Replication of a Research Claim from Malik et al. (2020),
from medRxiv

Replication Team: Sara Capitán, Tabaré Capitán, and Anna Fedor

Research Scientist: Nick Fox

Action Editor: Ernest O'Boyle

Independent Reviewers
(add name below when you initiate review, comment "DONE" on your name when you finish):

Reviewer #1: [Sam Smith DONE]

Reviewer #2: [NAME]

Reviewer #3: [NAME]

Review Period: November 2 - November 9

View-only links to: Original Paper, Original Materials, Replication Data, Replication Analysis

Privacy Statement: Other teams are making predictions about the outcomes of many different studies, not knowing which studies have been selected for replication. 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 replication and all aspects of the discussion about these replication designs.

Instructions for Data Analysts

The preregistration for this replication study was started by a separate team of researchers who were responsible for identifying data sources and constructing them into a replication 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 secondary data replications, 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.

Finally, in cases where the replication data combines observations from the original study with observations that were not used in the original study (what we are calling 'hybrid replications'), please perform up to three analyses (details immediately below). This will likely require you to subset your data, based on the description of the original analysis provided in the study.

- When the 'new' data alone can clear the minimum power threshold, please perform one analysis that relies only on the 'new data' (the focal replication analysis), one analysis that relies on all available data, and a third analysis that relies only on the original data (the focal reproduction analysis). Please make sure all three analyses are documented (with code) in section 19 below.
- When the 'new' data alone *cannot* clear the minimum power threshold, please perform one analysis that combines all available data (the focal replication analysis), and a second that only uses the old data (the focal reproduction analysis). Please make sure both analyses are documented (with code) in section 19 below.

Please contact Andrew if you have any questions. After you've completed the remaining sections of the preregistration and uploaded all the necessary materials to the OSF, please contact the SCORE coordinators regarding next steps.

Preregistration of Malik_covid_wx5k

Existing Data Replication

Study Information

1. Title (provided by SCORE)

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

Replication of a research claim from Malik et al. (2020).

2. Authors and affiliations

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

Sara Capitán, Data Finder[1]
Tabaré Capitán, Data Finder[2]
Anna Fedor, Data Analyst[3]

1 Independent researcher
2 University of Wyoming
3 Institute of Evolution, Centre for Ecological Research, Budapest, Hungary

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 replication from Malik et al. (2020) is that social distancing measures introduced by governments reduces the mobility in the concerned cities. This reflects the following statement from the paper's abstract: "Social distancing measures decreased the mobility by an additional 23% (95%CI: 20%, 27%)". A city was classified to have instituted social distancing measures if non-essential businesses were closed; these measures were further classified as moderate or intense based on the intensity of closure. The effect of time and social distancing measures was estimated using a multilevel mixed-effects linear regression model. Social distancing measures decreased the mobility by an additional 23% (95%CI: 20%, 27%).

Malik et al. claim that “There was no difference in mobility reduction based on whether the measures were moderate or intense”, but we do not plan to replicate this claim, i.e. the intensity of the closures will not be used in the replication as a variable.

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 replication/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 introduction of social distancing measures is associated with a decrease in mobility.
H1: The introduction of social distancing measures is associated with an increase in people staying at home.

Design Plan

5. Study type

NOTE: *The study type selected should be based on the data collected for the replication, and not necessarily the data used in 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.

7. Blinding

RR TEAM INSTRUCTIONS: *Since all existing data replications are based on data that has already been collected, in most cases it will not be necessary to comment on participant blinding. In the rare instance when an existing experiment is being re-analyzed for an existing data replication and blinding is a relevant consideration, please provide below any details regarding blinding that are important for a reviewer to be aware of.*

8. Study Design

RR TEAM INSTRUCTIONS: *Please describe how data was collected in the original study and how it compares to the data that was selected for the replication attempt. Explain why the data selected for the replication study is suitable for a replication and if any substantial deviations exist between the two.*

If the data used in the replication combines observations from the original study with new observations (e.g. if the data selected for the replication attempt comes from the same longitudinal survey as the original study), describe how ‘original’ and ‘new’ observations relate to each other and an estimate for what proportion of the final dataset’s observations will be comprised of original vs. new observations.

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 how the data selected for the replication attempt deviates from or is congruent with the data employed in the original study?*
- *Does the preregistration describe whether and how ‘original’ and ‘new observations’ are combined together for the replication dataset?*

This preregistration includes all required information, but we also include a document with an easier structure to follow that fully describes the replication dataset (here).

