Reproduction of a Research Claim from Siedner et al. (2020),
from medRxiv

Reproduction Team: Radoslaw Panczak

Research Scientist: Nick Fox

Independent Reviewer(s)
(add name below when you initiate review, comment “DONE” on your name when you finish and notify the reproduction team):

Reviewer #1: [Gustav Nilsonne]

As necessary:

Reviewer #2: [NAME]

Reviewer #3: [NAME]

Review Period: December 7 - December 10

View-only links to: Original Paper, Original Materials, Reproduction Materials

Privacy Statement: Other teams are making predictions about the outcomes of many different studies, not knowing which studies have been selected for reproduction. As a consequence, the success of this project requires full confidentiality of this peer review process. This includes privacy about which studies have been selected for reproduction and all aspects of the discussion about these reproduction designs.

Instructions for Data Analysts [Reproductions]

The preregistration for this reproduction study was started by a separate team of researchers who were responsible for identifying data sources and constructing them into a reproduction dataset(s) for your use in the analysis. They have completed sections 1-13 of the preregistration below, and included additional materials in the OSF project that document how the dataset was constructed.

You'll be responsible for filling out sections 16-25 of the preregistration below. Before you do so, **please review the original study, sections 1-15 of the preregistration, and the materials provided on the OSF**, so that you are familiar with all of the decisions that have been made to date. In many cases, the 'data preparer' will have left you instructions and suggestions on how the provided data can be used in the analysis, as well as idiosyncrasies and discrepancies in the data that you should be aware of. The data preparers have tried to be thorough in including all variables that you might need, but please keep in mind the following:

- Some of the variables included in the constructed dataset(s) may not be needed in the final analysis, so please do not feel the need to necessarily use all of the provided variables.
- Some of the variables needed might have mistakenly been excluded from the constructed datasets. If you find that this is the case, please let Andrew know, and he will work with you to supplement the datasets as needed.

For these reproduction analyses, we would like the analysis plan to be completed before the preregistration goes through review, so that after review, the only remaining steps are registration and running the analysis code on the full datasets. To facilitate that, we are asking that you include in section 19 a link to the code you will use that takes the constructed dataset(s) provided to you and produces the focal analyses (including all of the cleaning, merging, and transforming required). When developing your analysis plan and code, please randomly sample 5% of the data for use in your work and demonstrate that the focal analyses produce sensible results using just that random sample by providing a screenshot of the output (see section 19 for details). **Do not use the rest of the data until after your study is registered and it is time to run the final analyses**. In section 19, you will find a statement that we are asking you to bold that confirms you've only used 5% of the data when developing and testing your code. If this approach will not work for any reason, please let Andrew know and disclose deviations from this plan somewhere in the preregistration.

- In cases where we are providing you a complete dataset, you can just sample out 5% of the observations and hold the rest out until you are ready to perform the final analysis.
- In cases where we are providing you multiple datasets that need to be combined prior to analysis, please sample out 5% of the observations in whatever way is most sensible.
    - For example, in cases where each dataset contains complete observations on its own (a typical 'row bind' situation), it makes the most sense to sample out 5% of each dataset separately and then combine them together to develop and test your code.
    - In cases where datasets need to be merged in order to create complete observations (a typical 'column bind' situation), it makes the most sense to merge the separate datasets into a full dataset first, and then sample out the 5% before proceeding with the rest of the analysis code.
- We leave the decision on how to sample out the random subset of data to you, so long as (a) you are not performing any analyses on the complete dataset until after your study is registered and (b) whatever decision you make is documented in the preregistration.

Preregistration of Siedner_covid_P3NJ

Source Data Reproduction

Study Information

1. Title (provided by SCORE)

RR TEAM INSTRUCTIONS: *This has been determined by SCORE.*

Reproduction of a research claim from Siedner et al. (2020).

2. Authors and affiliations

RR TEAM INSTRUCTIONS: *Fill in the names and affiliations of your team below.*

RR LAB LEAD[1]
Radoslaw Panczak Data finder Data analyst [1]

[1] Institute of Social and Preventive Medicine, University of Bern, Mittelstrasse 43, CH-3012 Bern, Switzerland

3. Description of study (provided by SCORE)

RR TEAM INSTRUCTIONS: *This description has been provided by SCORE. Please review and make a SCORE project coordinator aware of any edits, additions, and corrections you would suggest to the paragraph. You are free to add additional descriptions of your project in a separate paragraph.*

The claim selected for reproduction from Siedner et al. (2020) is that implementation of social distancing measures is associated with a reduction in the mean daily growth rate of COVID-19 cases. This reflects the following statement from the paper's abstract: "*The mean daily COVID-19 growth rate decreased beginning four days after implementation of the first statewide social distancing measures, by an additional 0.8% per day; 95% CI, -1.4% to -0.2%; P=0.002)*".

