

Replication of a Research Claim from Gelfand et al. (2020),
from PsyArXiv

Replication Team: Andrew Tyner and Bob Reed

Research Scientist: Nick Fox

Action Editor: Nathaniel Porter

Independent Reviewers

(add name below when you initiate review, comment “DONE” on your name when you finish):

Reviewer #1: [Kai Jonas]

Reviewer #2: Elisabeth Julie VargoDONE

Reviewer #3: [NAME]

Review Period: October 23 - October 28

View-only links to: [Original Paper](#), [Original Materials](#), [Replication Materials](#)

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 Gelfand_covid_R3eV

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 Gelfand et al. (2020).

2. Authors and affiliations

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

Andrew Tyner [Data identification and preparation]¹

Bob Reed [Data analysis]²

¹ Center for Open Science, Charlottesville, VA

² University of Canterbury

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 Gelfand et al. is that cultural tightness and government efficiency should combine to predict the infection rate associated with COVID-19, such that the nations that fare the best may have both culturally tight norms and efficient governments. This reflects the following statement from the paper's abstract: "Nations with efficient governments and tight cultures have been most effective at limiting COVID-19's infection rate and mortality likelihood." The authors predicted that cultural tightness and government efficiency would predict slower growth rates of COVID-19 and lower mortality likelihoods, and that nations with high cultural tightness and high government efficiency would show especially slow growth rate and mortality likelihood. The authors captured infection rate by fitting regression equations for each nation, log-transforming the outcome variable (cases per million people) and the predictor variable (days) to account for the exponential growth rate of the virus. Log-transformation converts exponential growth rates into linear growth rates, which can be predicted in a general

linear model. This model found a significant interaction between tightness and efficiency, $b = -.17$, $SE = .07$, $t(41) = -2.23$, $p = .031$.

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 interaction between cultural tightness and government efficiency will be negative in its association with the COVID-19 infection rate.

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.

[QUESTION 6 - BOLD YOUR RESPONSE ABOVE]

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.*

No blinding was involved to the secondary data collectors' knowledge.

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?

The original study (https://osf.io/vptfx/?view_only=9b264556cde54e4b998cdf5a333cd368) was a preprint that was downloaded on April 6. In that study, the authors “predict that cultural tightness and government efficiency should combine to predict the infection rate and mortality likelihood associated with COVID-19, such that the nations that fare the best may have both culturally tight norms and efficient governments and the nations that fare the worst may have culturally loose norms and inefficient governments.” This was tested with COVID-19 data collected between March 21 - March 30 (p. 2). In the analysis from the preprint being replicated, countries’ COVID-19 growth rates were predicted from the following set of variables: GDP per capita, inequality (measured through the Gini coefficient), median age, cultural tightness, government efficiency, and the interaction of cultural tightness and government efficiency. The interaction term produced the focal finding reflected in section 3 above.

The replication study collects data from the same sources documented in the preprint. In some cases, information about the data sources was provided in greater detail in a later version of the preprint, dated July 29, 2020, and through correspondence with one of the original authors. Since the July 29 version of the preprint was consulted regularly when preparing the replication dataset, it has been preserved on the OSF:

https://osf.io/rbv49/?view_only=9a4bd62713784970a51497152f7a07f2

Notably, the July 29 version of the preprint adds additional context about the focal finding that was not present in the version downloaded on April 6. Specifically, the July 29 preprint focuses

their claims on the early period of the COVID-19 pandemic. For example: “Consistent with other analyses, we focus on data from the earliest available dates up to early April, a critical period during which populations around the world were almost entirely susceptible to catching the virus, causing exponential growth of COVID-19, yet where there was simultaneously tremendous variability across countries in growth curves and deaths” (p. 4). And later: “We also note that after early April, variability in cases was much less pronounced (see Supplemental Materials), and most countries worldwide had adopted stringent measures to contain the infection. Thus, by focusing on the early stage of the pandemic, we can test which countries were better able to flatten the curve and save lives during this critical period.” The title of the study changed as well, reflecting the focus on the early period (“The Importance of Cultural Tightness-Looseness and Government Efficiency for Understanding Early COVID-19 Growth and Death Rates”).

This explicit focus on the early period of the pandemic was not present in the preprint downloaded on April 6, which is the basis of this replication attempt. Accordingly, the data finder (A. Tyner) thought replicating the original analysis with COVID-19 data extended to more recent dates (end of August) represented a good-faith replication attempt and prepared a dataset that allowed for it. This assessment primarily reflects the version of the preprint that the SCORE team accessed on April 6, rather than later versions of the same study.

