

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

Data were collected using TPP SystmOne software (14th May maintenance release), for the purpose of direct clinical care. Data management was performed using Python 3.8 and SQL. All code for data management and analysis is archived online at <https://github.com/opensafely/risk-factors-research>.

Data analysis

Analysis was carried out using Stata 16.1 / Python 3.8. All code for data management and analysis is archived at <https://github.com/opensafely/risk-factors-research>.

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 data were linked, stored and analysed securely within the OpenSAFELY platform <https://opensafely.org/>. All code is shared openly for review and re-use under MIT open license. Detailed pseudonymised patient data is potentially re-identifiable and therefore not shared. All clinical and medicines codelists are openly available for inspection and reuse at <https://codelists.opensafely.org/>.

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)

Behavioural & social sciences study design

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

Study description	We conducted a quantitative cohort study using national primary care electronic health record data linked to COVID-19 death data.
Research sample	We used patient data from general practice (GP) records managed by the GP software provider The Phoenix Partnership (TPP), linked to Office for National Statistics (ONS) death data. The sample of patients represents approximately 40% of the population of England, spread geographically across the whole country.
Sampling strategy	Our study population consisted of all adults (males and females 18 years and above) currently registered as active patients in a TPP general practice in England on 1st February 2020. To be included in the study, participants were required to have at least 1 year of prior follow-up in the GP practice to ensure that baseline patient characteristics could be adequately captured, and to have recorded sex, age, and deprivation (see covariates, below).
Data collection	Data were collected by clinicians (e.g. doctors, nurses) and administrative staff, for the purpose of direct clinical care. This was carried out on computers using TPP SystmOne software. The researchers were not present for data collection into the TPP database. Data were then queried from the TPP database by the researchers, to create the study dataset. This was carried out using Python 3.8 and SQL software (available here https://github.com/opensafely/risk-factors-research). This study did not have an experimental condition or hypothesis.
Timing	Patients were observed from the 1st of February 2020 and were followed until the first of either their death date (whether COVID-19 related or due to other causes) or the study end date, 6th May 2020.
Data exclusions	To be included in the study, participants were required to have at least 1 year of prior follow-up in the GP practice to ensure that baseline patient characteristics could be adequately captured, and to have recorded sex, age, and deprivation. The total number of excluded patients was 6,322,225.
Non-participation	No participants dropped out.
Randomization	Participants were not allocated into experimental groups.

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	See above
Recruitment	This study uses data gathered during routine medical practice. We selected all patients except those <18 years old, anyone without a recorded sex, age, or deprivation score, and anyone without a year of prior follow-up (to ensure that baseline patient characteristics could be adequately captured). These inclusive criteria mean that bias is minimised.

This study was approved by the Health Research Authority (REC reference 20/LO/0651) and by the LSHTM Ethics Board (reference 21863).

In March 2020, the Secretary of State for Health and Social Care used powers under the UK Health Service (Control of Patient Information) Regulations 2002 (COPi) to require organisations to process confidential patient information for the purposes of protecting public health, providing healthcare services to the public and monitoring and managing the COVID-19 outbreak and incidents of exposure. Taken together, these provide the legal bases to link patient datasets on the OpenSAFELY platform and set aside the requirement for patient consent for COVID-19 related public health research. GP practices, from which the primary care data is obtained, are required to share relevant health information to support the public health response to the pandemic, and have been informed of the OpenSAFELY analytics platform.

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