The claim was tested by fitting mixed effects linear regression models, specifying the log difference in daily cases as the outcome of interest and a random effect for state. Explanatory variables included time in days, implementation period, and a time-by-implementation period product term. This resulted in, beginning four days after implementation of the first statewide social distancing measure, the mean daily case growth rate decreasing by an additional 0.8% per day (95% CI, -1.4% to -0.2%; P=0.002).

4. Hypotheses (provided by SCORE with possible Data Analyst additions)

RR TEAM INSTRUCTIONS: *The focal test for SCORE is indicated as H*. If you will test additional hypotheses (or use alternate analyses) that help you to evaluate the claim your reproduction is testing, number them H1, H2, H3 etc. (You can place H* in the list wherever makes sense). Please make sure that any additional hypotheses are logical deductions/operationalizations of the selected SCORE claim or are necessary to properly interpret the focal H* hypothesis. Research that is outside this scope should be described in a separate preregistration.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Are the listed hypotheses specific, concise, clearly testable, and specified at the level of operationalized variables?*
- *Are hypotheses identified as directional or non-directional, and, if applicable, have the direction of hypotheses been stated? (Example: "Customers' mean choice satisfaction will be higher in the CvSS architecture condition than in the standard attribute-by-attribute architecture condition.")*
- *Does the list of hypotheses/tests indicate whether additional hypotheses are taken from the original study or modified/added by the team?*

H*: The implementation of social distancing measures will be associated with a reduction in the mean daily growth rate of COVID-19 cases.

Design Plan

5. Study type

NOTE: *The study type selected should reflect how the data from the reproduction data sources was collected. In most cases, this selection will match the study type of the original study.*

- Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.
- **Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.**
- Meta-Analysis - A systematic review of published studies.
- Other

6. Blinding

RR TEAM INSTRUCTIONS: *Select any/all of the below that apply for your study by bolding them. You will give a longer description in the next question.*

- **No blinding is involved in this study.**
- For studies that involve human subjects, they will not know the treatment group to which they have been assigned.
- Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)
- Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

[QUESTION 6 - BOLD YOUR RESPONSE ABOVE]

7. Blinding

RR TEAM INSTRUCTIONS: *If blinding was used in collecting the original data, please comment on that below. In many cases, “Blinding was not used to the data finder’s knowledge” or “Blinding is not relevant to this study” will be acceptable answers.*

Blinding is not relevant to this study

8. Study Design

RR TEAM INSTRUCTIONS: *Please describe the data sources used to test the original finding and how they compare to the data sources identified for the reproduction attempt. Please state whether all of the original data sources could be identified and accessed. If any of the original data sources could not be identified or accessed, please explain how their absence might affect the reproduction attempt.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify the unit of analysis?*
- *Does the preregistration provide sufficient detail about each data source used in the original study, including whether it could be identified and accessed for the reproduction attempt?*
- *If relevant, does the preregistration sufficiently document the known deviations between the original data sources and the data sources used in the reproduction attempt, including how they might affect the final outcome?*

Siedner *et al.* attempt to quantify the impact of implementing social distancing measures on the progression of COVID-19 epidemic in the US states.

This is an observational study using the 'US state - day' as a unit of observation. The primary outcome is specified as "*the COVID-19 growth rate, calculated as the log of daily COVID-19 cases minus the log of daily COVID-19 cases on the prior day*". The primary exposure was defined as "*time in relation to implementation of the first statewide social distancing measure*". All of the original data sources could be identified and accessed.

The epidemiological data on the number of cases came from the New York Times COVID-19 database (https://github.com/nytimes/covid-19-data). Specifically Authors must have focused on the *State-Level Data* section of *Historical Data* available at (https://github.com/nytimes/covid-19-data/blob/master/us-states.csv). This file describes "*a series of data files with cumulative counts of coronavirus cases in the United States, at the state and county level, over time*". This dataset will be referred to as **nyt dataset** hereafter.

The Authors compiled information from multiple government websites and third-party sources to identify all statewide social distancing measures implemented between January 21 and March 30, 2020. Information about methodology is available in the eAppendix and full dataset is available in eTable 1 of the the Supplementary Material to the paper (https://www.medrxiv.org/content/medrxiv/suppl/2020/04/08/2020.04.03.20052373.DC1/2020.04.03.20052373-1.pdf). eTable 1 needs to be extracted from the pdf document and all dates needed to construct timings of the implementation of social distancing measures by state and type can then be constructed from there.This dataset will be referred to as **measures dataset** hereafter.

One of the sensitivity analyses performed in the paper adjusted models for “population density”. No further information is provided for this aspect of the analysis. Presumably Authors refer to the state’s population density. Such data can be relatively easily obtained from publicly available sources such as The Census Bureau's website for State Population Totals and Components of Change: 2010-2019; specifically *Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019* (NST-EST2019-01) table available here) for the numerator and data from the table available on the *State Area Measurements and Internal Point Coordinates* website for the denominator.