After receiving the dataset, the data analyst determined that the best faith replication attempt should limit each country’s observations to the first 30 days after meeting the inclusion criteria, so the analysis will only make use of a portion of the replication dataset that’s prepared and documented below.

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.*

As all the independent variables are country-level, randomization does not apply. To the secondary data collectors’ knowledge, the individual-level data used to construct the ‘cultural tightness’ variable did not involve randomization.

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

- 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.*

Data from the sources detailed in sections 12b and 12c below have been downloaded, merged, and cleaned in minimal ways prior to registration. Variables were selected based on their expected relevance to the replication analysis, as determined from the descriptions in the original preprint (downloaded April 6), a more recent version of the preprint (dated July 29), and correspondence with one of the original study authors.

Additionally, the original authors included a csv of country-level variables in the OSF project associated with their preprint (country.vars.csv, accessed in August 2020:

https://osf.io/qpdmj/?view_only=9a4bd62713784970a51497152f7a07f2). The data finder (A. Tyner) consulted that file to verify that the country-level independent variables were being recreated correctly. Because there were some discrepancies between values of the country-level variables that the data finder collected and the values in the country.vars.csv, the data finder contacted the corresponding author to investigate the differences. Details about that correspondence are included at the end of section 12c below.

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)

Infection rate of COVID-19

- In the April 6 version of the preprint, COVID-19 data was gathered between March 21-March 30. "We retrieved data on COVID-19 around the world from the European Center for Disease Control...In order to avoid confounding these COVID-19 data with nations' population sizes, we downloaded data on cases per million citizens, and indexed death rate through the number of mortalities divided by the number of total cases" (p. 2).
- The authors "focused on the rate of cases after the number of cases exceeded 1 per million people" and they "captured infection rate by fitting regression equations for each nation, log-transforming the outcome variable (cases per million people) and the predictor variable (days) to account for the exponential growth rate of the virus" (p. 3).
- The estimates from those regression equations became the dependent variable in the focal regression model.

Focal Independent Variable(s)

Government efficiency

- From the version of the preprint accessed on April 6: "We measured government efficiency using the World Bank's Government Efficiency Index, which assesses the public sector's performance in managing/regulating the political economy...The 2017 measure captures the government efficiency of 126 nations." (p. 2-3).
- The July 29 version of the preprint provides additional details about this variable, including that it was "originally compiled by the World Economic Forum as part of the 2017 Global Competitiveness report" (p. 4). Additional details about this measure were discovered from author correspondence (see below for details).

Cultural tightness

- This variable "captures the strength of norms in a nation and the tolerance for people who violate norms." It was gathered for 33 countries as part of Gelfand et al. 2011 (<https://science.sciencemag.org/content/332/6033/1100>), and has subsequently been expanded to 57 countries in Eriksson et al. (2020) ["Metanorms: The appropriateness of informal sanctions in 57 countries" (unpublished)].

The focal independent variable is an interaction of government efficiency and cultural tightness. The separate variables are also needed as independent variables in the regression.

Control Variable(s)

GDP per capita

- "Economic development was indexed through GDP per capita, which we retrieved from the International Monetary Fund's 2019 release" (p. 5).

Gini coefficient

- "Inequality was indexed through the nations' Gini coefficients, which we retrieved from the most recent World Bank estimate for each nation" (p. 5).

Median age

- "We retrieved data on nations' median ages from the 2018 CIA World Factbook, the most recent release where we could locate this information" (p. 5).

Sample parameter

Cases per million people

- From page 3: "We focused on the rate of cases after the number of cases exceeded 1 per million people, a CDC metric intended to track growth rates as the disease posed a risk to increasing numbers of people."

- The data finder (A. Tyner) interprets this to mean that, for each country, the first day included in the analysis is after its cumulative COVID-19 case count per million people is greater than 1.

Additional variables

Weight

- "...we weighted cases in our second set of regressions by number of observations across nations, so that nations with high numbers of observations (and more reliable estimates) would be weighted over and above nations with low numbers of observations (and less reliable estimates)" (p. 4).
- **Note:** It's unclear what the authors mean by "observations across nations" in this context. It's likely that it means number of days of reported COVID data in the first set of regressions, since they mention as a preface to this statement that "general linear models do not account for the error inherent in estimating growth curves" (p. 4).
- The data finder interprets this to mean the number of days that a country appears in the dataset between the first day when its cumulative COVID-19 case count per million people is greater than 1 and March 30, which is listed as the last day that COVID data was collected (p. 2).

Sample size of analysis has approximately 47 observations.

