

COVID-19 and Respiratory Infections Survey: QMI

Quality and Methodology Information (QMI) for the COVID-19 and Respiratory Infections Survey (CRIS), detailing strengths and limitations, methods, and data uses

Contact:

Dr Rhiannon Yapp, Astrid Dawes
and Benedetta Iametti
Health.Data@ons.gov.uk
+44 1633 656671

Release date:

10 July 2023

Next release:

To be announced

Table of contents

1. [Methodology background](#)
2. [About this Quality and Methodology Information report](#)
3. [Quality summary](#)
4. [Quality characteristics of the COVID-19 and Respiratory Infections Survey](#)
5. [Methods used to produce the data](#)
6. [Glossary](#)
7. [Related links](#)
8. [Cite this methodology](#)

1 . Methodology background

- Survey Name: COVID-19 and Respiratory Infections Survey (CRIS)
- Time period: The main CRIS survey was run throughout May and June 2023. A pilot survey was initially run in April 2023 to understand user experience, but findings from this have not been incorporated into our CRIS analysis or publications.
- How compiled: Estimates were derived from questionnaire responses of CRIS participants who were initially enrolled on the Office for National Statistics (ONS) Coronavirus (COVID-19) Infection Survey (CIS), which had been commissioned by the UK Health Security Agency (UKHSA).
- Geographic coverage: UK wide with selected analysis broken down to the 4 UK nations, and regions in England.
- Participants: Identified as eligible for CRIS if they had completed a CIS questionnaire in the past 90 days and had consented to further research. Further information on CIS participants is detailed in the [Coronavirus \(COVID-19\) Infection Survey quality report: December 2022](#).
- Number of participants: Approximately 330,000 participants were identified as being eligible for CRIS and invited to the survey.
- Achieved sample size: Throughout May 2023 and June 2023 156,261 responses were collected. Some participants responded twice, approximately 35 days apart.

2 . About this Quality and Methodology Information report

This quality and methodology report contains information on the quality (including the [European Statistical System \(PDF, 3.0 MB\)](#) five dimensions of quality) of the statistics produced as outputs from the COVID-19 and Respiratory Infections Survey (CRIS). The information presented is for data collected via a self-completed questionnaire, up until 12 June 2023. The methods used to create CRIS statistical outputs are also detailed. The information in this report will help you to:

- understand the strengths and limitations of CRIS statistics
- reduce the risk of misusing data
- help you to decide suitable uses for the data
- understand the methods used to create the data
- increase overall understanding of our data and help improve future pandemic preparedness

3 . Quality summary

Important points about the COVID-19 and Respiratory Infections Survey (CRIS)

CRIS was set up in April 2023 to:

- continue surveillance of COVID-19 as part of a Population Health Monitoring programme
- monitor the impact of self-reported COVID-19, long COVID and other respiratory infections on the lives of individuals, the community and on health services and how these are changing
- assess potential pressures to help support the NHS and other services to prepare for future stressors, acting as an early warning system

Only private residential households and their residents aged 16 years and over were included in the survey. People in hospitals, care homes and/or other communal settings were not included.

This survey was undertaken by the Office for National Statistics (ONS). We at the ONS collect information about the UK's society and economy, which provides evidence for policy- and decision-making, and for directing resources to where they are needed most.

Overview of CRIS

The main CRIS survey was launched in May 2023, across the UK. A small pilot survey was run in April 2023 to understand user experience, but responses to the pilot survey have not been incorporated into our main analyses or publications.

The CRIS sample was drawn from participants of the Coronavirus (COVID-19) Infection Survey (CIS). Eligibility to participate in CRIS was established if a CIS participant had both:

1. completed a CIS questionnaire in the past 90 days
2. consented to further research

All eligible participants aged 16 years and over were sent a CRIS survey invite by email or letter, depending on their communication preferences in the CIS. This invited participants to complete the CRIS questionnaire online or by telephone. The ONS processed and sent all letter communications. GOV.UK Notify, a service offered by the Cabinet Office, sent email communications on our behalf.

The CRIS survey questionnaire was designed to understand the impact of self-reported COVID-19, long COVID and other respiratory infections on people's lives, the community and health services. Questions included:

- general health, including any recent symptoms or vaccinations against COVID-19 and flu
- how much a person's health affected usual activities or use of health services
- work, or education

In many cases, more than one participant from a single household participated in CRIS. This was because the CIS was based initially on a random sample of households to provide a nationally representative survey.