9. Randomization (free response)

RR TEAM INSTRUCTIONS: *If the variables used for this reproduction attempt were randomized, state how they were randomized, and at what level. In many cases, “Randomization was not used to the data finder’s knowledge” or “Randomization is not relevant to this study” will be acceptable answers.*

Randomization is not relevant to this study

Sampling Plan

This section describes how the data sources for the reproduction were selected and how they were prepared into a reproduction dataset. Please keep in mind that the data described in this section are the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.

10. Existing data

1.1.1. Registration prior to creation of data
1.1.2. Registration prior to any human observation of the data
1.1.3. Registration prior to accessing the data
1.1.4. Registration prior to analysis of the data
1.1.5. Registration following analysis of the data

11. Explanation of existing data

RR TEAM INSTRUCTIONS: *Sections 10 and 11 section should describe how the reproduction data was used by the reproduction team prior to registration. Since no new data will be created for a reproduction study, 1.1.1 should never be selected. Since all analyses should occur after registration, 1.1.5 should also rarely be selected.*

Note: *When the original study has published summary statistics or analytic data, you are welcome to use it to cross-check your reproduction dataset against it. In cases when you cross-check the reproduction dataset against materials provided by the authors, please disclose that you've done so in the preregistration, and explain any ways in which it affected how you created the reproduction dataset. In general, summarizing single variables or computing approximate sample sizes is fine, but please refrain from running inferential tests prior to registration (e.g. please do not correlate two items to see if they show the expected relationship). There is no obligation to validate your reproduction dataset in this way, and in general can be avoided unless you encounter a serious ambiguity in how the original dataset was constructed.*

Collected data structure reflects the data described in the preprint. Given the state and date referenced cumulative case numbers reported by NYT and state and date referenced measures it is possible to construct all necessary variables needed to implement the analyses.

There are some discrepancies when it comes to descriptive statistics of the collected data in comparison to what was reported in the preprint. For instance, when talking about the implementation of the first measure, Authors report that

"*The most widely enacted measures on the first date of implementation were cancellations of public events (34/51 [67%]) and closures of schools (26/51 [51%])*."

In the reproduction dataset these figures stand at 24 and 26 respectively (with one more extra measure, thus summing up to correct N=51). Since analysis describes 51 states this seems to be a typo in the preprint or the conceptualization of measure differs from what can be inferred from the preprint.

Similarly the measures of the state of the epidemic when measures were implemented are not possible to be replicated - neither for the first implemented measure and for lockdown (see sections 3.1. And 3.2 (bottom) of the 02-data-sourcing.html file for details). For instance Authors specify that:

The first social distancing measures were implemented when the median epidemic size was 36 cases (interquartile range [IQR],17-72).

In the prepared dataset these figures are 60 and 43-176 respectively.

This might arise from vague descriptions of how, for which state and at which stage of epidemic these summaries were calculated, how missing data was handled and what were inclusion & exclusion criteria for these summary stats.

12. Data collection procedures

RR TEAM INSTRUCTIONS: *Please describe the process for constructing the reproduction dataset in as much detail as you can. The sections below should be used to provide the following information:*

- *Which variables are needed from the original study to perform a good-faith, high-quality reproduction.*
- *Which data sources were used and the procedures to access the data, including differences in dataset versions between the original data and the reproduction data, if they are known.*
- *Which items from the reproduction data sources were used to create the reproduction variables.*
- *The procedure for creating the reproduction dataset, in both narrative and script form.*
- *A data dictionary that documents each variable included in the reproduction dataset.*
- *Whether there are known deviations between the data used in the original study and the reproduction dataset you've prepared.*

In the sections below, please provide links to the source materials whenever possible -- including descriptions of the source datasets and corresponding codebooks. If materials can be shared on the OSF, please do so, and provide view-only links to those materials.

Specific points to keep in mind for reviewers:

- *Does the preregistration describe which data sources were used in the reproduction study?*
- *Does the preregistration make clear how the data sources were used to construct the reproduction dataset?*
- *Does the preregistration describe any known deviations between the data used in the original study and the reproduction dataset they have prepared, including known deviations in the dataset(s) version?*

From the **nyt** dataset we extracted the variables *date, state and cases*.

Timing of the state's measures was extracted from the table in supplement. `extract_tables` function of the `tabulizer` package in `R` software environment was used to convert pdf appendix tables into R data frame. Short letter codes of implemented measures were labelled using full name using the information form the appendix table footnote. Dataset was converted to 'long format' with one state level intervention per line to allow for easier data linkages further on.

Data is described in the final data dictionary linked below.

