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

Replication team:

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

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

Project ID: Malik_covid_wx5k_996g
OSF project: https://osf.io/zd2uq/
Preregistration: https://osf.io/u5z7m

Claim summary:

The claim selected for replication from Malik et al. (2020) is that social distancing measures introduced by governments during the 2020 pandemic reduced the mobility in the concerned cities. 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.

Replication criteria:

H*: The introduction of social distancing measures is associated with a decrease in mobility.

Criteria for a successful replication attempt 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.

Replication result:

The replication included 39 geographical areas in the analytic sample (mapped to 39 of the 41 cities in the original study) which reaches the requirements for Stage 2 data collection (the stage 2 required sample size would be 8; the stage 1 required sample size would be 4; the minimum viable sample size would be 2 geographic areas - with 25 observations per geographic area, corresponding to the 25 days between March 2nd and March 26th 2020). With 39 geographic areas and 25 days, we had 975 observations.
We found that social distancing measures (“lockdown”) significantly decreased the mobility by 26% (95%CI: 23%, 28%), z = -22.08, p<.001 (reported as .000 by Stata 13.0 statistics software). The replication of the claim was successful according to the SCORE criteria.

Table of the model output (see the full output of the analysis script in the Appendix, at the end of this document):

CMRT_transit	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
date2	-2.094034	.0648392	-32.30	0.000	-2.221116	-1.966951
lockdown	**-25.68071**	1.162885	**-22.08**	**0.000**	**-27.95992**	**-23.4015**
_cons	46021.72	1425.455	32.29	0.000	43227.88	48815.56

Deviations from preregistration:

None. We commented out one line in the analysis script (“sample 5”) which was responsible for randomly sampling 5% of the data for the preregistration, thus we performed the analysis on all of the data as planned.

Description of materials provided:

Our project stored and preregistered on OSF contains all materials that we used: https://osf.io/zd2uq/

The root directory includes:

- SCORE Report - Malik_covid_wx5k_996g.pdf: final SCORE report (this document)
- Malik_covid_wx5k_996g (Capitán & Capitán_Fedor) Preregistration.pdf: the preregistration document
- Malik_996g_prereg_commentary.pdf: a commentary from one of the original authors, Amyn Malik, her concerns regarding the data source

For repeating our analysis one would only need the **Analysis** component on OSF and Stata 13.0 statistics software (we verified that Stata 15.0, the most current version, produces the same results)

- Analysis/replicationDataset_Malik2020_with.year.csv: the data file used for the replication. The data analyst modified the date format in the data file provided by the data finders. Otherwise, it contains the same data as Data/data/replicationDataset_Malik2020.csv and .dta
- Analysis/mycode_for.replication.dataset.do: analysis script for the replication written by the data analyst based on the analysis script used by the original authors
- Analysis/results_new.log: the results produced when running the above analysis script (on 5% of the replication data). This includes an additional analysis that is not part of the replication attempt.
- Analysis/screenshot.jpg: screenshot proof of running the script on 5% of the data for the preregistration

For replicating our results one would have to comment out the “sample 5” line from the script in order to perform the analysis on all of the data.

Other components (subfolders) on the OSF project that are not empty are Power Analysis and Data.

The Data component includes:

- Data/replicationDataset.docx: an explanation and summary about the replication dataset provided by the data finders
- Data/replicationDataset_appendix_1.xlsx: a mapping between the original dataset and the replication dataset with notes justifying the mappings

- Data/code/createReplicationDataset.do: the script that the data finders used to create the replication dataset
- Data/communications/gmail_SCORE_proposal.pdf: copy of email correspondence between the data finders (Tabaré Capitán and Sara Capitán) and the project coordinators (Andrew Tyner and Anna Abatayo). The data finders proposed the dataset for replication, which the project coordinators first rejected, but then after some clarifications, accepted.
- Data/communications/CityGoogleComparison.xlsx: copy of first attachment of the email correspondence above. A table listing the cities in the original study and the corresponding geographic areas in the replication dataset.
- Data/communications/CityGoogleComparison.docx: copy of second attachment of email correspondence above, where the data finders address the project coordinators' concerns
- Data/communications/proposalDefense.docx: duplicate of previous document
- Data/data/dataDictionary.xlsx: the data dictionary provided by the data finders
- Data/data/replicationDataset_Malik2020.csv and .dta: the dataset provided by the data finders in two different formats
- Data/data/raw/Citymapper_Mobility_Index_20200701.csv: Citymapper data (original data) downloaded from the Citymapper app
- Data/data/raw/malik2020.xlsx: the Citymapper data organized in an Excel sheet by the original authors
- Data/data/raw/Overview - Community Mobility Reports Help.pdf: Help document downloaded from Google about their community mobility reports
- Data/data/raw/Understand the data - Community Mobility Reports Help.pdf: another help document downloaded from Google about their community mobility reports

The Power Analysis component includes a zipped folder with 3 files that were used for the power analysis by the SCORE team.