Some of the CRIS participants who responded to the initial invite and completed the questionnaire or telephone interview, were sent a second invite approximately 35 days later.

In contrast to the CIS, no bio-samples (swabs or bloods) were undertaken as part of CRIS. In addition, no financial incentives were offered for participation.

Uses of CRIS

CRIS provides important information about the characteristics of people and households who self-reported COVID-19, long COVID or other respiratory 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. This will help UK governments understand the impact of self-reported COVID-19 infections, long COVID and other respiratory infections on life, the community, and health services. It could further assist with service planning and vaccination roll out.

The data can be used for:

- estimating the number and proportion of self-reported respiratory infections in the community
- identifying differences in numbers of self-reported cases of respiratory illness and changes in them over time
- identifying characteristics associated with self-reporting symptoms consistent with influenza-like illnesses (ILI)
- identifying the most common symptoms associated with [long COVID](#) experienced by CRIS participants

The data cannot be used for:

- estimating the number of respiratory illness cases or their prevalence in care homes, hospitals and/or other communal settings
- providing information about recovery times for those infected
- accurately identifying the type of respiratory infection reported
- estimating the prevalence of long COVID in the general population
- estimating asymptomatic infections

Strengths and limitations

Some of CRIS' main strengths include:

- a large sample of participants
- high levels of participant engagement
- the collection of data from all 4 UK nations
- a questionnaire which examines various symptoms and characteristics across all age groups

An additional strength of our survey was that all participants were a subset of the CIS sample. This enabled us to collect and maintain longitudinal data which we could link back to for COVID-19 monitoring, alongside examining other respiratory infections and vaccination histories. Notably, in CRIS, any new episodes of COVID-19 and respiratory infections were self-reported as no biological samples were collected or processed as part of this survey, in turn reducing cost. This may mean that some participants reported having COVID-19, long COVID or a respiratory infection, when they may have not tested positive for such infections. Also, the symptoms identified as being part of a current illness may not be specific to that illness.

In addition, as with all surveys, the CRIS sample is subject to possible bias. More so, as the CRIS study population is a subset of the CIS population and selection for CRIS was driven by underlying biases within the CIS cohort. As such, it was observed within CRIS, that those with long COVID were more likely to respond to the survey than those not reporting long COVID. This created challenges for producing long COVID prevalence estimates.

All estimates presented in our publications 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. For further information on uncertainty, please see [Section 4. Quality characteristics of the COVID-19 and Respiratory Infections Survey](#).

4 . Quality characteristics of the COVID-19 and Respiratory Infections Survey

Relevance

The COVID-19 and Respiratory Infections Survey (CRIS) sought to understand the impact of self-reported COVID-19, long COVID and other respiratory infections.

Data were collected through self-completion of online questionnaires. Data were analysed to understand the impact of self-reported COVID-19 and respiratory infections on the lives of individuals and on the community, including absence from work, and the impact on health services.

These experimental statistics can be used to highlight potential pressures for the NHS and to help support wider services, for example, assisting governments with informed decisions on important policies, such as service planning and vaccination rollouts.

Uncertainty

Estimates in our publications contain some uncertainty. There are many sources of uncertainty, but the main possible sources in our CRIS publications would include:

- quality of data collected in the questionnaire
- the data are based on a sample of people rather than the whole population
- potential non-response bias, which may not be fully mitigated by the methods used to adjust for this including weighting
- 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

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. We also minimised this during data collection whereby participants were only able to submit one answer for each question. For more information on uncertainty please see our publication on [Uncertainty and how we measure it for our surveys](#).

Communicating uncertainty

To quantify uncertainty in our analyses, we present 95% confidence intervals in our data. This was because the data were drawn from a sample and the published estimates are modelled data produced by analyses, which are based on a number of assumptions.

Confidence intervals give an indication of the degree of uncertainty of an estimate, with a wider interval indicating more uncertainty in the estimate. Overlapping intervals indicate that there may not be a true difference between two estimates. Further information on confidence intervals can be found in [Section 6. Glossary](#).

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 to help us understand trends in different population sub-groups across the UK.