(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 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.

Data has been accessed from the following sources:

Government Efficiency

- Correspondence with the corresponding author suggested that the specific item used to measure government efficiency was the 2016 government efficiency value row from the following source:
https://tcdat360.worldbank.org/indicators/h8125e315?country=BRA&indicator=40979&viz=line_chart&years=2007,2016
- Data was downloaded 10/13/20 via the ‘Download source data’ link. Data downloads as ‘data.csv’ but has been renamed as ‘efficiency_data_world_bank.csv’ for documentation purposes.

Cultural tightness

- The Eriksson et al. (2020) unpublished study that is the source of the 57 countries’ values for cultural tightness could not be found through an online search. Instead, the replication dataset relies directly on the ‘Tightness’ variable provided in the ‘country.vars.csv’ file associated with the July 29 version of the preprint. Note that this file contains 63 valid country values for ‘Tightness,’ which reflects all the observations in the dataset.
 - The discrepancy between 57 countries in the preprint and 63 countries in the csv is likely accounted for by the exclusion of 6 countries that the corresponding author noted via email: Belgium, France, New Zealand, Norway, Pakistan, and Venezuela.
 - For these six countries, the data collection that produced the tightness measure was conducted earlier than for the rest of the countries. The corresponding author recommended excluding the six countries from the replication attempt, as they did in the original analysis.
- As a small verification check, the 52 values for ‘Cultural Tightness’ made available in the Appendix of the July 29 preprint (see pages 23-25) were directly compared to the values in the ‘country.vars.csv’ file. All were off by less than .01, likely reflecting rounding errors.

GDP per capita

- The source of the GDP per capita data is listed as “the International Monetary Fund’s 2019 release.” The closest available option from an online search is the IMF’s World Economic Outlook database, available here:
<https://www.imf.org/external/pubs/ft/weo/2019/02/weodata/index.aspx>
- The version downloaded is the ‘October 2019 Edition’ [click ‘Entire Dataset’, then click ‘By Countries’].
- Note:** There are many measures of GDP per capita available for each country and many years of data contained in the October 2019 dataset. It was not clear from either version of the preprint which measure of GDP per capita should be used.
 - By comparing the values of ‘GDP.capita’ in the country.vars.csv file to values present in the WEO dataset, the data finder found that most countries’ measures of GDP per capita were from the NGDPDPC measure in 2019. NGDPDPC is defined as follows in the dataset: “GDP is expressed in current U.S. dollars per

- person. Data are derived by first converting GDP in national currency to U.S. dollars and then dividing it by total population.”
- For Belgium, France, New Zealand, and Norway, the measure of GDP per capita appears to be from the 2020 measure of PPPC, which is defined in the dataset as: “Expressed in GDP in PPP dollars per person. Data are derived by dividing GDP in PPP dollars by total population.”
 - For two countries in the country.vars.csv file -- Pakistan and Venezuela -- the data finder was unable to find corresponding GDP per capita values in the WEO dataset.
 - As mentioned above, correspondence with the original author made clear that these six countries were excluded from the original analysis. Regardless, *all* countries in the replication dataset have GDP per capita values from the 2019 NGDPDPC measure.

Gini coefficient

- The source of the Gini coefficient data is listed as “the most recent World Bank estimate for each nation” (p. 5 of April 6 preprint).
- Accordingly, we downloaded the GINI index from the World Bank, available as a csv from here: <https://data.worldbank.org/indicator/SI.POV.GINI>
 - Click on CSV, which downloads the ‘API_SI.POV.GINI_DS2_en_csv_v2_1217501’ zip file.
 - Within the file, use ‘API_SI.POV.GINI_DS2_en_csv_v2_1217501.csv’.
- As detailed in section 12d below (in Section “Discrepancies between re-collected data and authors’ values”), for each country in the country.vars.csv file, the most recent non-NA Gini value was used. This resulted in a set of values that were -- for the most part -- very similar but rarely equivalent to the values present in the ‘Gini’ column of the country.vars.csv file. Some notable deviations:
 - The value for Brazil listed in the country.vars.csv file is 46.9, though the most recent year of Gini data for Brazil in the World Bank data is 53.9 (from 2018).
 - New Zealand, Qatar, Saudi Arabia, and Singapore all have valid responses in the country.vars.csv file, but only contain NA values for all years in the World Bank data (1964-2019).