Citation:

Amyn A. Malik, Chandra Couzens, Saad B. Omer (2021). COVID-19 related social distancing measures and reduction in city mobility. medRxiv preprint: https://doi.org/10.1101/2020.03.30.20048090

Appendix: Output log of the analysis script

-

name: <unnamed>

```
      log:  C:\Users\fedor\OneDrive\Documents\DOKUMENTUMOK\Reproducibility
Project\SCORE\Malik2020_replication_using.existing.datasets\data\replication_from.new.data\r
esults_new.log
  log type:  text
 opened on:  13 Jan 2021, 14:13:56

.
. * Import data
. import delimited "replicationDataset_Malik2020_with.year.csv", varnames(1) case(preserve)
clear
(12 vars, 975 obs)
.
. * The "date" variable is a string -> make it to a date type variable called date2
. generate date2=date(date,"MDY")

.
. * Take 5% random sample of the observations
. *sample 5
.
. * Focal analysis: Multilevel mixed-effects linear regression model to estimate the effect of time
and governmental social distancing measures on mobility.
. xtmixed CMRT_transit date2 lockdown ||city:, var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -3743.9982
Iteration 1:   log likelihood = -3743.9982

Computing standard errors:

Mixed-effects ML regression           Number of obs      =       975
Group variable: city                  Number of groups   =        39

                                      Obs per group: min =        25
                                                    avg =      25.0
                                                    max =        25


                                      Wald chi2(2)       =   4738.41

Log likelihood = -3743.9982           Prob > chi2        =    0.0000
```

```
------------------------------------------------------------------------------
CMRT_transit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
date2        |  -2.094034   .0648392   -32.30   0.000    -2.221116   -1.966951
lockdown     |  -25.68071   1.162885   -22.08   0.000    -27.95992    -23.4015
_cons        |   46021.72   1425.455    32.29   0.000     43227.88    48815.56
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.      [95% Conf. Interval]
-----------------------------+------------------------------------------------
city: Identity               |
                  var(_cons) |   77.34121   18.63525    48.22977       124.0243
-----------------------------+------------------------------------------------
               var(Residual) |   112.8675   5.218446    103.0892       123.5732
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   372.30 Prob >= chibar2 = 0.0000

.
. * Additional analysis: Multilevel mixed-effects linear regression model to estimate the effect of
time and governmental social distancing measures on people staying at home.
. xtmixed CMRT_residential date2 lockdown ||city:, var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood =  -3356.584
Iteration 1:   log likelihood =  -3356.584  (backed up)

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =       975
Group variable: city                      Number of groups   =        39

                                          Obs per group: min =        25
                                                         avg =      25.0
                                                         max =        25


                                          Wald chi2(2)       =   1142.23
Log likelihood =  -3356.584               Prob > chi2        =    0.0000

------------------------------------------------------------------------------
```

```
CMRT_residential |     Coef.   Std. Err.          z         P>|z|    [95% Conf. Interval]
-----------------+------------------------------------------------------------------
           date2 |   .6049939   .043579513.88   0.000   .5195798        .6904081
        lockdown |   10.08324   .7815825      12.90   0.000   8.551369        11.61512
           _cons |  -13296.15   958.0705   -13.88   0.000        -15173.93   -11418.36
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.          [95% Conf. Interval]
-----------------------------+------------------------------------------------
city: Identity               |
                  var(_cons) |    34.8733   8.367715    21.78966        55.81303
-----------------------------+------------------------------------------------
           var(Residual) |   50.98797   2.357013          46.5714         55.82337
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   390.86 Prob >= chibar2 = 0.0000

.
. * Close log file
. log close
           name:  <unnamed>
           log:  C:\Users\fedor\OneDrive\Documents\DOKUMENTUMOK\Reproducibility
Project\SCORE\Malik2020_replication_using.existing.datasets\data\replication_from.new.data\r
esults_new.log
  log type:  text
 closed on:  13 Jan 2021, 14:13:57
-------------------------------------------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------------------------------------------
-
```