nature research

Corresponding author(s):   Jure Leskovec

Last updated by author(s):   Oct 7, 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	No data collection was performed for this study; all analysis relied on previously collected datasets, as described in the Data section below.
Data analysis	All data analysis was performed using Python with standard libraries. Code is available at http://snap.stanford.edu/covid-mobility and https://github.com/snap-stanford/covid-mobility.

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

Census data (https://www.census.gov/programs-surveys/acs), case and death counts from The New York Times (https://github.com/nytimes/covid-19-data), and Google mobility data (https://www.google.com/covid19/mobility/) are publicly available. Cell phone mobility data is freely available to researchers, non-profits, and governments through the SafeGraph COVID-19 Data Consortium (https://www.safegraph.com/covid-19-data-consortium).

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	This is a quantitative epidemiological modeling study.
Research sample	We study previously-collected data on 10 of the largest American metropolitan statistical areas. The data came from the following: 1) Anonymized mobility data from smartphone users from SafeGraph. SafeGraph data is obtained by "partnering with mobile applications that obtain opt-in consent from its users to collect anonymous location data", per official SafeGraph documentation. 2) US Census 3) Case and death counts from The New York Times 4) Google mobility data
Sampling strategy	We did not perform sampling, but relied on previously collected datasets. All datasets were chosen as they were comprehensive: the case and death counts and Census data were designed to cover the entire population; Google mobility data covers a large subset of the population; and previous analyses have shown that the SafeGraph anonymized cell phone mobility dataset is geographically representative: for example, it does not systematically over-represent individuals from higher-income areas (https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset).
Data collection	We did not perform data collection, but relied on previously collected datasets. Census data is collected as described here: https://www.census.gov/programs-surveys/acs. The cell phone mobility data is collected from cell phone geolocation information, and is anonymized and aggregated. We accessed and downloaded the datasets in May 2020.
Timing	We make use of data in the following ranges: 1) Safegraph mobility data: Jan 1, 2019 - May 2, 2020 2) US Census: 2013-2018 3) Case and death counts from The New York Times: March 8 - May 9, 2020 4) Google mobility data: March 1 - May 2, 2020
Data exclusions	The original SafeGraph dataset contains 5.4 million points of interest (POIs). We retain a POI in our final dataset if it satisfies the following criteria: (1) it lies within one of the 10 American metropolitan areas that we analyze (out of 384 metropolitan statistical areas total); (2) SafeGraph has visit data for this POI for every hour that we model, from 12am on March 1, 2020 to 11pm on May 2, 2020; (3) SafeGraph has recorded the home CBGs of this POI's visitors for at least one month from January 2019 to February 2020; (4) the POI is not a "parent" POI, as defined in the Methods section. After applying these filters, our dataset contains 553k POIs. Most POIs are filtered out because they do not lie within the 10 large metropolitan statistical areas that we study; this filtering decision was made prior to any analysis for computational tractability reasons. In our analysis of POI-specific category risks, we do not analyze 6 categories of POIs because we wish to be conservative and only focus on categories where we are most confident we are fully capturing transmission at the category: Child Day Care Services, Elementary and Secondary Schools, Drinking Places (Alcoholic Beverages), Nature Parks and Other Similar Institutions, General Medical and Surgical Hospitals, and Other Airport Operations. The justifications for these exclusions, which are based on prior work, are given in the Methods section.
Non-participation	Because we relied on previously collected anonymized, aggregated data from cell phone mobility tracking, we did not have access to individual-level data and do not know how many participants dropped out/declined participation.
Randomization	This is not a randomized controlled trial and participants were not randomized 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	See above. Prior work has looked into biases in the SafeGraph dataset (https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset).
Ethics oversight	The dataset from The New York Times consists of aggregated COVID-19 confirmed case and death counts collected by journalists from public news conferences and public data releases. For the mobility data, consent was obtained by the third-party sources collecting the data. SafeGraph aggregates data from mobile applications that obtain opt-in consent from their users to collect anonymous location data. Google's mobility data consists of aggregated, anonymized sets of data from users who have chosen to turn on the Location History setting. Additionally, we obtained IRB exemption for SafeGraph data from the Northwestern University IRB office.

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