round 4	Number of participants in round 4*	% of eligible
Primary				
	Staff	2,800	1,410	50%
	Pupils	17,200	3,762	22%
Secondary				
	Staff	10,500	2,744	26%
	Pupils	53,500	7,271	14%
Totals				
	Staff	13,300	4,154	31%
	Pupils	70,700	11,033	16%

Source: Office for National Statistics - COVID-19 Schools Infection Survey

Table 4: Response rates for participants in Round 5 by participant and school type

School type	Participant type	Estimates number eligible* round 5	Number of participants in round 5*	% of eligible
Primary				
	Staff	2,800	1,491	53%
	Pupils	17,200	4,369	25%
Secondary				
	Staff	10,300	3,548	34%
	Pupils	52,100	9,026	17%
Totals				
	Staff	13,100	5,039	38%
	Pupils	69,300	13,395	19%

Source: Office for National Statistics - COVID-19 Schools Infection Survey

Table 5: Response rates for participants in Round 6 by participant and school type

School type	Participant Type	Estimates number eligible* round 6	Number of participants in round 6*	% of eligible
Primary				
	Staff	3,100	1,318	43%
	Pupils	17,600	4,243	24%
Secondary				
	Staff	11,000	2,505	23%
	Pupils	49,900	8,242	17%
Totals				
	Staff	12,900	3,819	30%
	Pupils	67,400	12,485	19%

Source: Office for National Statistics – COVID-19 Schools Infection Survey

Notes

1. Number of eligible *: all primary and secondary school staff; all primary pupils; and secondary pupils in year groups selected for participation estimated from the 2020 to 2021 academic year school census and school workforce census collected by Department for Education.
2. Participation is defined as any person who provided at least one coronavirus current infection or SARS-CoV-2 antibody test.
3. All-through schools; participants from all-through schools have been spilt into primary and secondary school type according to the school year of the pupil, or the year(s) the staff member interacts with most.
4. In Round 4 participating secondary schools extended pupil participation to other year groups and this is reflected in an increase in the estimated number eligible for Round 4.

6 . Weighting

Weighting is applied to the data collected from responding schools, pupils and staff to make the data representative of the wider, target population of the local authority from which they have been drawn. The weighting takes account of the design of the sample and reflects the response patterns observed and the total numbers of staff and pupils in schools and the local authorities selected.

Accounting for response patterns in the weighting is particularly important. Response rates to the study may differ between various subgroups of the eligible population, and if this response propensity is correlated with the study outcomes, positivity rates and other estimates will be biased if they are computed from the unweighted, observed data. It is important to note that weighting can only be carried out to adjust for observed biases in response rates. There may be other unobserved biases that impact on an individual's likelihood of taking part, which cannot be controlled for by the weights calculated.

Generally, our aim of weighting the Schools Infection Survey data has been to achieve representativeness at the individual local authority level. More specifically, separate sets of weights have been computed for each local authority, participant group (pupils, staff), school type (primary or secondary schools), type of test (antibody or current infection) and target population (local authority level-enrolled pupil population for pupils and local authority level school staff).

For each of these sets, weighting has been performed as a single calibration step with uniform input weights, which reflect the sample design within each local authority. After calibration, the weights are assigned to particular subgroups of the sample and sum to prespecified population totals. The following calibration groups were used:

- sex (male, female)
- ethnicity (White British, not White British)
- age group for staff (under 25 years, 25 to 29, 30 to 39, 40 to 49, 50 to 59, 60 years and over)
- year group for pupils (Reception and years 1 to 2, years 3 to 4, years 5 to 6, years 7 to 8, years 9 to 10, years 12 to 13)

Because of the lack of data for the current academic year, totals for the local authority level pupil and staff populations were computed from the 2019 to 2020 school census tables [published by the Department of Education](#) and applied without further correction. In several cases, calibration groups had to be combined or omitted because of low counts in certain categories.

For calibration to staff and pupil totals, the effect of weighting on the positivity estimates was found to be moderate in most cases. This implies that either the amount of oversampling or under-sampling with respect to calibration groups has usually been limited, or that the positivity estimates are similar across calibration groups. For example, for staff antibody testing in Rounds 1 and 2, the absolute difference between weighted and unweighted estimated prevalence rates was at most 0.5 percentage points (pp) in 24 out of 51 cases and at most 1 pp in a further 9. The largest difference was 4.1 pp. For Waves 4 and 5 the corresponding figures are 15 out of 56 cases below 0.5 pp and a further 10 below 1 pp; the largest difference was 12.6 pp.

Weighting for seroconversion rates was created in June 2021 and published within the Round 5 bulletin. This weighting for seroconversion rates was calculated for participants with a negative test result in the first round of test and a positive or negative test result in the second round of testing. This weighting consists of a singular calibration step allied to all participants with a negative test in the first round of testing, using our standard calibration groups and calibration totals.

7 . Seroconversion

In the case of the coronavirus (COVID-19), seroconversion is the incidence of antibody test results changing from negative to positive and will capture both symptomatic and asymptomatic infections that may have been missed between testing rounds. During an infection, antigens enter the blood, and the immune system begins to produce antibodies in response. In the testing of pupils for COVID-19 antibodies, we are using oral fluid tests as an indicator of blood antibodies.

