

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection De-identified study data were accessed through the Office for National Statistics (ONS) Secure Research Service (SRS). The data available in SRS were prepared for data analysis using Stata MP 16.

Data analysis All statistical analyses were performed using standard functions in the following R packages: ggplot2 (version 3.3.2), rms (version 6.0-1), dplyr (version 1.0.2), emmeans (version 1.5.1), haven (version 2.3.1), sandwich (version 3.0-0), ggeffects (version 1.0.1), broom (version 0.7.2), multcomp (version 1.4-14), and Epi (version 2.44). Code used for data analysis is available upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data are still being collected for the COVID-19 Infection Survey. De-identified study data are available for access by accredited researchers in the ONS Secure Research Service (SRS) for accredited research purposes under part 5, chapter 5 of the Digital Economy Act 2017. For further information about accreditation, contact Research.Support@ons.gov.uk or visit the SRS website.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This analysis included all visits with positive or negative swab results obtained between 1 December 2020 to 3 April 2020 from participants of the ongoing COVID-19 Infection Survey. In total, 1,945,071 RT-PCR results from nose and throat swabs from 383,812 individuals were included.
Data exclusions	This analysis included participants aged 16 years or over (i.e. those who theoretically could have received vaccination), and all visits with positive or negative swab results from 1 December 2020 to 8 May 2021.
Replication	The statistical analyses have been successfully replicated by two individuals.
Randomization	The following potential confounders were adjusted for in all models as potential risk factors for acquiring SARS-CoV-2 infection: geographic area and age in years (see below), sex, ethnicity (white vs non-white as small numbers), index of multiple deprivation (percentile, calculated separately for each country in the UK), working in a care-home, having a patient-facing role in health or social care, presence of long-term health conditions, household size, multigenerational household, rural-urban classification, direct or indirect contact with a hospital or care-home, smoking status, mode of travel to work, work location, and visit frequency. Details are shown in Supplementary Table 1.
Blinding	Since we compared multiple exposure categories to the same reference and also included splines modeling time since vaccination blinding was not feasible in this observational study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Private households are randomly selected on a continuous basis from address lists and previous surveys to provide a representative sample across the UK. The characteristics of the study population are provided in table S1.
Recruitment	<p>The Office for National Statistics (ONS) COVID-19 Infection Survey (CIS) is a large household survey with longitudinal follow-up (ISRCTN21086382, https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/protocol-and-information-sheets) (details in20). The study received ethical approval from the South Central Berkshire B Research Ethics Committee (20/SC/0195). Private households are randomly selected on a continuous basis from address lists and previous surveys to provide a representative sample across the UK. Following verbal agreement to participate, a study worker visited each selected household to take written informed consent for individuals aged 2 years and over. Parents or carers provided consent for those aged 2-15 years; those aged 10-15 years also provided written assent. All participants who completed the enrolment visit was offered a £50 voucher, and one £25 voucher for each further visit. For the current analysis we only included individuals aged 16 years and over.</p> <p>While certain factors might drive non-response to invitations to participate, adjustment for covariates that may influence selection into the sample ensures that estimates of relative effects are not biased by factors that both influence selection into the sample and the risk of the outcome (model-based inference). Factors that were included in the model included:</p>

geographic area and age in years (see below), sex, ethnicity (white vs non-white as small numbers), index of multiple deprivation (percentile, calculated separately for each country in the UK), working in a care-home, having a patient-facing role in health or social care, presence of long-term health conditions, household size, multigenerational household, rural-urban classification, direct or indirect contact with a hospital or care-home, smoking status, mode of travel to work, work location, and visit frequency.

Ethics oversight

The study received ethical approval from the South Central Berkshire B Research Ethics Committee (20/SC/0195).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

ISRCTN21086382

Study protocol

<https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/protocol-and-information-sheets>

Data collection

The Office for National Statistics (ONS) COVID-19 Infection Survey (CIS) is a large household survey with longitudinal follow-up (ISRCTN21086382, <https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/protocol-and-information-sheets>) (details in 20). Following verbal agreement to participate, a study worker visited each selected household to take written informed consent for individuals aged 2 years and over. Parents or carers provided consent for those aged 2-15 years; those aged 10-15 years also provided written assent. For the current analysis we only included individuals aged 16 years and over. Individuals were asked about demographics, behaviours, work, and vaccination uptake (<https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/case-record-forms>). At the first visit, participants were asked for (optional) consent for follow-up visits every week for the next month, then monthly for 12 months from enrolment. At each visit, enrolled household members provided a nose and throat self-swab following instructions from the study worker, which is comparable to or even more sensitive than swabs done by healthcare workers. From a random 10-20% of households, those 16 years or older were invited to provide blood monthly for antibody testing.

From 1st December 2020 to 8th May 2021, 383,812 individuals from 216,953 households provided 1,945,071 RT-PCR results from nose and throat swabs in the COVID-19 Infection Survey

Outcomes

Analysis was based on visits, since these occur independently of symptoms and are therefore unbiased. Only the first test-positive visit in the first new positive infection episode starting after 1 December was used, dropping all subsequent visits in the same infection episode, to avoid misattributing ongoing PCR-positivity to visit characteristics. Primary analysis included all first new positive infection episodes. Secondary analyses considered the impact of vaccination, as being implemented in the UK, on infection severity, by classifying positives by cycle threshold (Ct) value (<30 or ≥30) and self-reported symptoms. The threshold of a Ct value of 30 is somewhat arbitrary, but corresponds to ~150 copies/ml, and is consistently used in the UK for many purposes, including algorithms for review of low level positives at the laboratories where the PRC tests were performed and a threshold for attempting whole genome sequencing. For each positive test, a single Ct was calculated as the arithmetic mean across detected genes (Spearman correlation > 0.98), then the minimum value was taken across positives in the infection episode to reflect the greatest measured viral burden within an episode. To allow for pre-symptomatic positives being identified in the survey, any self-reported symptoms at any visit within 0 to 35 days after the index positive in each infection episode were included (questions elicit symptoms in the last 7 days at each visit). Finally, positive infection episodes were classified as compatible with the B.1.1.7 SARS-CoV-2 variant (those positive at least once for ORF1ab+N across the episode and never S-positive) and those that were incompatible (ORF1ab+N+S or ORF1ab+S or N+S at least once). B.1.1.7 has deletions in the S gene leading to S gene target failure, and ORF1ab+N positivity only remains a good proxy for B.1.1.7 from whole-genome sequencing from mid November 2020. Positives where only a single N or single ORF1ab gene were detected were excluded from this secondary analysis.