Median age

- According to the April 6 preprint, the authors “retrieved data on nations’ median ages from the 2018 CIA World Factbook, the most recent release where we could locate this information” (p. 5).
- A link to the data source was not provided in either version of the preprint, but the latest available Factbook appears to be 2018 according to this website (<https://www.cia.gov/library/publications/download/download-2018/index.html>), which serves as the replication data source.
 - File downloaded (8/27) through the factbook.zip link within ‘Single .Zip file for high-bandwidth users: factbook.zip (1.6GB)’

- File path to median age data is factbook -> fields -> rawdata_343.txt.
- As with the Gini coefficient data above, the values in the ‘rawdata_343.txt’ file were -- for the most part -- very similar but rarely equivalent to the values present in the ‘Median_Age’ column of the country.vars.csv file. Some notable deviations (with more details in section 12c in Section “Discrepancies between re-collected data and authors’ values” below):
 - The values for Ireland are off by 17.1 (20 in the country.vars.csv file and 37.1 in the ‘rawdata_343.txt’ file). The values for the United Arab Emirates are off by 6.9 (30.3 in the country.vars.csv file and 37.2 in the ‘rawdata_343.txt’ file).
 - Of note, 16 observations were off by exactly 0.4, with the version in country.vars.csv always smaller than the version from the factbook.

COVID-19 data

- Per the April 6 preprint, “We retrieved data on COVID-19 around the world from the European Center for Disease Control, which provides daily updates of the number of COVID-19 documented cases and the number of documented deaths due to COVID-19” (p. 2).
- The ECDC provides a file of COVID data that is updated daily, available here: <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>. The version used for the replication data contains country-day data between 2019-12-31 and 2020-08-27.
- It was unclear what the authors meant by, “In order to avoid confounding these COVID-19 data with nations’ population sizes, we downloaded data on cases per million citizens,” since there does not appear to be a measure of ‘cases per million citizens’ in the ECDC data. However, the ECDC data does contain a ‘popData2019’ column, so a cases per million citizens variable can be derived.

(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.

COVID-19 Infection Rate

From the April 6 preprint: "We focused on the rate of cases after the number of cases exceeded 1 per million people, a CDC metric intended to track growth rates as the disease posed a risk to increasing numbers of people...We captured infection rate by fitting regression equations for each nation, log-transforming the outcome variable (cases per million people) and the predictor variable (days) to account for the exponential growth rate of the virus." (p. 3)

For use in the procedure above, the following variables have been included in the replication dataset, unaltered from the ECDC:

- cases
- popData2019
- day
- month
- year
- date: This is a slightly adjusted version of the original 'dateRep' from the ECDC. The 'date' variable in the replication dataset has been converted from a character string to a Date class variable.

Additionally, the following variables have been created based on raw data from the ECDC:

- running_total_by_country: Cumulative sum of cases in each country.
- pop_per_million: popData2019 divided by 1000000
- Total_covid_per_million: Computed as running_total_by_country divided by pop_per_million. According to the preprint, country-days are only included for calculating the infection rate after the number of cases exceeded 1 per million people in the country.

The following variables were retained from the ECDC data, but don't have an obvious use in the replication analysis:

- deaths
- geoid (2-character country ID)
- countryterritoryCode (3-character country ID)

Special note on the 'cases' and 'deaths' variables

There are 8 country-days with a negative values on the cases variable:

- Spain, 2020-04-19: -713
- Portugal, 2020-05-03: -161
- Ecuador, 2020-05-07: -2461
- Ecuador, 2020-05-09: -1480
- Ecuador, 2020-05-12: -50
- Spain, 2020-05-25: -372
- France, 2020-06-03: -766
- Italy, 2020-06-20: -148

NOTE: All instances of negative cases occur after 30 days following attainment of 1 case per million.

There are 5 country-days with a negative value on the deaths variable:

- Spain, 2020-05-25: -1918
- Italy, 2020-06-25: -31
- Czech Republic, 2020-07-05: -1
- Czech Republic, 2020-07-06: -3
- Spain, 2020-08-12: -2

It's unclear what these negative values represent, though they're most likely typos. Alternatively, for the cases variable, the negative values could be previously classified positive cases that have been reclassified as negative. For the deaths variable, they could be previously classified COVID-19 deaths that have been reclassified to a different cause. But in the latter scenario, it's not clear whether, e.g., 713 is the number of positive cases that were reclassified as negative in Spain on 4/19, or whether 713 is the difference between the number of positive cases and the number of positive cases that were reclassified as negative on 4/19.