Data source 1: The authors of the original study downloaded data from Citymapper.com. The data is from the Citymapper Mobility Index from March 2, 2020 to March 26th, 2020. In the analysis, they use 1,025 observations across 41 cities (25 observations per city).

We propose to use the Google Community Mobility Reports (available here), which includes mobility reports in 6 categories, one of which is transit stations (the analyst should use the transit variable, but we include the remaining to allow for additional analysis). Another category, "residential" will be used to investigate H1.

The Citymapper Mobility Index (hereafter CMI) and the COVID-19 Community Mobility Reports: Transit (hereafter CMRT) are alternative measures of the same conceptual variable: the change in people's mobilityduring the pandemic relative to "normal" times before the pandemic. However, both measures operationalize the conceptual variable in different ways that reflect differences in their primary data sources, baseline definition, and unit of analysis (geographic level).

First, the CMI is calculated using primary data from planned trips in the Citymapper app, and the CMRT is calculated using primary localization data from Google users in transit stations. Second, Citymapper defines the baseline for each day based on their data during the 4-week period between Jan 6th and Feb 2nd, 2020. But they adjust the baseline in Paris (Feb 3rd to March 1st) and Hong Kong and Singapore (both Dec 2nd to Dec 22nd) to reflect that these cities were exposed to the impact of the pandemic earlier than most. On the other hand, Google defines the baseline as a normal value for that day of the week, in which normal is defined as the median value from the 5-week period Jan 3rd to Feb 6th, 2020. Google does not adjust the baseline for any region. Third, Citymapper reports data at the city level and Google reports data at different levels depending on their data availability and privacy policies. As a result, Google's region is sometimes larger than the corresponding city in Citymapper, and sometimes smaller.

The first difference – primary data sources – implies that the indexes may be reflecting the behavior of different people if there is no significant overlap between those who use Citymapper and those who use Google services. Despite the lack of available data to test the overlap, we believe that this issue should not be a major concern since most phones run on Android. Moreover, since the original claim is general enough, it should be true for a different group of people too. A second implication regarding the different primary data sources is that Citymapper might be capturing planned trips and Google is capturing realized trips. We again do not have data to test, but Citymapper mentions that, as expected, there is a correlation in their data between planned trips and realized trips (unfortunately, they do not indicate the strength of this correlation).

The second difference – baseline definition – implies that the usage of transit stations during the pandemic is compared to a different measure of "normal" times, leading to differences in the measure of change. We believe this difference is not a major concern because Malik et al. (2020) analyze the data using regression analysis, which is driven by variation in the data instead of levels.

The third difference – unit of analysis – is our main concern because we are unable to perfectly match each city in the CMI with the exact region in the CMRT. However, most of the matches are reasonable. Our strategy to match cities is to use the region in the CMRT dataset closest to the city in the CMI, that is, we use the CMRT data to proxy the (smaller) region reported in the CMI. Our rationale is as follows. Google's CMRT likely contains a weighted average of mobility within the region they report. For example, within a county, they would report a weighted average of mobility in the cities within that county. Since the CMI focuses on highly populated cities (i.e., the city should have a big weight in the Google's mobility report), the data generating process for the data in the CMRT would likely be driven by mobility in the city and so the data from the CMRT would be a good proxy of the CMI data. After matching the original and the proposed dataset, we believe that we get a good proxy for most cities except the two in Russia, the one in South Korea, and Monaco. Still, we kept Moscow and Seoul, because these are the most populated cities of Russia and South Korea, respectively, and we assumed that the country data provided by Google proxies these cities. We excluded St. Petersburg and Monaco from the proposed dataset. Appendix 1 describes each match.

After matching the new data with the original, we are left with 975 observations across 39 cities (25 observations per city). We could easily extend the mobility data until today, but since the original paper is about the short-term reaction to the initial social distance measures, we do not think it is appropriate.

Data source 2: As opposed to mobility data, there is no alternative data on lockdowns. Here we rely on the same data collected by the authors because there are no alternative measures of the conceptual variable lockdown. The authors of the original paper describe this variable like this: "We classified a city to have instituted social distancing measures if non-essential businesses were closed; these measures were further classified as moderate or intense based on the intensity of closure". We have obtained the original data (and their primary sources) from the authors and merged it with the new measures of mobility. We only used the variable of "partial lockdown" for the replication. This is a binary variable which takes the value of 1 if any kind of lockdown (moderate or intense) was in place on the given day in the given city, and 0 otherwise.