The most recent data up to 12 June 2023, show that within the CRIS sample:

- the overall sample was representative of all of Wales, Northern Ireland and Scotland in terms of population share along with the majority of English regions
- females were slightly over-represented, while males were slightly under-represented at the UK level (UK 52% female and 48% male, CRIS 56% female and 44% male) and for each UK country (England 52% female and 48% male, CRIS 56% female and 44% male; Wales 52% female and 48% male, CRIS 55% female and 45% male; Scotland 52% female and 48% male, CRIS 58% female and 42% male; Northern Ireland 51% female and 49% male, CRIS 55% female and 45% male)
- younger age groups (aged 16 to 24 years, 25 to 34 years, and 35 to 49 years) were under-represented when compared with older age groups (aged 50 to 69 years and 70 years and over), which were over-represented
- those reporting white ethnicity were largely over-represented in England (83%, CRIS 95%) and slightly over-represented in Wales (95%, CRIS 98%)
- those living in two-person households were largely over-represented in both England (34%, CRIS 52%) and Wales (36%, CRIS 56%)

The following tables show the representativeness analysis of the CRIS sample who provided survey responses from the start of the UK-wide survey in April 2023 up to 12 June 2023. The unweighted response population is the actual number of people taking part in the survey.

Table 1a: Actual population and CRIS response population by sex, UK, up to 12 June 2023

Category	Actual relevant proportion population profile			CRIS unadjusted population sample profile		
	Number of people	Proportion	Cumulative proportion	Number of people (excluding missing)	Proportion	Cumulative proportion
Female	27,610,048	0.52	0.52	68,069	0.56	0.56
Male	25,896,195	0.48	1.00	52,561	0.44	1.00

Source: COVID-19 and Respiratory Infections Survey from the Office for National Statistics

Notes

1. Population figures used for the "actual relevant population" are drawn from Census 2021 for England and Wales and ONS mid-year population estimates for Northern Ireland and Scotland.
2. Response population figures are unadjusted and represent the actual response numbers.

Table 1b: Representativeness of CRIS response population compared with the actual relevant population by sex, UK, up to 12 June 2023

**CRIS Data Collection to actual relevant population
Absolute and relative difference in unadjusted population samples (%)**

Category	Absolute difference to actual population	Relative difference to actual population
Female	0.05	9%
Male	-0.05	-10%

Source: COVID-19 and Respiratory Infections Survey from the Office for National Statistics

Notes

1. Population figures used for the "actual relevant population" are drawn from Census 2021 for England and Wales and ONS mid-year population estimates for Northern Ireland and Scotland.
2. Response population figures are unadjusted and represent the actual response numbers.
3. Absolute difference represents the difference in the unweighted proportion in our sample to the actual population.
4. Relative difference represents the proportion of difference attributed to each group. This is calculated by taking the absolute difference and dividing by the proportion of the actual population.

Table 1c: Actual population and CRIS response population by age, UK, up to 12 June 2023

Category (years)	Actual relevant proportion population profile			CRIS unadjusted population sample profile		
	Number of people	Proportion	Cumulative proportion	Number of people	Proportion (excluding missing)	Cumulative proportion
16 to 24	6,594,169	0.12	0.12	2,798	0.02	0.02
25 to 34	8,935,166	0.17	0.29	4,384	0.04	0.06
35 to 49	12,828,535	0.24	0.53	19,798	0.16	0.22
50 to 69	16,347,936	0.31	0.84	59,031	0.49	0.71
70 +	8,800,436	0.16	1.00	34,619	0.29	1.00

Source: COVID-19 and Respiratory Infections Survey from the Office for National Statistics

Notes

1. Population figures used for the "actual relevant population" are drawn from Census 2021 for England and Wales and ONS mid-year population estimates for Northern Ireland and Scotland.
2. Response population figures are unadjusted and represent the actual response numbers.

Table 1d: Representativeness of CRIS response population compared with the actual relevant population by age, UK, up to 12 June 2023

CRIS Data Collection to actual relevant population		
Absolute and relative difference in unadjusted population samples (%)		
Category (years)	Absolute difference to actual population	Relative difference to actual population
16 to 24	-0.1	-81%
25 to 34	-0.13	-78%
35 to 49	-0.08	-32%
50 to 69	0.18	60%
70 +	0.12	74%

Source: COVID-19 and Respiratory Infections Survey from the Office for National Statistics

Notes

1. Population figures used for the "actual relevant population" are drawn from Census 2021 for England and Wales and ONS mid-year population estimates for Northern Ireland and Scotland.
2. Response population figures are unadjusted and represent the actual response numbers.
3. Absolute difference represents the difference in the unweighted proportion in our sample to the actual population.
4. Relative difference represents the proportion of difference attributed to each group. This is calculated by taking the absolute difference and dividing by the proportion of the actual population.