Additional notes on territories

The ECDC data organizes geographic units as 'countries and territories,' and thus includes territories like Puerto Rico, the US Virgin Islands, and Guam as separate units. Absent any indication from the preprint about whether territories' case counts were added to the countries they are a part of, the case counts from territories have **not** been combined with countries' case counts in the replication dataset. Further, the only units retained in the replication dataset were ones with values for 'cultural tightness' in the authors' original data (see below). Since none of the territories have observations in the authors' country data, the COVID case data from territories has not been included in the replication dataset in any way.

Additionally, one of the values of the 'countriesAndTerritories' variable from the ECDC data is Cases_on_an_international_conveyance_Japan. Based on Google searching, it appears this refers to cases reported from a cruise ship. These case counts have not been added to Japan's cases, nor have they been included in the replication dataset in any way.

Government efficiency

- The corresponding author shared the exact items used via email. These items were re-collected directly from the World Bank.
- This item appears as 'efficiency' in the replication dataset, and it represents the 2016 values for the '4. Government Efficiency' indicator from the World Bank data source.

Cultural tightness

As mentioned above, the data file associated with the authors' July 29 version of the preprint appears to be the only source available for cultural tightness, so it has been included directly in

the replication dataset. The variable name in the original data file is ‘Tightness’ and has been renamed as tightness in the replication dataset.

Because cultural tightness is only available with the authors’ data, it is the limiting factor in the number of countries that can be included in the replication dataset. Accordingly, the replication dataset created below only includes countries with an entry in the authors’ country.vars.csv dataset. Additionally, as mentioned in a few places in the preregistration, the corresponding author recommended excluding Belgium, France, New Zealand, Norway, Pakistan, and Venezuela, since the data collection that produced the tightness measure for those countries was conducted earlier than for the rest of the countries.

GDP per capita

GDP per capita data was accessed from the October 2019 version of the IMF’s World Economic Outlook database. The specific item selected for the replication dataset is the 2019 value for NGDPDPC [GDP expressed in current U.S. dollars per person].

Gini coefficient

Data was accessed from the World Bank’s Gini index. For each country, the most recent non-NA value was included as the country’s value for ‘gini_val’ in the replication dataset. Note that New Zealand, Qatar, Saudi Arabia, and Singapore have no valid Gini values for any years in the World Bank data, accounting for the 885 NA values in the replication dataset. See notes below about the ‘alternative_gini’ column, which supplies data for some of these countries.

Median age

The value for ‘median_age’ in the replication dataset was found in ‘rawdata_343.txt’ from the 2018 CIA World Factbook.

Sample parameter and weights

From pages 3-4 of the preprint: “We focused on the rate of cases after the number of cases exceeded 1 per million people, a CDC metric intended to track growth rates as the disease posed a risk to increasing numbers of people.... we weighted cases in our second set of regressions by number of observations across nations, so that nations with high numbers of observations (and more reliable estimates) would be weighted over and above nations with low numbers of observations (and less reliable estimates).”

The replication dataset contains a variable -- total_covid_per_million -- that tracks the number of cumulative COVID-19 cases per million in a country, in case the analyst wants to apply a similar filter in the replication.

Additionally, the replication dataset contains three observation count variables for purposes of weighting, corresponding to three different analyses the analyst might perform:

- obs_count_full, measuring the number of times that each country appears in the full dataset (12/31/19 - 08/27/20).
- obs_after_one_per_million, measuring the number of times that each country appears in the dataset when its value of total_covid_per_million is higher than 1.
- obs_count_original, measuring the number of times that each country appears in the dataset between the first date where its value of total_covid_per_million is higher than 1 and 03/30/20. This assumes the original period of data collection ended on March 30, as indicated on page 2 of the preprint.

Discrepancies between re-collected data and authors' values

GDP per capita & countries to include

As mentioned above, the GDP per capita value selected is the 2019 value for NGDPDPC [GDP expressed in current U.S. dollars per person] from the IMF. When compared with the values in the country.vars.csv file, all values match exactly except for Belgium, France, New Zealand, Norway, Pakistan, and Venezuela.

Our emails with the corresponding author revealed that those six countries were not included in the original analysis anyways, for reasons detailed above. The corresponding author recommended excluding the six countries from the replication attempt, as they did for the original analysis.

- The data finder (A. Tyner) has preserved them in the replication dataset, and it is up to the data analyst to decide whether a good faith replication attempt should include or exclude them. If they're included, it should just be noted that the values for GDP per capita reflect the 2019 values for NGDPDPC for each of these six countries, rather than the values present in the country.vars.csv file.
- The data finder's recommendation is to exclude them, per the corresponding author's guidance.