9. Randomization (free response)

RR TEAM INSTRUCTIONS: *If the variables used for this replication attempt were randomized, state how they were randomized, and at what level.*

Sampling Plan

This section describes how the data sources for the replication were selected, how they were prepared into a replication dataset, and the number of observations that will be analyzed from these data. 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 (multiple choice question, provided by SCORE)

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

NOTE: *For replications that rely on existing data sources, this question refers to the data that will be used for the replication analysis (i.e. the final replication dataset), and not (a) the data from the original study or (b) the data sources accessed to construct the replication dataset. Since no new data will be created for 'existing data replications,' 1.1.1 should never be selected. Since all analyses will occur after registration, 1.1.5 should also never be selected.*

Mobility data: We propose to use the Google Community Mobility Reports (available here), which includes mobility reports in 6 categories, one of which is transit stations (the analyst should use the transit variable, but we include the remaining to allow for additional analysis). After matching the new data with the original, we are left with 975 observations across 39 cities (25 observations percity). See more details in section 8.

Social distance measures: We rely on the same data collected by the authors because there are no alternative measures of the conceptual variable lockdown. We have obtained the original data (and their primary sources) from the authors and merged it with the new measures of mobility.

12. Data collection procedures

RR TEAM INSTRUCTIONS: *Please describe the process for constructing the replication 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 replication.*

- *Which data sources were used, why they were selected, any deviations between the original study design and the replication study design that these selections present, and the procedures used to access the data.*
- *Which of the variables from the original study are available in the replication data sources, including relevant details about each measure.*
- *The procedure for creating the replication dataset, in both narrative and script form.*
- *A data dictionary that documents each variable included in the replication dataset.*

In the sections below, please provide links to the original materials whenever possible -- including descriptions of the original 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 selected for the replication study and why each is suitable?*
- *Does the preregistration make clear how the data sources were used to construct the replication dataset?*

(a) Data Needed

RR TEAM INSTRUCTIONS: *List below the datasets and variables the original author used to analyze the focal claim. Include details regarding the sample size, waves or years used, and other details pertinent to finding an existing dataset for replication. 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, include the sample size of the original study's focal analysis, if it is available.*

Dependent Variable(s)

% Moving: City mobility index
- Variable Data Source
  - https://cdn.citymapper.com/data/cmi/Citymapper_Mobility_Index_20201105.csv
- Waves / Years
  - March 2 - March 26 (2020)
- Other Information
  - Time-series for 41 cities (25 days per city, total 1025 observations)

Focal Variable(s)

Lockdown: (Binary) Social measures implemented
- Variable Data Source

    - Constructed dataset of measures instituted in each city. For each city and for each day, the variable takes the value of 1 if lockdown measures (moderate or intense) were in effect during that day in that city, and 0 otherwise.
- Waves / Years
    - March 2 - March 26 (2020)
- Other Information
    - The authors collected the information from official websites and the media (see data request email here).
    -

Sample Parameters

- Sample size of analysis has 1025 observations across 41 cities.
    - The original data (citymapper) is on the city level, but the proposed replication dataset (Google mobility report) has mobility data on various levels (usually a country / sub-region level). We were able to match 39 of the 41 cities in the original data to a corresponding geographic area in the new data. As expected, we do not find perfect matches since google does not report data by city. However, most of the matches are reasonable. For example, in the US, we can always match the city (which is always the most populated within a county) to its corresponding county. The same is true for the corresponding administrative regions for cities in most countries.

(b) Data Access

RR TEAM INSTRUCTIONS: *Describe below the data sources that will provide the replication variables. Include information such as the name of the data source (e.g., Indonesian Family Life Survey), the description and link of the data source, and the waves needed to create a final replication dataset.*

Also describe the process for accessing the data sources that will be used to create the final replication dataset; specify how long it took for the registration to be approved and what information was required (e.g., writeup of the purpose of the project, email address from an IPCSR institution, etc.); 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 replication study and final report (e.g. 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 data source 1 (Google Mobility Report) can be freely downloaded from https://www.google.com/covid19/mobility/index.html?hl=en right away. The data source 2 (list of cities with measures and related intensity) can be either obtained from the authors via email (as described in here), or manually visiting media and governmental sources to collect the same data as the original authors. We have included the dataset used by the authors.