Table 1e: Actual relevant population by ethnicity, England and Wales, up to 12 June 2023

Actual relevant proportion population profile						
Category	England			Wales		
	Number of people	Proportion	Cumulative proportion	Number of people	Proportion	Cumulative proportion
White	37,460,203	0.83	0.83	2,375,038	0.95	0.95
Asian/Asian British	4,020,728	0.09	0.92	64,087	0.03	0.97
Black/African /Caribbean /Black British	1,737,559	0.04	0.96	19,279	0.01	0.98
Mixed/Multiple ethnic groups	905,211	0.02	0.98	27,095	0.01	0.99
Other ethnic groups	931,904	0.02	1.00	18,661	0.01	1.00

Source: COVID-19 and Respiratory Infections Survey from the Office for National Statistics

Notes

1. Population figures used for the "actual relevant population" are drawn from Census 2021 for England and Wales.

Table 1f: CRIS response population by ethnicity, England and Wales, up to 12 June 2023

CRIS unadjusted population sample profile						
Category	England			Wales		
	Number of people	Proportion (excluding missing)	Cumulative proportion	Number of people	Proportion (excluding missing)	Cumulative proportion
White	95,684	0.95	0.95	6,240	0.98	0.98
Asian/Asian British	2,970	0.03	0.98	49	0.01	0.99
Black/African /Caribbean /Black British	734	0.01	0.98	8	0.00	0.99
Mixed/Multiple ethnic groups	941	0.01	0.99	35	0.01	1.00
Other ethnic groups	598	0.01	1.00	17	0.00	1.00

Source: COVID-19 and Respiratory Infections Survey from the Office for National Statistics

Notes

1. Population figures used for the "actual relevant population" are drawn from Census 2021 for England and Wales.
2. Response population figures are unadjusted and represent the actual response numbers.

Table 1g: Representativeness of CRIS response population compared with the actual relevant population by ethnicity, England and Wales, up to 12 June 2023

CRIS Data Collection to actual relevant population				
Absolute and relative difference in unadjusted population samples (%)				
Category	England		Wales	
	Absolute difference to actual population	Relative difference to actual population	Absolute difference to actual population	Relative difference to actual population
White	0.12	14%	0.03	4%
Asian/Asian British	-0.06	-67%	-0.02	-70%
Black/African /Caribbean /Black British	-0.03	-81%	-0.01	-84%
Mixed/Multiple ethnic groups	-0.01	-54%	-0.01	-49%
Other ethnic groups	-0.01	-71%	0.00	-64%

Source: COVID-19 and Respiratory Infections Survey from the Office for National Statistics

Notes

1. Population figures used for the "actual relevant population" are drawn from Census 2021 for England and Wales.
2. Response population figures are unadjusted and represent the actual response numbers.
3. Absolute difference represents the difference in the unweighted proportion in our sample to the actual population.
4. Relative difference represents the proportion of difference attributed to each group. This is calculated by taking the absolute difference and dividing by the proportion of the actual population.

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 completing the survey

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.

Participants are asked to provide employment data. Occupation is provided in a free-text box, whilst employment sector is selected from 15 categories which are coded using the [Standard Occupation Classification](#). This again allows for consistency and comparability of outputs across the UK.

Accessibility and clarity

The ONS-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 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. More information on accessibility and clarity is available in [Section 6, Glossary](#).

CRIS 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

This publication provides [timely and punctual](#) information from the CRIS survey, detailing our analysis on the impact of self-reported COVID-19, long COVID and other respiratory infections. These data were collected, processed, and published within a short time frame.

For more details on related releases, the [GOV.UK release calendar](#) is available online and provides advance notice of release dates.

Why you can trust our data

The ONS is the UK's largest independent producer of statistics and its national statistical institute. Our [Data Policies and Information Charter](#), details how data are collected, secured and used in the publication of statistics. We treat the data that we hold with respect, keeping 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.

Provisional estimates and revisions

The general principle applied to CRIS 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 potentially being discovered in raw or derived variables.