02-data-sourcing.Rmd script in the osf project documents all steps performed using R programming language. 02-data-sourcing.html file presents these steps together with output and narrative in form of a literate programming report. View only versions of the files are available on osf https://osf.io/kh93w/?view_only=1fc936b2a9114163862dd96ce7f11da4

(a) Data Needed

RR TEAM INSTRUCTIONS: *List below the variables the original authors used to analyze the focal claim. For each variable, list the data source it is contained in and (as relevant) the specific wave/year used. As needed, include additional information about each variable that would be relevant for a data analyst or external reviewer. Please include page numbers when excerpting from the original article. If possible, categorize the list of variables as one of the following: dependent variable, focal independent variable, control variable, or sample parameters/clustering variable. Finally, at the bottom of the section, include the sample size of the original study's focal analysis, if it is available.*

Dependent Variable(s)

(names as referred in the final table presenting results in the preprint)

Authors define the outcome of the analyses as:

the COVID-19 growth rate,calculated as the log of daily COVID-19 cases minus the log of daily COVID-19 cases on the prior day

Daily growth rate

- NYT Covid database repository
- Calculated as the log of daily cumulative COVID-19 cases minus the log of daily cumulative cases on the prior day

Focal Independent Variable(s)

Authors define their primary exposure as:

time, measured as a continuous variable, in relation to implementation of the first statewide social distancing measure

Time (days relative to intervention)

- Created from the date of measurement of the number of covid casesand the date of measure implementation as collected in the appendix of the preprint
- Either first implemented measure in the first analysis or lockdown in the second analysis
- Collated from multiple sources

Control Variable(s)

Post intervention period

- Binary indicator dividing time (above) into two periods: pre- (14 days prior to three days after) and post-intervention (four or more days)
- Sensitivity analyses were done with various levels of that indicator modifying various incubation times (three days mentioned above)

Time x post intervention period 'Time' and 'Post intervention period'

- Interaction variable created from 'Time' and 'Post intervention period' variables
- Does not need to be created per se since most of the statistical packages would allow creating it on the fly during the analysis

State

- Level variable of mixed model
- US state reported in the Author's data and nyt dataset
- Used also to link datasets and derived final datasets for the analyses

Sample Parameters

Sample size of analysis is unknown due to varying observation periods for each state. The unit of analysis is state-date tuple.

(b) Data Access

RR TEAM INSTRUCTIONS: *Describe below the data sources that will be used to produce the reproduction variables. Each of the data sources listed in section 12a above should be included in this section. For each data source, include information such as its name (e.g., Indonesian Family Life Survey), the description of and a link to the data source, and the waves needed to create a final reproduction dataset. For data sources that provide version numbers or other identifiers, please document which version of the data you have accessed and whether it deviates from the version used in the original study, if the original version is known. If the version used in the original study can be accessed instead, please do so.*

Also describe the process for accessing the data sources that will be used to create the final reproduction dataset (e.g. did they require registration; what information had to be provided); and verify that the data can be opened as expected. If applicable, provide a link to the page where you registered to access the data.

Describe in detail any restrictions on data access and data-sharing, as well as any additional terms of data use that will be relevant for the reproduction study and final report, ***including***

***citations that will need to be made**. If you were able to access the data because of special permissions that you have, but that you expect other researchers might not have, please document those as well.*

The New York Times COVID-19 database is available as github repo (https://github.com/nytimes/covid-19-data). The dataset is openly available and has been released under the license that is co-extensive with the Creative Commons Attribution-NonCommercial 4.0 International license. The repository doesn't provide any details regarding the preferred citation and a standard description with url and date of access should then be used.

The appendix of the preprint is freely available on the medRxiv repository (https://www.medrxiv.org/content/medrxiv/suppl/2020/04/08/2020.04.03.20052373.DC1/2020.04.03.20052373-1.pdf). Care must be t aken to use the correct version of the appendix matching the version chosen for SCORE replication. Copyright section of the preprint specifies that *the copyright holder for this preprint is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license*.

It is safe to assume that in both cases once the attribution and non commercial requirements are fulfilled it is safe to reuse the data.

(c) Variable Availability

RR TEAM INSTRUCTIONS: *For each variable required for the reproduction analysis (listed in section 12a above), describe the items from the data sources that will be used to measure it, including which data files or sources the item is found in. As relevant, please document any known deviations between the original variable and the reproduction variable due to data availability or version differences.* ***Please also include any notes a data analyst should consider when using the items in the reproduction analysis.***

If there are multiple items in a data source that could be used to measure a required variable (e.g. two different plausible items for measuring education), include all of those options below. If a variable from the original study ***cannot*** *be measured using the reproduction data sources, please make that clear as well.* ***Finally, include a description of the identifiers used to merge multiple datasets, if applicable.***

log_diff

- Item from original data source: nyt.Rds
- File name: nyt_first.Rds & nyt_lockdown.Rds