(c) Variable Availability

RR TEAM INSTRUCTIONS: *For each variable required for the replication analysis (listed above), describe the variables from the replication data that can be used to measure it (including which data files or sources each measure is found in),* ***any notes a data analyst should consider when using the measure in a replication analysis****, and any important differences between the original variable and the proposed replication variable.*

If there are multiple variables in the replication data that correspond to a required variable (e.g. two different measures of education in the replication data), include all of those options below. If a variable from the original study ***cannot*** *be measured using the replication data, please make that clear as well.* ***Finally, include a description of the identifiers used to merge multiple datasets, if applicable.***

CMRT_transit

- Uses Google Mobility Index variable in **Google Mobility Report** dataset from **Google Mobility Reports** data source for **the same period of the original study (March 2 - March 26)**
- Description of **Change in visits with respect to visits before COVID19 became widespread** variable. That is, this variable reports the percentage change in visits with respect to the baseline described in section 8.

Lockdown:

- For the lockdown variable there is only one data source and it's the one used by the authors. That is, beyond the official websites, any other source would be a source repeating the official source. That's why we propose to use the original author's constructed data. They defined lockdown like this: "We classified a city to have instituted social distancing measures if non-essential businesses were closed"

(d) Data Creation

RR TEAM INSTRUCTIONS: *Create a dataset using the data sources and variables listed above. Provide a detailed narrative describing how the various datasets were cleaned and merged into a final replication dataset. Provide a view-only link to a clearly commented script on the OSF that produces the replication 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. Instead, include either below or in your script (commented out at the bottom) two pieces of information that will help an independent team verify they have created the dataset according to your instructions:*
    - *The dimensions of the final dataset(s) you've created (# of rows, # of columns)*
    - *A summary of 8-10 variables in the replication dataset. For numeric variables, the summary should include the mean, standard deviation, and count of NAs. For categorical variables, the summary should include each level present in the data and its count, as well as a count of NAs. If multiple datasets are submitted as part of your work, at least one variable should be included from each dataset.*

The data from the replication 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). 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 variable in the new dataset.

When combining multiple datasets by binding rows, please be sure that the data type and measurement units are equivalent across each dataset. If there is a discrepancy in how a variable is measured across datasets, rename the variable in each dataset to indicate the original dataset, and then carefully document the resulting measures below and in the data dictionary. See here for an example of how this should work.

Please also use this section to describe:

- *Any deviations between the original study design and the replication design that would result from using this replication dataset.*
- *Any notes about using these variables that you would like to pass along to the data analyst.*

We used Stata to create the replication dataset (here is the code). The code simply takes the raw data (mobility data and social distance measures data) and merges it. We used the exact same name for the cities as in the original article. The correlation between the original measure of mobility (CMI) and our replication measure (CMRT) is 0.935. We saved the dataset both in stata format (.dta) and CSV.

(e) Data Dictionary

RR TEAM INSTRUCTIONS: *Create a data dictionary following this template. Provide below a view-only link to the completed data dictionary included in the OSF project. If the Data Analyst will need to create new variables using the variables in the final replication 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. Please also document any additional notes regarding the variables in the dataset that do not fit within the provided data dictionary template or the other sections above.

Here is the dictionary.

13. Sample size

RR TEAM INSTRUCTIONS: *Please report below the analytic sample size(s) in the replication dataset, with reference to however many units or levels are in the data. 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).* ***Finally, when the replication combines observations from the original study with new observations, please estimate what proportion of the analytic sample's observations will be comprised of original vs. new observations.***

We obtained the same 25 observations per city corresponding to the 25 days between March 2nd and March 26th used in the original article. However, we were unable to find data for 4 cities in the original data. We recovered 2 out of these 4 cities by assuming that the country-level data is a good proxy for the most populated city of the country (Moscow and Seoul).Therefore, we included 39 cities instead of 41 and we ended up with 975 observations.

Required sample size [to be filled out by the SCORE team]: The primary unit of analysis is the geographic area. An estimate of the minimum viable sample size for the data analytic replication is: 2 [with 25 observations per geographic area]. For comparison, the stage 1 required sample size would be: 4 [with 25 observations per geographic area] and the stage 2 sample size would be: 8 [with 25 observations per geographic area].

