

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: No software was used by investigators in collecting data, as this is a secondary data analysis using data that have already been collected.
- Data analysis: R version 3.6.3 was used to analyse data.

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

An anonymised data set allowing replication of the analysis is available at <https://zenodo.org/record/4579857>.

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	The sample size was determined by the full data set available, as this was an observational study rather than a clinical trial. The full sample size used in the paper was 2,245,263 subjects.
Data exclusions	We excluded individuals with missing data for sex (n = 14), age (n = 171), and IMD and regional covariates (n = 3,817). We also excluded age-0 individuals as there were 17,913 age-0 individuals compared to 10,312 age-1 individuals in the dataset, suggesting that some age-0 individuals may have been miscoded.
Replication	As this was not an experimental study, it was not possible to replicate any experiment. However, we performed multiple sensitivity analyses considering different time period, covariates, and stratification methods and found consistent results for each replication. We reported all sensitivity analyses performed. We also analysed the data using both complete-cases and inverse probability weighting methods.
Randomization	Individuals were not randomised, as the exposure of interest was infection with one of two variants of SARS-CoV-2. Our statistical analysis controls for age, sex, ethnicity, type of residence (residential housing, care/nursing home, or other), index of multiple deprivation, date of virus test and local authority of residence.
Blinding	Individuals were not allocated to groups by the investigators. Investigators were not blinded to the exposure groups during analysis.

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

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

Population characteristics	Subjects were aged 1-110; male and female; White, Asian, Black, and other/mixed/unknown ethnicities; from all deciles of the Index of Multiple Deprivation in England, from all 7 NHS England regions, residing in residential housing, care homes, or other accommodation, and tested for SARS-CoV-2 between 1 September 2020 and 14 February 2021. Full details of the study population are presented in Extended Data Table 1.
Recruitment	This is a secondary data analysis study using linked data from all individuals who took a SARS-CoV-2 community test in England between 1 September 2020 and 14 February 2021.
Ethics oversight	Approved by the Observational / Interventions Research Ethics Committee at the London School of Hygiene and Tropical Medicine (reference number 24020).

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