- Daily cumulative cases are listed in the NYT github dataset (variable **cases**); lagged variable can then be created created from daily cases (**cases_daybefore**); logs (**cases_ln** & **cases_daybefore_ln**) and difference (**log_diff**) are then readily available as simple operations

state

- Item from both original data sources: nyt.Rds & measures.Rds
- File name: nyt_first.Rds & nyt_lockdown.Rds
- US state reported in the Author's data and nyt dataset
- Used to link datasets and derived final datasets for the analyses

time_to_intervention

- Item from original data source: measures.Rds
- File name: nyt_first.Rds & nyt_lockdown.Rds
- Derived from the date of the first intervention in **appendix** dataset from **the preprint** and date of the cases counts of the NYT dataset (above)

post_intervention

- Item from original data source: NA (derived from the two datasets)
- File name: nyt_first.Rds & nyt_lockdown.Rds
- Uses **time_to_intervention** variable described above, recategorized into binary indicator

time_X_post

- Item from original data source: NA (derived from the two datasets)
- File name: nyt_first.Rds & nyt_lockdown.Rds
- Interaction of the two variables above: **time_to_intervention** and **post_intervention**

(d) Data Creation

RR TEAM INSTRUCTIONS: *Create a dataset using the data sources and items listed above. Provide a detailed narrative describing how the various datasets were cleaned and merged into a final reproduction dataset. Provide a view-only link to a clearly commented script on the OSF that produces the reproduction data as described in the narrative. Our preference is that this be either an R script or a script from another language that similarly allows for open and reproducible analyses. Please let the SCORE team know if this is not possible.*

- *If the data can be freely shared and posted to OSF, please post it in your OSF project and provide a link to the completed dataset below.*

- *If any part of the dataset cannot be shared between researchers or posted to the OSF, please leave the final dataset off the OSF.*

The data from the original sources should be preserved in as 'raw' a form as possible, in order to give the data analyst the most latitude to clean the variables as they see fit. Variables from the original source should be preserved in their original form (e.g. do not recode values of 99 to NA). In general, new variables should only be created when they're needed to complete the merge or combine the datasets; in those cases, please preserve a version of the original, unaltered item(s) in the reproduction dataset whenever possible.

Please also use this section to describe:
- *Any known deviations between the original study data and the reproduction dataset.*
- *Any notes about using these variables that you would like to pass along to the data analyst, if they are not already documented in section 12c above.*

First, **nyt** data are first limited to the 50 states plus DC - that was achieved by excluding `fips` codes `66, 69, 72, 78` for Guam, Northern Mariana Islands, Puerto Rico and Virgin Islands.

Next, data is limited to match the time frame of the analysis ie. until `2020-03-30` and daily cumulative cases are used for each day & state. Next cumulative cases from each day and cumulative cases from the previous day logarithms and their differencescalculated. Finally only days when a cumulative number of cases reached 30 are kept in the dataset.

First implemented measure is selected using the date of implementation from the **measures** dataset. It is linked to **nyt** dataset using the name of the state as a key. Date of measuring the counts of cases and date of first intervention are first used to calculate time to intervention. Finally this last variable is recoded to binary indicator or pre- post-intervention using the specification described above.

The script (02-data-sourcing.Rmd) to implement this data preparation can be found on osf https://osf.io/kh93w/?view_only=1fc936b2a9114163862dd96ce7f11da4

(e) Data Dictionary

RR TEAM INSTRUCTIONS: *Create a data dictionary following this template. Provide upload the complete data dictionary to the OSF and include a view-only link to it below. If the Data Analyst will need to create new variables using the items in the final reproduction dataset (e.g. recoding the provided education variable to be in a better format for analysis), please document below your recommendation on how the analyst should do so unless that is already documented elsewhere in the preregistration. Please also document any additional notes regarding the items in the dataset that do not fit within the provided data dictionary template or the other sections above.*

Data dictionary has been uploaded to the osf project.

(f) Known Deviations

RR TEAM INSTRUCTIONS: *Please use this section to document known deviations between the data used to test the selected claim from the original study and the dataset you have prepared for the reproduction analysis. If already documented above, please summarize them here as well. Please explain how the deviation was discovered (e.g. cross-checked against summary statistics in the original study) and what effect it might have on the reproduction analysis.*

There are no known deviations between data used by Authors and that used by the replication team.

13. Sample size

RR TEAM INSTRUCTIONS: *Please report below the sample size of the focal analysis selected from the original study, as well as the approximate sample size of the reproduction analysis based on the reproduction dataset constructed above. Please report as much information here as will be helpful for the review committee to be aware of, including differences in sample size resulting from various analytic decisions (e.g. listwise deletion vs multiple imputation).*

Original study' sample size and units: The exact number of sample size is not reported. The Authors used 50 US states and DC for the analyses. These could have various number of data points depending on the date of the implementation of the measure. Such variability can be glanced from the Figure 1A-B of the preprint where it is clear that various states start and end data series at different dates.

Approximate sample size and units of the reproduction analysis: For the reproduction analyses we prepared the dataset of the 770 records of 51 states from 2020-03-01 to 2020-03-30 period.