14. Sample size rationale

For data analytic replications in SCORE, three sample sizes are calculated:

- *A minimum threshold sample size, defined as the sample size required for 50% power of 100% of the original effect*
- *A stage 1 sample size, defined as the sample size needed to have 90% power to detect 75% of the original effect*
- *A stage 2 sample size, defined as the sample size needed to have 90% power to detect 50% of the original effect*

Details about how those sample sizes were calculated for this project are found here.

15. Stopping rule (provided by SCORE)

RR TEAM INSTRUCTIONS: *For replications and reproductions involving existing data, this section describes which analyses the SCORE team is recommending be performed. Most often, this corresponds to analyses involving new data, original data, or a combination of new and old data.*

Since the replication data does not directly overlap with the original data, the SCORE team recommends that a single analysis that includes all eligible observations be performed. This will be the focal replication analysis of the study.

Note: In this case the data is similar to the original study but not equivalent, since mobility is measured at a different level. Accordingly, there will be no reproduction analysis (i.e., an analysis involving only the original observations), since the study design intentionally deviates from the original study.

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 variables). For existing data replications, 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 replication, 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. The use of data from an experiment should be rare in existing data replications. If your existing data replication 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 existing data replication preregistrations.

N/A -- not documented for existing data replications.

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 variable in the replication dataset that it corresponds to, and explain any deviations between the two. In cases where an equivalent replication variable was not found, explain how, if at all, you expect it will affect the replication attempt. In cases where you are adding a variable that was not present in the original study, please explicitly state that you are doing so, and explain how, if at all, you expect it will affect the replication attempt.*

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

- *Does the preregistration surface all of the variables needed to replicate the focal analysis?*
- *Are deviations between the original variables and replication variables documented when needed?*

VARIABLE NAME

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

City

- Clustering/sample parameter
- City: geographical area of the given city
- city: geographical area (usually at the county or sub-region level) that proxies the given city
- The unit of analysis is city in CMI data and region in CMRT data. The variable "city" corresponds to the best match of CMRT data to CMI data, see sections 8 and 12.

Date

- Sample parameter
- Calendar days

- date2
- The data finders provided the replication data in dta (Stata v16) and csv file formats. The data analyst did not have access to Stata v16. She used Stata v13, which cannot read dta files created by newer versions of the software, so she had to import the csv file to Stata instead of using the dta file. For ensuring that the dates were imported correctly, she took the following steps:
    - Opened the csv file in Excel
    - Changed the format of the entire column of the date variable to show years (e.g., originally it showed 26-Mar, after the change it showed 3/26/2020)
    - Imported the date variable as a string, then converted it to a date type variable (called date2) using the analysis script provided

% Moving

- Dependent variable
- "The CMI is based on planned trips on the Citymapper application"
- CMRT_transit
- In the original study, the dependent variable is the Citymapper Mobility Index (CMI) from the Citymapper app, whereas in the replication the dependent variable is the Community Mobility Report: Transit from Google, see section 8.

Lockdown

- Focal independent variable
- "We classified a city to have instituted social distancing measures if non-essential businesses were closed"
- lockdown
- The lockdown variable was imported without change from the original study.

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?*

Analysis Plan

19. Statistical models

RR TEAM INSTRUCTIONS: *This section should describe in detail the analysis that will be performed to replicate 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 replication 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.

For each analysis specified in section 15 (and particularly the analyses labeled as 'focal'), please test that the code runs without error on a random subset of 5% of the relevant data. When more than one analysis is listed in section 15, this could require separate 5% samples (e.g. a replication sample and a reproduction sample). Please provide verification that the code has produced sensible results by providing a screenshot(s) of the output (please upload the screenshot(s) to the OSF as well). Finally, please confirm that you have only developed and tested your analysis plan and code using 5% of the dataset (noting that that could be 5% of the replication observations; 5% of the reproduction observations; and/or 5% of the combined observations, as relevant).

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 replication dataset? Is there a separate verification for each analysis specified in section 15?*

The authors of the original paper provided us with their code that produced the focal analysis. The focal analysis was a multilevel mixed-effects linear regression model to estimate the effect of time and governmental social distancing measures on mobility. The following line of code was used to run this analysis:

xtmixed Moving Date Lockdown ||city1:, var

Where Moving is the dependent variable representing change in mobility based on the CMI, and Date and Lockdown are the independent variables. City1 is the level variable for the random-effects equation. (The same analysis can be run from the GUI like this: Statistics -> Multilevel mixed-effects models -> Linear regression; Dependent variable: Moving; Independent variables: Date Lockdown; Random effects equations -> Create -> Level variable for equation: city1)

First, we tried the original code on the original data to make sure that v13 of Stata produced the same results as the original authors' software (which was probably v16). It produced exactly the same results as those in the log file provided to us by the original authors.