Antibodies remain in the blood at low levels after infection, although these levels can decline over time to the point that tests can no longer detect them. The length of time antibodies remain at detectable levels in the blood is not fully known.

To account for the different follow-up times between the rounds, the seroconversion rate has been expressed per 1,000 person-weeks. This calculation takes the number of participants who seroconverted between two testing rounds and divides this by the sum of the weeks that each participant had between the two rounds of testing.

Antibodies are also produced when a person is vaccinated, however, the type of antibody tests used in this study only detects the N-protein antibody, which is produced following natural infection not after vaccination.

The seroconversion figures in our analysis do not control for other factors that may affect the probability of seroconverting; for example, different distributions between local authorities of those sampled, and different age distributions.

8 . Linking survey data and biological samples

Information collected from each participant who agreed to take part is anonymised. An individual serial number or identifier (ParticipantID) is used. This allows for the differentiation of data collected between each pupil or staff member. Each ParticipantID is linked to their school by the school's unique reference number.

The biological samples are given a barcode and this barcode is also recorded against the ParticipantID by the study team. This allows the test results to be matched to the correct individual. Personal identifiers (for example, name) are not used to link the data.

9 . Linking COVID-19 Schools Infection Survey data to vaccination data

Information collected from each participant was linked to the vaccination data from the National Immunisation Management System (NIMS).

Participants in the COVID-19 Schools Infection Survey (SIS) were matched to the NHS Personal Demographic Service (PDS) to attach the individual's NHS number. This then enabled the SIS participants to be linked to the vaccination data via the individual's NHS number.

Vaccination records can only be attached if an NHS number is found. In our latest round of matching we found NHS numbers from the PDS for approximately 94% of staff that participated in Round 4. We are continuing to refine our matching process to improve the match rate between SIS and the PDS. Only staff with an NHS number are included in the denominator when working out the proportion of staff having received at least one dose of vaccine.

10 . Test sensitivity and specificity

The coronavirus (COVID-19) infection estimates provided in the COVID-19 School Infection Survey bulletin are the percentage of the school-based population testing positive for current COVID-19 infection on the day of testing. The proportion testing positive for current COVID-19 infection should not be interpreted as being the prevalence rate. To calculate prevalence rates, we would need an accurate understanding of the swab test's sensitivity (true-positive rate) and specificity (true-negative rate).

Results calculated as the proportion of individuals with a positive result can be adjusted to account for the tests specificity and sensitivity using:

$$p = (q + \textit{specificity} - 1) / (\textit{sensitivity} + \textit{specificity} - 1)$$

where p is the adjusted proportion positive and q is the observed proportion positive.

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% \(PDF, 1.82MB\).](#)

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 can assume that the specificity of our test must be very close to 100% because the numbers of positive tests in our study is low, meaning that specificity would be very high if all positives were false. 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 any of the small number of positive tests were false-positive.

Type of tests

The nasal swabs were sent to one of the UK's national laboratories for COVID-19 detection using an accredited reverse transcriptase polymerase chain reaction (RT-PCR) test. This assay has been shown to have a 70% sensitivity and a high 95% specificity.

Capillary blood samples from staff were collected and tested using a commercial immunoassay for antibodies against SARS-CoV-2 (Roche cobas® Elecsys Anti-SARS-CoV-2 assay). The assay has been shown to have a [high sensitivity \(97.2%\) and specificity \(99.8%\).](#)

Oral fluid samples from students were collected and sent for detection of antibodies against the SARS-CoV-2 Nucleoprotein (NP) using an Immunoglobulin G (IgG)-capture-based enzyme immunoassay (EIA). The assay has been shown to have 80% sensitivity and 99% specificity. Although the presence of immunoglobulins in oral fluids are in concentrations of at least 1 per 1,000th of that found in blood, the reactivity of salivary immunoglobulins mirrors that of serum. Therefore, oral fluids are an attractive non-invasive alternative sample to blood, particularly in children.

11 . Uncertainty in the data

The estimates presented in the Schools Infection Survey statistical bulletin are subject to [uncertainty](#). There are many causes of uncertainty, but the main sources of uncertainty in the analysis and data presented include each of the following.

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

These results derive directly from the tests, and no test is perfect: there will be false-positives and false-negatives from the tests. In addition, false-negatives could also arise because participants in this study are self-swabbing and some may not produce a sample that can provide a conclusive result.

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

Any estimate based on a sample contains some uncertainty as to whether it reflects the broader population of interest because of its smaller sample size. A [confidence interval](#) gives an indication of the degree of uncertainty of an estimate, showing the precision of a sample estimate. The 95% confidence intervals are calculated so that if we repeated the study many times, 95% of the time the proportion testing positive would lie between the lower and upper confidence limits. A wider interval indicates more uncertainty in the estimate. Overlapping confidence intervals indicate that there may not be a true difference between two estimates.

Pupils and staff who chose to enrol in the study may be different to those who do not enrol

As well as uncertainty, samples can be affected by non-response bias. This can occur when there is a systematic difference between those who take part in the study and those who do not, meaning participants are not representative of the study population. If this difference is also associated with the likelihood of contracting the coronavirus (COVID-19) then the estimates produced from the data collected cannot be generalised to the study population as a whole. Weighting is used to compensate for non-response.