There are no missing values in such a dataset.

All 51 states are represented in the measures and will therefore be able to be linked to the NYT dataset of cases.

14. Sample size rationale (provided by SCORE)

This section is not relevant to a reproduction analysis.

15. Stopping rule (provided by SCORE)

This section is not relevant to a reproduction analysis.

Variables

RR TEAM INSTRUCTIONS: *The preregistration form divides variables across three questions: manipulated variables, measured variables, and indices (i.e. analytic variables derived from raw items). For source data reproductions, only fill out the "Measured variables' and 'Indices' sections. Please do not fill out anything in the 'Manipulated variables' section.*

The raw data of any transformed variable (e.g. reaction time → log reaction time) or any created index should be defined in the 'Measured variables' section. Details regarding the variable transformation should be specified in the 'Transformations' section. Details regarding the creation of an index should be specified in the 'Indices' section.

Across these questions, you should define all variables that will later be used during your analysis (including data preparation/processing). You can describe all variables in the preregistration and/or summarize and link to a data dictionary (codebook) in your repository to answer these questions.

If you will share data from your reproduction, this is also the place to state whether any variables will be removed prior to sharing the dataset (e.g. to reduce risk of participant identification or comply with copyright restrictions on scale items.)

16. Manipulated variables

RR TEAM INSTRUCTIONS: *Manipulated variables in this preregistration refer specifically to variables that have been randomly assigned in an experiment. We expect source data reproductions of experiments will be rare. If your source data reproduction relies on experimental data, please document each manipulated variable as a measured variable, and use the codebook to indicate what each level of the variable corresponds to (e.g. participants assigned to the treatment condition = 1; participants assigned to the control condition = 0). The default language in bold below has been copied into all source data reproduction preregistrations.*

N/A -- not documented for source data reproductions.

17. Measured variables

RR TEAM INSTRUCTIONS: *Please use this section to document each variable that was used in the original study's analysis and the role it served (e.g. dependent variable, control variable, sample parameter, etc). For each variable, provide the description of the variable offered in the paper and/or codebook of the original study, the item(s) in the reproduction dataset that it corresponds to, and explain any deviations between the two. In cases where an equivalent*

reproduction variable could not be found or accessed, explain how, if at all, you expect its absence will affect the reproduction attempt.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration document all of the variables needed to reproduce the focal analysis?*
- *Are deviations between the original variables and reproduction variables documented when needed?*

VARIABLE NAME

- [Use in the analysis]
- [Description from the original study]
- [Variables used in the reproduction (if it needs to be constructed from multiple items, include all of them here)]
- [Deviations between the original study and the reproduction study]

log_diff

- Outcome of the regression model
- Described as: *the COVID-19 growth rate,calculated as the log of daily COVID-19 cases minus the log of daily COVID-19 cases on the prior day*
- Calculated from cumulative cases from the NYT github dataset (variable cases); lagged variable can then be created from daily cases (cases_daybefore); logs (cases_ln & cases_daybefore_ln) and difference are then readily available as simple operations

time_to_intervention

- Main exposure of the regression model
- Described as *time, measured as a continuous variable, in relation to implementation of the first statewide social distancing measure*
- Derived with simple arithmetics from the date of the first intervention in appendix dataset from the preprint (measures dataset) and date of the cases counts of the nyt dataset

post_intervention

- Exposure of the regression model
- Described as an indicator of: *two periods: pre-implementation(14 days prior, through three days after, implementation of the first statewide social distancing measure) versus post implementation (four or more days after implementation).*
- Uses time_to_intervention variable (above), recategorized into binary indicator

time_X_post

- Exposure of the regression model
- Interaction of the two variables above: **time_to_intervention** and **post_intervention** - will be created automatically during the analyses

state

- Level variable of mixed model
- US state reported in the Author's data and nyt dataset

18. Indices

RR TEAM INSTRUCTIONS: *If any of the measured variables described in Section 17 will be combined into a composite measure (including simply a mean), describe in detail what measures you will use and how they will be combined. Please be sure this preregistration includes a link to a clearly commented script that constructs the index according to the narrative.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify each of the composite measures (e.g. mean scores, factor scores) that are needed for the focal analysis, and which of the measured variables in Section 17 are used in each one (e.g. the happiness, joy, and satisfaction items will be used to create the 'positive feelings' measure)?*
- *Does the preregistration link to a clearly commented script that constructs the indices according to the narrative description?*

NA unless population density is treated as index and sensitivity analyses implemented.

