nature portfolio

Corresponding author(s): Yong Chen, PhD

Last updated by author(s): Nov 25, 2024

Reporting Summary

Nature Portfolio 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 Portfolio 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 is collected from electronic health data RECOVER program.
Data analysis	All analyses were performed using R version 4.1.2. The code used for the analysis in this study is available and can be accessed in the code ocean repository at https://doi.org/10.24433/CO.0460366.v1.

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The results reported in this study are based on detailed individual-level patient data compiled as part of the RECOVER program. Due to the high risk of reidentification based on the number of unique patterns in the date, patient privacy regulations prohibit us from releasing the data publicly. The data are maintained in a secure enclave, with access managed by the program coordinating center to remain compliant with regulatory and program requirements. Please

direct requests to access the data, either for reproduction of the work reported here or for other purposes, to the RECOVER EHR Pediatric Coordinating Center (recover@chop.edu).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	The study involved 225,723 children and adolescents under age of 21 with COVID-19 and 677,448 without COVID-19. Of these, 50.2% were female. 33.5% were <5 years, 28.5% were between age 5 to 11 years, 38.0% above age 12 years. The sex variable used in our analysis is based on the person domain in PEDSnet CDM table. A gender_concept_id of 8532 indicates "Female", and a gender_concept_id of 8507 indicates "Male". All other values are categorized as "Other/Unknown", and we only included Male and Female in our analysis. More details can be found here: https://github.com/PEDSnet/Data_Models_Public/blob/master/PEDSnet/docs/Conventions%20Docs/v5.5_PEDSnet_CDM_ETL_Conventions.md.
Reporting on race, ethnicity, or other socially relevant groupings	The COVID-19-positive cohort comprised 109,022 NHW patients (48.3%), 45,823 NHB patients (20.3%), 60,012 Hispanic patients (26.6%), and 10,866 AAPI patients (4.8%) with SARS-CoV-2 infection from March 2020 to October 2022. The raw data for all ethnic/racial groups are from the "person" CDM table from the PEDSnet CDM in the EHR data. The ethnic and racial information is identified from columns "ethnicity_concept_id" and "race_concept_id". More details can be found here: https://github.com/PEDSnet/Data_Models_Public/blob/master/PEDSnet/docs/Conventions%20Docs/v5.5_PEDSnet_CDM_ETL_Conventions.md.
Population characteristics	We provided the following population charateristics: age at the cohort entry date (<5, 5-12, 12-21), sex (female, male), cohort entry month (from March 2020 to October 2022), site indicators, obesity (obese, non-obese), a chronic condition indicator as defined by the Pediatric Medical Complexity Algorithm (PMCA, no chronic condition, non-complex chronic condition, and complex chronic condition), healthcare visits (inpatient, outpatient, and emergency department visits), medications (0, 1, 2, ≥3), negative tests (0, 1, 2, ≥3), vaccine doses (0, 1, ≥2), and immunization duration during the baseline period (no vaccine, <4 months, ≥ 4 months). The severity of COVID-19 at the cohort entry date was stratified into the following categories: asymptomatic, mild (symptomatic), moderate (involving moderately severe COVID-19-related conditions like gastroenteritis, dehydration, and pneumonia), and severe (comprising unstable COVID-19-related conditions, ICU admission, or mechanical ventilation).
Recruitment	The National Institutions of Health (NIH) launched the new RECOVER initiative in 2021 to leverage electronic health record (EHR) data to better identify and characterize patients with post-acute sequelae of SARS-CoV-2 infection (PASC). RECOVER obtains EHRs from three large national healthcare networks within the United States, covering regional catchment areas across 41 states. These networks collectively hold the EHRs of over 60 million patients, including records from more than 7 million individuals who have been affected by COVID-19.
Ethics oversight	- This study constitutes human subject research. Institute Review Board (IRB) approval was obtained under Biomedical Research Alliance of New York (BRANY) protocol #21-08-508. As part of the Biomedical Research Alliance of New York (BRANY IRB) process, the protocol has been reviewed in accordance with the institutional guidelines. The Biomedical Research Alliance of New York (BRANY) waived the need for consent and HIPAA authorization. - Institutional Review Board oversight was provided by the Biomedical Research Alliance of New York, protocol # 21-08-508-380.

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

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

Behavioural & social sciences study design

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

Study description	Quantitative observational study
Research sample	The study involved 225,723 children and adolescents under age of 21 with COVID-19 and 677,448 without COVID-19. Of these, 50.2% were female. 33.5% were <5 years, 28.5% were between age 5 to 11 years, 38.0% above age 12 years. The sample is representative.
Sampling strategy	We conducted a retrospective study from March 1, 2020, to October 3, 2022, with at least 6 months of follow-up time. We included patients under the age of 21 who had at least one visit within 18 months to 7 days before the index date (defined as the baseline period) and at least one encounter within 28 days and 179 days after the index date (defined as the follow-up period).
Data collection	The National Institutions of Health (NIH) launched the new RECOVER initiative in 2021 to leverage electronic health record (EHR) data to better identify and characterize patients with post-acute sequelae of SARS-CoV-2 infection (PASC). RECOVER obtains EHRs from

	three large national healthcare networks within the United States, covering regional catchment areas across 41 states. These networks collectively hold the EHRs of over 60 million patients, including records from more than 7 million individuals who have been affected by COVID-19. Only the participants and the researcher were present during the study. The researcher was blinded to the study hypothesis.
Timing	From March 1, 2020, to October 3, 2022
Data exclusions	The pre-specified exclusion criteria is established as follows: (1) 245,822 COVID-19 positive patients and 942,295 COVID-19 negative patients were excluded as they had no visit 18 months to 7 days before cohort entry, ensuring the included population have interaction with the health system; (2) 232,071 COVID-19 positive patients and 713,239 COVID-19 negative patients were excluded as they had no visit 28 to 179 days after cohort entry, ensuring the included population have interaction with the health system; (3) 51,725 COVID-19 positive patients and 104,126 COVID-19 negative patients were excluded as they aged > 21 at cohort entry; (4) 260,625 COVID-19 positive patients and 603,785 COVID-19 negative patients were excluded because they had incomplete variables or with race/ethnicity "Multiple" or "Other/Unknown". After applying the exclusion criteria (1)-(4), we have 225,723 children and adolescents with COVID-19 and 677,448 without COVID-19.
Non-participation	No participants dropped out.
Randomization	Non-random. We employed the propensity score model approach to adjust for covariates stemming from various patient characteristics.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A