nature research

Corresponding author(s): Claire J. Steves

Last updated by author(s): Feb 15, 2021

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	The data collection is made through the Zoe Covid Symptom Study App with code available under the following link: https://github.com/zoe/covid-tracker-react-native
Data analysis	The following packages were used for the analysis of the data - all the analysis was performed using python3.7 numpy package version 1.16.4 pandas package version 0.25.0 Statsmodels v0.11.1 kmode package v0.10.2 networkx package version 2.3 scipy package version 1.3.1 sklearn package version 0.22.2.post1 exetera package Link to the code for the data extraction is also available through the data access link and also available at https://github.com/KCL-BMEIS/ExeTera

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 and materials availability: Data used in this study is available to bona fide researchers through UK Health Data Research using the following link https://web.www.healthdatagateway.org/dataset/fddcb382-3051-4394-8436-b92295f14259 -

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

Life sciences study design

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

Sample size	No sample size was calculated in this study as we considered all participants matching the inclusion criteria
Data exclusions	To be included in the analysis, users of the COVID Symptoms Study app were selected based on the following criteria defined prior to the running of any of the analysis: Inclusion criteria: Age >=18 yrs; reporting a positive SARS-CoV-2 swab test (PCR) confirming the diagnosis of COVID-19; disease onset between 14 days before and 7 days after the test date, and before the 30th June 2020 (to limit right censoring). Exclusion criteria: individuals who started app reporting when already unwell; users reporting exclusively healthy throughout the study period; users with gaps of more than 7 days after an unhealthy report and not reporting any hospital visit (to account for gaps due to hospitalisation). In addition, individuals reporting for less than 28 days but reporting more than 5 symptoms at their last log were excluded, as duration could not be ascertained.
Replication	Replication of the modelling was ensured by inputing a seed as parameter of the fold generation allowing for reproduction of the results. Due to the necessity for regular logging, there was no equivalent external dataset on which to reproduce the findings
Randomization	No group randomization was necessary to the design of the study as it was an observational study
Blinding	There were no random group allocation in this study so no blinding was necessary

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	Design. Prospective observational study Setting. The Covid Symptom Study app, launched on March 24th 2020

	Participants. SARS-CoV2 incident swab test PCR+ve adults logging their symptoms prospectively. Among the 4182 selected individuals, there were 1190 males and 2292 females with mean age 42.8 (SD 13.4)
Recruitment	Participants were recruited through national media campaigns, existing cohort studies and through sharing on social media from the start of the pandemic. Recruitment necessitates access to a mobile electronic device, although individuals without access could report through another. Representativeness of the population is presented in the manuscript, and adjustments were made to balance gender and age. Specific analyses (supplementary table 5 and 6) were performed to assess the influence of exclusion groups on the estimates of LC28 as well as the characteristics of these groups. Since regular logging was a necessary condition for the inclusion in the study, there may be a systematic bias in the symptom reporting. Additionally, one must keep in mind that at the early stages of the pandemic, PCR testing was restricted to those more severely unwell.
Ethics oversight	Ethics: Informed consent for participation was integrated into the App by ZOE Global and given by all participants. Ethical approval was granted KCL ethics Committee REMAS ID 18210, review reference LRS-19/20-18210. In Sweden, ethics approval for the study was granted by the central ethics committee (DNR 2020-01803). In USA, this study was approved by the Partners Human Research Committee (Protocol 2020P000909).

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