While every effort is made to thoroughly check the data before they are published or released for dissemination, errors do occasionally occur. When errors occur, corrections are made in a timely manner, announced and clearly explained to users in line with our [Guide to statistical revisions](#). Work is also undertaken to mitigate the same error happening again, for example by reviewing and improving code.

5 . Methods used to produce the data

The data collected by the COVID-19 and Respiratory Infections Survey (CRIS) enables us to estimate symptoms, respiratory infections and long COVID by important characteristics and analyse the impact on work, education, and healthcare.

How we collect the data

Sampling method

The [Overview of the COVID-19 and Respiratory Infections Survey](#) outlines the criteria on which CRIS participation was based. Here, Coronavirus (COVID-19) Infection Survey (CIS) participants were considered eligible to be invited to CRIS if all participating household members had been active CIS participants; completed a CIS questionnaire in the 90 days prior to 13 March 2023; and had consented to be contacted about participating in future research studies.

Households where all individuals within the household responded via online in the last 90 days were selected, along with a small number of telephone participants. The small number of telephone participants were selected at individual level from those who stated telephone as their preferred method of communication during the CIS.

More information on the initial CIS sampling method from which our participant pool was selected is available in the [Coronavirus \(COVID-19\) Infection Survey quality report: December 2022](#), which was last updated 30 March 2023.

Data we collect

We collected data from each participant by using a questionnaire. This asked participants questions on vaccinations (COVID-19, flu), respiratory infections, symptoms experienced in the past seven days, including long COVID, use of healthcare services and time off work (because of a respiratory infection or for general health reasons), working from home, social contact within a work or healthcare setting and travel. The questionnaire was carried out online, with a small number of participants completing the questionnaire by telephone interview. All symptoms and respiratory infections were self-reported.

How we analyse the data

The primary objective of the study was to understand the impact of self-reported respiratory infections, including COVID-19 and long COVID.

The analysis of the questionnaire data focused on three primary areas: reported symptoms, impact of respiratory infections and long COVID.

Reported symptoms and impact

The main aim of the reported symptoms analysis was to identify factors associated with reporting influenza-like illnesses (ILI). This was achieved through monitoring changes in self-reported symptoms over time and examining how the likelihood of reporting symptoms varies by important characteristics (for example, ethnicity and age).

Data on self-reported respiratory infections were used to determine the characteristics of people who reported:

- symptoms consistent with ILI, as classified by the [United States Centers for Disease Control and Prevention \(CDC\)](#), in the last seven days
- symptoms consistent with ILI, as classified by the [European Centre for Disease Control \(ECDC\)](#), in the last seven days
- one or more days off work because of a respiratory infection, in the last 28 days
- one or more GP appointments for a respiratory infection, in the last 28 days
- knowing or thinking they have had a respiratory infection, in the last 28 days

Participants were asked to report whether they had experienced a number of symptoms over the past seven days, including those such as abdominal pain, cough, diarrhoea, fever, loss of taste, loss of smell and shortness of breath. Survey weights were applied to make the sample representative of the population in terms of sex, age and region.

As well as calculating the percentage of participants reporting each symptom, indicators for participants reporting ILI were produced.

Symptoms consistent with ILI were defined as follows:

- by the [European Centre for Disease Prevention and Control \(ECDC\)](#) to include at least one of fever, fatigue, headache or myalgia and at least one of cough, sore throat or shortness of breath.
- by the [Centres for Disease Control and Prevention \(CDC\)](#) to include a fever (temperature of 100 degrees Fahrenheit or greater), and a cough or sore throat (or both).

Indicators for both definitions of ILI were calculated but as the CDC definition, required a fever to be recorded, the number of participants fitting this ILI definition was low and the ECDC definition was used for analyses.

Weekly estimates of the percentage reporting each symptom were produced, for the main 6-week period from week beginning 30 April 2023 to week beginning 4 June 2023. This related to just over 118,000 participant responses.

For more detailed breakdowns (for example by age and country), data for a five-week period from 30 April 2023, were aggregated to ensure the sample size was large enough. We were able to do this because the weekly trends for each symptom were largely flat for the period during which CRIS data were collected. This time period was used to ensure each participant was only included once.

Long COVID

The data were also analysed, using descriptive statistics, to identify the characteristics of people who report having long COVID and the symptoms they experience.