Analysis Plan

19. Statistical models

RR TEAM INSTRUCTIONS: *This section should describe in detail the analysis that will be performed to reproduce the focal result. This analysis must align as closely as possible with the original study's analysis, even if you have identified limitations in the original study. The level of detail should allow anyone to reproduce your analyses from your description below. Examples of what should be specified: the model; each variable; adjustments made to the standard errors and to case weighting; additional analyses that are required to set up the focal analysis; and the software used.*

Beyond the reproduction of the focal analysis from the original study, it is at your discretion to test the claim using other analytic approaches as a check of the robustness of the claim. The original test should be listed first and be clearly distinguished from any other tests. If you are testing additional confirmatory hypotheses, describe them in the same order as you numbered

them in the "Hypotheses" section above and make clear reference to the specific hypothesis being tested for each.

Please provide a link to a clearly commented script that performs the analysis described in the narrative provided below. Our preference is that this be either an R script or a script from another language that similarly allows for open and reproducible analyses. Please let the SCORE team know if this is not possible. Please also test that the code runs without error on a random subset of 5% of the reproduction dataset, and provide verification that the code has produced a sensible result below by providing a screenshot of the output (please upload the screenshot to the OSF as well). Finally, please confirm that you have only developed and tested your analysis plan and code using 5% of the data.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify which statistical model will be used to provide the 'focal evidence' for the SCORE test (e.g. a regression coefficient in a larger multiple regression model), and does it correspond closely to the model and evidence from the original study?*
- *Does the preregistration describe each variable that will be included in the focal analysis, and what role each variable has (e.g. dependent variable, independent variable)?*
- *Does the preregistration include a detailed specification of the focal analysis, including interactions, lagged terms, controls, etc., in both narrative form and in a clearly commented script?*
- *Does the preregistration verify that the code runs without error on a random subset of the reproduction dataset?*

(a) Model description, analysis code, and test verification

The claim was tested by fitting mixed effects linear regression models, specifying the log difference in daily cases as the outcome of interest and a random effect for state. Explanatory variables included time in days, implementation period, and a time-by-implementation period product term. Code provided by the Authors indicates that Stata of unknown version was used with few user written commands.

We will analyse prepared datasets using R software environment (version 4.0.3). Mixed effects models are not part of the standard R implementation and are usually implemented via an external package. There are multiple solutions possible and authors of preprint do not give any hints about that. Although they might differ in details, the choice of the package should not have an impact on directions and strength of the association in case of such a simple model. We will use the `lme4` package (version 1.1-25) which is a standard and established way of working with mixed models in R.

The model will use the `log_diff` variable as outcome and will be adjusted for `time_to_intervention`, `post_intervention` variables and their interaction term. Random effect will be specified on the `state` variable using `lme4` package specification (ie.` `(1 | state)`). Interaction effects will be created automatically during the analysis using the form `time_to_lockdown : post_intervention`.

Two almost identical analyses will be performed using two datasets: `nyt_first` & `nyt_lockdown` in order to test the effects of first measure and full lockdown respectively.

We will use a 5% sample of states to test the code. Having 51 states in the dataset that would amount to using three states rounding up to achieve slightly bigger rather than lower sample in order to have more stability in mixed models

All code is provided in literate programming set up of R markdown in a file `03_data-analysis-sample.Rmd` with output presented in `03_data-analysis-sample.html`. And available on osf https://osf.io/wrb36/?view_only=a59f80b2f5094abdbf00ac664a984126

(b) Known Deviations

RR TEAM INSTRUCTIONS: *Please use this section to document known deviations between the analysis used to test the selected claim from the original study and the analysis documented above. Please explain why these deviations are necessary and what effect they may have on the reproduction outcome.*

There are no deviations from the planned analysis using the sample data.

(c) Confirmation of random subset

RR TEAM INSTRUCTIONS: *Please bold the statement below to indicate your confirmation.*

This statement confirms that only 5% of the data have been randomly sampled in developing the analysis plan and code contained in this preregistration.

20. Transformations

RR TEAM INSTRUCTIONS: *This section should describe how any of the measured variables or composite measures mentioned above will be transformed prior to the analyses listed in Section 19. These are adjustments made to variables **after** measurement or measure creation, and might include centering, logging, lagging, rescaling etc. Please provide enough detail such that anyone else could reproduce the transformations based on the description below. Please be sure this preregistration includes a link to a clearly commented script that performs the transformations described in the narrative provided below.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify which of the measured variables or composite measures will need to be transformed prior to the focal analysis?*
- *For each variable needing transformation, does the preregistration adequately describe the transformations, including any centering, logging, lagging, recoding, or implementation of a coding scheme for categorical variables?*
- *Does the preregistration link to a clearly commented script that performs each transformation?*

log_diff

- calculated from cumulative cases from the NYT github dataset (variable `cases`); lagged variable can then be created created from daily cases (`cases_daybefore`); logs (`cases_ln` & `cases_daybefore_ln`) and difference are then readily available as simple operations

21. Inference criteria

RR TEAM INSTRUCTIONS: *This section describes criteria that will be used to assess whether the hypotheses listed above were confirmed by the analyses in Section 19. The default language below only applies to the test of the SCORE claim, **H***. It is at your discretion to describe the inferential criteria you will use for any additional analyses.*

If the additional analyses will use multiple comparisons, the inference criteria is a question with few "wrong" answers. In other words, transparency is more important than any specific method of controlling the false discovery rate or false error rate. One may state an intention to report all tests conducted or one may conduct a specific correction procedure; either strategy is acceptable.

