nature research

Corresponding author(s): Giorgio Quer

Last updated by author(s): Oct 1, 2020

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 MyDataHelps smartphone based app was developed by CareEvolution, and it includes the DETECT study since March 25, 2020. The participants have agreed to share the historical data collected by their eligible devices, including heartrate, sleep and activity measures. On first use, the app records self-reported age, gender and location. With continued use the users are encouraged to share additional information through the completion of targeted surveys, including manifestation of symptoms and COVID-19 test outcomes. Data from the participants is automatically uploaded to a protected server.
Data analysis	Analyses were carried out using Python version 3.8.5. The Python packages pandas version 1.1.2 and numpy version 1.19.1 have been used for data processing. Receiver operating characteristic (ROC) curves were computed using the Python package scikit-learn version 0.23.2. Statistical tests and p-values have been evaluated using the Python package scipy version 1.5.2.

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

All interested investigators will be allowed access to the analysis data set following registration and pledging to not re-identify individuals or share the data with a third party. All data inquiries should be addressed to the corresponding author.

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	Between March 25, 2020 and June 7, 2020, 30,529 US individuals were enrolled in the DETECT study. During the study, 3,811 individuals reported symptoms through the DETECT smartphone app. 480 individuals reported having had a SARS-CoV-2 test, 333 of which having received the outcome (54 individuals tested positive; 279 tested negative). We applied bootstrap resampling method with 10,000 independent iterations to estimate the uncertainty of the outcomes due to the sample size. Due to the observational nature of the study, we did not perform any statistical analysis to predetermine the sample size, which was enforced by the number of active participants enrolled.
Data exclusions	The 3,811 individuals who completed the report for symptoms in the DETECT app were included in the analysis of frequency of symptoms. All the 333 individuals who had received the outcome of the SARS-CoV-2 test were included in the analysis of the proposed models. The data exclusion criteria for the analysis were pre-established. Due to the nature of the study, only individuals owning a smartwatch or activity tracker device were able to be enrolled. No data samples that were eligible for the analysis has been excluded.
Replication	The study has not been replicated, due to the nature of the data collection process, requiring the enrollment of thousands of individuals. The methods have been described in detail in the paper with all the software packages adopted, so it is possible to replicate the study by accessing a similar but independent dataset.
Randomization	Randomization was not relevant for this study, as participants have not been divided in two or more cohorts at the beginning of the study. The study focuses on the discrimination between individuals who tested positive or negative to COVID-19, so no randomization is needed.
Blinding	The group allocation was solely related to the data shared by the participants. Due to the observational nature of the study, the investigators were not involved in any arbitrary group allocation.

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	Between March 25, 2020 and June 7, 2020, our research study enrolled 30,529 individuals with representation from every state in the United States. Among the consented individuals, 62.0% are female and 12.8% are 65 or more years old. 78.4% of participants connected their Fitbit device to the study-app, 31.2% connected the data from Apple Health Kit while 8.1% connected data from Google Fit (note that one individual can connect to multiple platforms). The most important covariates available at the time of analysis are sex, age and device used by each participants. We did not observe any statistically significant difference among subjects with reported positive and negative COVID-19 test.
Recruitment	Any person living in the United States over the age of 18 years old is eligible to participate in the DETECT study by downloading the iOS or Android research app, MyDataHelps. Scripps Research, along with outreach partners, conducted a multi-faceted outreach campaign including social media, educational web-based content, email outreach and in-app notifications to recruit participants for the DETECT Study. The study was also featured in several national media outlets, which increased visibility to larger section of the U.S. population.

	Individuals owning a smartwatch or activity tracker and having access to COVID-19 diagnostic testing may not be fully representative of the general population.
Ethics oversight	The protocol for this study was reviewed and approved by the Scripps Office for the Protection of Research Subjects (IRB 20-7531). All individuals participating in the study provided informed consent electronically.

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