For more detailed information on how we investigated long COVID, please refer to our article, [Symptoms of those with self-reported long COVID in the UK](#).

Statistical testing

Logistic regression

Multivariable statistical modelling can be used to estimate the relative effect of each characteristic on the likelihood of each outcome of interest (for example, reporting an influenza-like illness (ILI)), while controlling for other factors.

We used logistic regression to estimate the [odds ratios](#) for:

- reporting an ILI (ECDC) in the previous seven days
- missing at least one day of work or school because of respiratory illness
- attending a GP appointment

Each were examined in terms of selected demographic, socio-economic and geographical factors. Models were run initially with a core set of variables. To include work sector in the model, a separate model was produced, with participants filtered to include only those working.

Generalised Additive Models (GAMs)

GAMs are a framework for modelling outcomes with the flexibility to allow for non-linear trends. The GAMs model a trend as a smooth non-linear function of time.

GAMs were measured on data from 2 May 2023 to 3 June 2023 and were calculated from daily responses. GAMs were modelled using a negative binomial distribution with log link, modelled with thin plate splines (k=30). The models produced are weighted to address unrepresentative non-response patterns, and 95% confidence intervals are included.

The models that are presented show the predicted percentage reporting an outcome over time. The outcomes measured are symptoms consistent with an influenza-like illness (ECDC definition), self-reported respiratory infection in the previous 28 days, and, in the previous 28 days, whether time was spent off work or education because of respiratory infection.

All models have been subjected to diagnostic tests to establish whether the number of basis dimensions are suitable for that model and whether models show concurvity. Concurvity tests show whether smooths relate to each other in the same non-linear trend, similar to measures of collinearity but allowing for non-linear relationships. Diagnostic tests for all models presented suggest the number of basis dimensions is sufficient for the model and that concurvity remains low.

Outcomes and predictors

The outcome variables for all models were a binary measure (yes or no) of whether a participant reported experiencing the outcome, in the last seven days for symptoms consistent with the ILI models, and in the last 28 days for all other models.

For all models, five predictor variables were included in the model irrespective of statistical significance - sex, age, ethnicity, region and deprivation - creating a core model. Based on previous CIS findings and exploration of the CRIS data, the following predictor variables were examined for their association with the outcome variables, after controlling for the core variables:

- survey week
- self-reported long COVID
- self-reported long-term health condition/s
- smoking status
- work sector

Statistical analysis

Forward selection was used to build the regression models. Starting with the core model, for each forward step, the variable that gave the single best improvement to the model, based on the Akaike Information Criterion (AIC) was retained. This process was repeated until adding more variables no longer led to a statistically significant reduction in the AIC. The variance inflation factor (VIF), which measures the strength of correlation between predictor values in a regression model, was used to check for multicollinearity in the final models.

To facilitate interpretation of the results, the model coefficients were exponentiated to odds ratios, and confidence intervals calculated.

Given the large proportion of participants who were neither in work, or education - and the consequent reduction in sample sizes - separate models were built to examine the relationship between work sector and the outcome variables. Odds ratios and confidence intervals for work sector, only, are presented from this model.

Statistical significance

The statistical tests produced 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 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 are considered to provide relatively strong and moderate evidence of difference between the groups being compared, respectively.

Weighting

The weighting strategy accounts for:

- the probability of selection in CRIS
- the probability of responding to CRIS
- known population totals

Probability of selection in CRIS

The CRIS sample is a subset of the CIS population. Therefore, the probability of being selected for CRIS is linked to the probability of being selected for CIS. For this reason, CIS design weights are used as a basis for the CRIS design weights.

For most cases inclusion in CRIS was based on fitting certain requirements making their probability of selection 1. However, there are factors that influence their likelihood of fitting those selection criteria, for example, some participants being more likely to respond online or consent to further research. To account for this, a logistic model was constructed to identify the probability of a participant being selected for CRIS given they were in CIS. This included:

- Age (restricted cubic spline)
- Sex (male and female)
- Region (9 English regions, Wales, Scotland, and Northern Ireland)
- Previous long COVID status

For each participant, the CIS design weight was multiplied by the inverse of the probability of selection for CRIS given they were in CIS to produce the CRIS design weights.