Gini data

When comparing the Gini values as sourced from the World Bank data with the values available in the country.vars.csv file, we found that 11 countries have exact matches and 26 that are within 0.5. Some of the countries have larger deviations; for example, the value for Brazil listed in the country.vars.csv file is 46.9, though the most recent year of Gini data for Brazil in the World Bank data is 53.9 (from 2018). Additionally, New Zealand, Qatar, Saudi Arabia, and Singapore all have valid responses in the country.vars.csv file, but only contain NA values for all years in the World Bank data (1964-2019).

Emails with the corresponding author clarified that the original authors relied largely on the World Bank values in the Wikipedia article for “List of countries by income equality,” which accounts for the discrepancies. They noted as well that some of the values in the Wikipedia entry have since been updated. As of 10/13, the value for Brazil now matches what’s in the World Bank data (53.9, from 2018).

Regarding the countries without valid values in the World Bank data: the corresponding author shared the following sources below that were used for Qatar, Saudi Arabia, and Singapore, as well as Slovakia and South Korea (which do have valid values in the World Bank data). The values in the sources below for Qatar, Saudi Arabia, Singapore, and Slovakia are provided in an ‘alternative_gini’ column in the replication dataset, which the data analyst can combine with the ‘gini_val’ column if they wish to. The value for South Korea in the source below was not added to the ‘alternative_gini’ column, since it matches the value that was already present in the ‘gini_val’ column. Slovakia has different values in the ‘gini_val’ and ‘alternative_gini’ columns, reflecting the different sources. All other countries are NA for the ‘alternative_gini’ column, including New Zealand, which does not have a valid value in either column:

- Qatar: <http://hdr.undp.org/en/content/income-gini-coefficient>
 - Value is 41.1 as of date accessed (10/13).
- Saudi Arabia:
<https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html>
 - Value is 45.9 as of date accessed (10/13).
- Singapore [CIA column]:
https://en.wikipedia.org/wiki/List_of_countries_by_income_equality
 - Value is 46.4 as of date accessed (10/13).
- Slovakia:
<https://knoema.com/atlas/Slovakia/GINI-index#:~:text=Slovakia%20GINI%20index%20was%202026.1.0.00%25%20from%20the%20previous%20year>
 - Value is 26.1 as of data accessed (10/13).
- South Korea [World Bank column]:
https://en.wikipedia.org/wiki/List_of_countries_by_income_equality
 - Value is 31.6 as of date accessed (10/13). Because that value matches the value already present in the ‘gini_val’ column, it has not been included in the ‘alternative_gini’ column.

Government efficiency

The government efficiency values collected from the World Bank source are nearly equivalent to the values of ‘Government.efficiency’ from the country.vars.csv file, save for a few exceptions:

- The values of New Zealand, Norway, France, and Belgium are off by a tiny amount (all less than .01), though notably the corresponding author recommends that these four counties be excluded anyways.

- Pakistan and Venezuela (also countries that the corresponding author has recommended be excluded) appear to be inverted: they're listed as 1.41 and 3.05 in the authors' data file and as 3.05 and 1.41 in the World Bank data.
- UAE, Botswana, and Ivory Coast are missing in the authors' data but available in the World Bank data; the corresponding author noted that they've since added entries for UAE and Botswana to their own file.
- Kazakhstan and Sweden are NA in both the authors' data and the World Bank data.

Median age

As with the Gini coefficient data above, the values in the 'rawdata_343.txt' file were -- for the most part -- very similar but rarely equivalent to the values present in the 'Median_Age' column of the country.vars.csv file. Some notable deviations:

- The values for Ireland are off by 17.1 (20 in the country.vars.csv file and 37.1 in the 'rawdata_343.txt' file). The values for the United Arab Emirates are off by 6.9 (30.3 in the country.vars.csv file and 37.2 in the 'rawdata_343.txt' file).
- Of note, 16 observations were off by exactly 0.4, with the version in country.vars.csv always smaller than the version from the factbook.

The corresponding author shared that the source they used was the Wikipedia version of the CIA estimates, available here: https://en.wikipedia.org/wiki/List_of_countries_by_median_age. The author noted that their value for Ireland was a typo, but it's unclear why there are so many small discrepancies between the Wikipedia version and the original data from the CIA.

(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.

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.*

Manual handling of GDP per capita data

As mentioned in section 12b above, the October 2019 Edition of the IMF’s World Economic Outlook database is what’s included in the replication dataset for the GDP per capita measures. It was downloaded in raw form through this page:

<https://www.imf.org/external/pubs/ft/weo/2019/02/weodata/index.aspx>

- Click on ‘Entire Dataset’ under ‘October 2019 Edition’ and then click on ‘By Countries’ to download. File was downloaded 8/27 as ‘WEOOct2019all.xls.’
- For easier handling, the excel file was uploaded to Google Sheets, saved as a Google Sheet, and then re-downloaded as a csv (WEOOct2019all.csv).