To analyze the new data, we only modified this line of code to include the new variable names:

xtmixed CMRT_transit date2 lockdown ||city:, var

Apart from the focal analysis we planned to do an additional analysis to estimate the effect of time and governmental social distancing measures on people staying at home. For this we used a similar multilevel mixed-effect linear regression model as for the focal analysis:

xtmixed CMRT_residential date2 lockdown ||city:, var

The only difference is that the dependent variable here is CMRT_residential which represents change in people staying at residential areas (supposedly their own homes) during the pandemic vs. before the pandemic. We predict the change to be positive, i.e., that people stayed at home more during the pandemic and as a result of lockdowns.

All files necessary for the analysis are uploaded here:
https://osf.io/t4qh5/?view_only=0fd2d8c3b225462c9dfae525baf41a73

- replicationDataset_Malik2020_with.year.csv: This file contains the data provided by the data finders, where the date variable is already changed to show the year.
- mycode_for.replication.dataset.do: This file contains the analysis script.
- results_new.log: This file was produced by the script above and it contains the results of the analyses on a 5% random sample of the data.

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?*

We did not transform any of the variables. Only the format of the date variable was changed in Excel to make sure that the imported dataset correctly includes years. Actually, this change was not necessary for the analyses because all the dates were in the same year, but we still thought that it is best if we import the data without loss.

21. Inference criteria

RR TEAM INSTRUCTIONS: *This section describes the precise 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. They need not rely on p-values and/or the same alpha level we have specified for **H***. Following section 15, if you are performing multiple analyses corresponding to different subsets of the data, please specify whether the same criteria will be used for each analysis (e.g. the same coefficient is expected to be positive and significant in each subset). If the inference criteria differ across analyses, please make that clear below.*

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.

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, this criteria is met by a p value < .05 for the coefficient of the lockdown variable, which should be negative.

For H1, the coefficient of the lockdown variable should be positive and p should be < .05 to conclude that social distancing measures were associated with an increase in people staying at home. H1 is not related to the successfulness of the replication attempt. We decided to do the additional analysis solely out of curiosity.

22. Data exclusion

RR TEAM INSTRUCTIONS: *The section below should describe the rules you will follow to exclude collected cases from the analyses described in Section 19. Note that this refers to exclusions **after** the creation of the replication dataset; exclusion criteria that prevent a case from entering the replication 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 cases included in the replication 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?)*

We did not exclude data from the dataset collected by the data finders.

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?*

There was no missing data in the dataset provided by the data finders.

24. Exploratory analysis (Optional)

RR TEAM INSTRUCTIONS: *If you plan to explore your data set 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.*

25. Other

RR TEAM INSTRUCTIONS: *This section serves two purposes. First, please use this section to discuss any features of your replication 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 replication 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 (including deviations due to differences in location or time compared to the original study)?*
- *Does the preregistration comment on plans to make the data and materials from the replication study public?*

Final review checklist

REVIEWER INSTRUCTIONS: *For the following questions, reviewers 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 pre-registration are specific materials needed to create a replication dataset:
    - Is the final replication dataset that the research team constructed suitable for performing a high-quality, good-faith replication of the focal claim selected from the original study?
    - Is the procedure for constructing the final replication dataset sufficiently documented that an independent researcher could construct the same dataset following the procedures and code they lay out?
- Included with this pre-registration is a narrative description of how the replication dataset will be used to perform the focal replication 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 replication 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 replication dataset?
- I have reviewed all sections of this pre-registration, and I believe it represents a good-faith replication attempt of the original focal claim.

Additionally, please consider the following if the preregistration includes a reproduction analysis:

- The observations used for the reproduction analysis were collected and measured in the same way as the original study.
- The observations used for the reproduction analysis were analyzed in the same way as the original study.
- The data analyst has demonstrated that their analysis code works as expected on a random 5% of the reproduction data.
- I believe this preregistration represents a good-faith reproduction attempt of the original focal claim.