Probability of responding to CRIS

In CRIS, those with long COVID were seen to be more likely to respond to the survey than those not reporting long COVID. This was a main consideration when producing the model to identify variables that contribute to participants' choice of answers provided in the CRIS survey. The model was then used to calculate the probability of response. The model included:

- age (restricted cubic spline)
- sex (male and female)
- region (London and non-London within England, Wales, Scotland, and Northern Ireland)
- ethnicity (white and non-white)
- deprivation (deciles of the Index of Multiple Deprivation)
- disability (2021 Census definition; no conditions, with conditions but no impact of day-to-day activities, with conditions and a little impact, with conditions a lot of impact)
- previous long COVID status

The CIS design weights were multiplied by the inverse of the probability of responding to create an initial weight. These weights were standardised to be used in GAMs. They are also the basis used for calibration.

Calibration and population totals

Initial weights were bounded, to reduce variability by removing extreme weights, and calibrated for some CRIS datasets.

For the weighting of the symptoms analyses, seven-day symptom files were created including participants that responded in the corresponding seven-day window. If any instances where more than one visit was recorded for a participant, the participant was removed from the file as the true response was unknown. Weeks beginning 30 April 2023 to 4 June 2023 were weighted.

For England, the calibration groups were age group by sex and region separately. For the Devolved Administrations, the calibration groups were age group and sex separately. In most instances the age groups were 16 to 24 years, 25 to 34 years, 35 to 49 years, 50 to 69 years and 70 years and over. For Wales, the under 35 years age groups were combined for the first two weeks and for the week beginning 28 May 2023. The under 50 years age groups were also combined for the last week. For Scotland, the under 35 years age groups were combined for the first week and the last two weeks. For Northern Ireland, the under 35 years age groups were combined for the first two weeks. The under 50 years age groups were also combined for the last two weeks. These groups were combined because the sample sizes in the lower age groups were too small.

6 . Glossary

Relevance

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

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.

Confidence Interval

A 95% confidence interval is the range of values that you would expect your estimate to fall between 95% of the time if you were to repeat the study. For confidence intervals, the probability that the population estimate lies between the upper and lower limits of the interval is based upon hypothetical repeats of the study. For instance, in 95 out of 100 studies, we would expect that the true population estimate would fall within the 95% confidence intervals. While the remaining five studies would deviate from the true population estimate. Here we assume the population estimate is fixed and any variation is because of differences within the sample in each study.

Odds ratio

The odds ratio is a measure of how likely an outcome is, given a particular characteristic, compared with a baseline.

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.

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.

Influenza-like illness (ILI)

ILI is a term used to describe a diagnosis of possible influenza or other illness causing a set of common symptoms.

Long COVID

Long COVID was self-reported according to the following CRIS question: "Would you describe yourself as currently having long COVID? Long COVID can be described as still experiencing symptoms more than 4 weeks after you first had COVID-19, that are not explained by something else."

7 . Related links

[Consumer price inflation, updating weights](#)

Article | Released 18 March 2019

An overview of the latest annual update of Consumer Prices Index including owner occupiers' housing costs (CPIH) weights.

[Coronavirus \(COVID-19\) Infection Survey: methods and further information](#)

Methodology | Last revised 20 April 2023

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 Coronavirus (COVID-19) Infection Survey.

[Protocol and information sheets - Nuffield Department of Medicine \(ox.ac.uk\)](#)

Webpage

This study protocol specifies the research for the Coronavirus Infection Survey study.

[Coronavirus \(COVID-19\) Infection Survey, UK: 24 March 2023](#)

Bulletin | Released 24 March 2023

Percentage of people testing positive for coronavirus (COVID-19) in private residential households in England, Wales, Northern Ireland and Scotland, including regional and age breakdowns.

[Self-reported long COVID symptoms, UK: 10 July 2023](#)

Bulletin | Released 10 July 2023

Estimates of self-reported long COVID symptoms and associated activity limitation, using COVID-19 and Respiratory Infections Survey (CRIS) data. Experimental Statistics.

[Prevalence of symptoms and impact of respiratory infections, UK: 10 July 2023](#)

Bulletin | Released 10 July 2023

Percentage of people reporting symptoms of respiratory infections in private households in the UK and the impact on work, education and healthcare services, from the COVID-19 and Respiratory Infections Survey.

8 . Cite this methodology

Office for National Statistics (ONS), released 10 July 2023, ONS website, methodology, [COVID-19 and Respiratory Infections Survey: QMI](#)