Raw data files

- Authors’ original country-level variables:
https://osf.io/qpdmj/?view_only=9a4bd62713784970a51497152f7a07f2
- COVID data: https://osf.io/zf8ma/?view_only=9a4bd62713784970a51497152f7a07f2
- Government efficiency:
https://osf.io/szudm/?view_only=9a4bd62713784970a51497152f7a07f2
- GDP per capita data:
https://osf.io/3s85p/?view_only=9a4bd62713784970a51497152f7a07f2
- Gini data: https://osf.io/xtejr/?view_only=9a4bd62713784970a51497152f7a07f2
- Median age data: https://osf.io/j63ha/?view_only=9a4bd62713784970a51497152f7a07f2

Creating the replication dataset

All data cleaning and merging was performed in R. Please see the R code linked below for details on packages and versions.

The following steps create the replication datasets:

- Load data from the following sources:
 - country.vars.csv from the original authors
 - ECDC COVID data
 - Government efficiency data from the World Economic Forum
 - Economic data from the IMF
 - Gini data from the World Bank
 - Median age data from the CIA World Factbook
- Reconcile all datasets' country names with the spelling in the authors' country.vars.csv file.
- With the government efficiency data, filter down to just indicator 40979 ['Value' row for the 'Government efficiency' measure] and for just the countries included in the authors' variable file.
- With the IMF data, select each country's measure of NGDPDPC from 2019 and filter to just include the countries in the authors' variable file.
- With the Gini data, filter to just include the countries in the authors' variable file and then select the rightmost value that is not NA as the measure of Gini to include the replication dataset.
 - Additionally create an 'alternative_gini' variable, documented above, that contains non-NA values only for Qatar, Saudi Arabia, Singapore, and Slovakia.
- With the COVID data, perform some light cleaning on the ECDC data:
 - Rename the country variable.
 - Create a new date variable of the class Date.
 - Create a cumulative sum of the COVID cases in each country (running_total_by_country) and a measure of the population count expressed in millions (pop_per_million), and then use those two measures to compute total COVID cases per million in each country (total_covid_per_million).
- From the authors' country.vars.csv file, select the tightness and country variables, and use them as the basis of the replication dataset by left-joining data from all the other cleaned datasets. This means that only countries with a value for 'cultural tightness' are included in the replication dataset.
- Finally, create three observation counts that sum the number of times each country appears in the dataset within different parameters. These can be used to weight the countries according to the original study's procedures, and they correspond to different analyses that might be performed:
 - obs_count_full: count each date the country appears in the dataset
 - obs_after_one_per_million: begin counting each date the country appears in the dataset only after its value of total_covid_per_million is greater than 1.
 - obs_count_original: begin counting each date the country appears in the dataset after its value of total_covid_per_million is greater than 1 and then stop on March 30.

The R code to produce the replication datasets is found here:
https://osf.io/6xkd7/?view_only=5188d39edd104098be929dff3a896e20. Please consult this file for details on its use: https://osf.io/dvtyw/?view_only=5188d39edd104098be929dff3a896e20

(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.

The data dictionary for the replication dataset is available here:

https://osf.io/ztswg/?view_only=5188d39edd104098be929dff3a896e20

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.

Data finders' response goes here: The number of countries available for the replication analysis depends on the inclusion criteria.

- In the full dataset -- spanning 12/31/19 - 8/27/20 -- there are 63 distinct countries. All have at least 133 observations in the sample period where the total COVID cases per million in the country are greater than 1. 60 of these countries have non-missing responses for all of the independent variables.
- Of the 60 countries without missing data, 5 are countries that the corresponding author recommended not be included in the replication analysis: Belgium, France, Norway, Pakistan, and Venezuela. If the data analyst follows that recommendation, the number of available countries is 55.

Required sample size [to be filled out by the SCORE team]: The primary unit of analysis is the country. An estimate of the minimum viable sample size for the data analytic replication is: 38. For comparison, the stage1 required sample size would be: 122 and the stage2 sample size would be: 180.

Notes: The analysis in this paper happens in 2 steps. Initially a GLM is fit for each country to obtain the infection rate for each country. In a second step, the estimates for each country from this initial model are used as the DVs in a weighted OLS with only country level predictors, with each country weighted by the number of observations in each country. The current power analysis does not take into account the first step in this process, and so implicitly assumes that the estimate of the country infection rates in the replication will be approximately as accurate as those in the original (so based on about the same number of observations).