Inference criteria for a reproduction will be based on a subset of the following criteria:

- Sample size
- Focal variable p value
- Control variable(s) coefficient(s) in focal model
- Focal variable coefficient value
- Effect size of focal variable
- Subjective interpretation by analyst

Accordingly, the reproduction analysis will report outcomes for each of these criteria whenever possible and relevant.

Criteria for a successful replication attempt for the SCORE project is a statistically significant effect (alpha = .05, two tailed) in the same pattern as the original study on the focal hypothesis test (H*). For this study, these criteria are met by obtaining a statistically significant (p-value

specified in the preprint is 0.002 for first measure implemented) negative regression coefficient on the interaction between time and post implementation period from the unadjusted model run on the full sample of states (reported coefficient: -0.008; -0.014, -0.002 95%CI).

22. Data exclusion

RR TEAM INSTRUCTIONS: *The section below should describe the rules you will follow to exclude observations from the analyses described in Section 19. Note that this refers to exclusions **after** the creation of the reproduction dataset; exclusion criteria that prevent a case from entering the reproduction dataset in the first place should be detailed in the 'Data Collection Procedure' section above. Please be as detailed as possible in describing the rules you will follow (e.g. What is the specific definition of outliers you will use? Exactly how many attention checks does a participant need to fail before their removal from the analytic sample?).*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on whether any observations included in the reproduction dataset will be excluded prior to data analysis?*
- *If yes, does the preregistration provided detailed instructions on how the exclusions will be performed (e.g. Is the definition of outlier provided? Is the number of attention checks failed before a participant is excluded specified?)*

As per preprint specifications only days when cumulative number of cases reached 30 are kept in the dataset.

23. Missing data

RR TEAM INSTRUCTIONS: *The section below should describe how missing or incomplete data will be handled. Please be as detailed as possible in describing the exact procedures you will follow (e.g. last value carried forward; mean imputation) and any software required (e.g. We will use Amelia II in R to perform the imputation).*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on how missing or incomplete data will be addressed (e.g. casewise removal, missing data imputation)?*
- *If applicable, does the preregistration specify how many missing variables will lead to a case's removal (e.g. If a subject does not complete any of the three indices of tastiness, that subject will not be included in the analysis.)?*
- *If applicable, does the preregistration describe how missing data imputation will be performed, including relevant software?*

24. Exploratory analysis (Optional)

RR TEAM INSTRUCTIONS: *If you plan to explore your dataset to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test*

where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time. If any exploratory analyses involve additions to the data collection procedure beyond what was performed in the original study (e.g. additional items on the survey; running another condition in the experiment), please describe them below.

We will graph values of `cases_day, cases_day_percent` and `log_diff` across days of analysis for the sampled state for visual inspection.

25. Other

RR TEAM INSTRUCTIONS: *This section serves two purposes. First, please use this section to discuss any features of your reproduction plan that are not discussed elsewhere. Literature cited, disclosures of any related work such as replications or work that uses the same data, plans to make your data and materials public, or other context that will be helpful for future readers would be appropriate here. Second, please also re-surface any major deviations from earlier in the preregistration that you expect a reasonable reviewer could flag for concern. Give a summary of these deviations, focusing on larger changes and any possible challenges for comparing the results of the original and the reproduction study.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration reference other sections of the preregistration where substantial deviations from the original study have been described?*
- *Does the preregistration comment on plans to make the data and materials from the reproduction study public?*

Final review checklist

REVIEWER INSTRUCTIONS: *Reviewers -- for the following questions, please indicate whether you can 'sign off' on the following items by adding a comment. You can update this response as the lab moves through revisions during the review period!*

- Included in this preregistration are specific materials needed to create a reproduction dataset:
    - Is the final reproduction dataset that the research team constructed suitable for performing a high-quality, good-faith reproduction of the focal claim selected from the original study?
    - Is the procedure for constructing the final reproduction dataset sufficiently documented that an independent researcher could construct the same dataset following the procedures and code they provide?
- Included with this preregistration is a narrative description of how the reproduction dataset will be used to perform the focal reproduction analysis, as well as the specific analytic scripts/code/syntax that will be used:
    - Is the analysis plan (including code) that's documented in the preregistration consistent with a high-quality, good-faith reproduction of the focal claim selected from the original study?
    - Has the data analyst demonstrated that the analysis code works as expected on a random 5% of the final reproduction dataset?
- I have reviewed all sections of this preregistration, and I believe it represents a good-faith reproduction attempt of the original focal claim.