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:

https://osf.io/h38n4/?view_only=5188d39edd104098be929dff3a896e20

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.*

The replication data combines country-days from the April 6 preprint with more recent country-days. Accordingly, the SCORE team recommends that two analyses be performed:

- An analysis only using eligible country-days that were **not** used in the original study. This will be the focal replication analysis of the study.
- A second analysis using all eligible country-days in the replication data.

Because there are known deviations between the original data sources and the replication data sources, the SCORE team does not recommend performing 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 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]

Dependent Variable(original study)

Infection rate of COVID-19

- In the April 6 version of the preprint, COVID-19 data was gathered between March 21-March 30. "We retrieved data on COVID-19 around the world from the European Center for Disease Control...In order to avoid confounding these COVID-19 data with nations' population sizes, we downloaded data on cases per million citizens, and indexed death rate through the number of mortalities divided by the number of total cases" (p. 2).
- The authors "focused on the rate of cases after the number of cases exceeded 1 per million people" and they "captured infection rate by fitting regression equations for each nation, log-transforming the outcome variable (cases per million people) and the predictor variable (days) to account for the exponential growth rate of the virus" (p. 3).
- The estimates from those regression equations became the dependent variable in the focal regression model.

Dependent Variable(replication)

Infection rate of COVID-19

- Per the April 6 preprint, "We retrieved data on COVID-19 around the world from the European Center for Disease Control, which provides daily updates of the number of COVID-19 documented cases and the number of documented deaths due to COVID-19" (p. 2).
- The ECDC provides a file of COVID data that is updated daily, available here: <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>. The version used for the replication data contains country-day data between 2019-12-31 and 2020-08-27.
- The ECDC data contains a 'popData2019' column, so a cases per million citizens variable can be derived.

- The replication dataset contains a variable -- total_covid_per_million -- that tracks the number of cumulative COVID-19 cases per million in a country, in case the analyst wants to apply a similar filter in the replication.

Deviation: The data sources of Covid infection rates are identical between the original study and replication.

Focal Independent Variable1(original study)

Government efficiency

- From the version of the preprint accessed on April 6: "We measured government efficiency using the World Bank's Government Efficiency Index, which assesses the public sector's performance in managing/regulating the political economy...The 2017 measure captures the government efficiency of 126 nations." (p. 2-3).
- The July 29 version of the preprint provides additional details about this variable, including that it was "originally compiled by the World Economic Forum as part of the 2017 Global Competitiveness report" (p. 4). Additional details about this measure were discovered from author correspondence.

Focal Independent Variable1(replication)

Government efficiency

- The corresponding author shared the exact items used via email. These items were re-collected directly from the World Bank.
- This item appears as 'efficiency' in the replication dataset, and it represents the 2016 values for the '4. Government Efficiency' indicator from the World Bank data source.

Deviation: The Data Finder reports that the government efficiency values are "nearly equivalent."

Focal Independent Variable2(original study)

Cultural tightness

- This variable "captures the strength of norms in a nation and the tolerance for people who violate norms." It was gathered for 33 countries as part of Gelfand et al. 2011 (<https://science.sciencemag.org/content/332/6033/1100>), and has subsequently been expanded to 57 countries in Eriksson et al. (2020) ["Metanorms: The appropriateness of informal sanctions in 57 countries" (unpublished)].

Focal Independent Variable2(replication)

Cultural tightness

The data file associated with the authors' July 29 version of the preprint appears to be the only source available for cultural tightness, so it has been included directly in the replication dataset. The variable name in the original data file is 'Tightness' and has been renamed as tightness in the replication dataset.

Deviation: The variables are identical.

NOTE: The focal independent variable is an interaction of government efficiency and cultural tightness.

Control Variable1(original study)

GDP per capita

- “Economic development was indexed through GDP per capita, which we retrieved from the International Monetary Fund’s 2019 release” (p. 5).

Control Variable1(replication)

GDP per capita

- GDP per capita data was accessed from the October 2019 version of the IMF’s World Economic Outlook database. The specific item selected for the replication dataset is the 2019 value for NGDPDPC [GDP expressed in current U.S. dollars per person].

Deviation: For the countries included in the focal regression equation, the GDP values are identical.

Control Variable2(original study)

Gini coefficient

- “Inequality was indexed through the nations’ Gini coefficients, which we retrieved from the most recent World Bank estimate for each nation” (p. 5).

Control Variable2(replication)

Gini coefficient

- Data was accessed from the World Bank’s Gini index. For each country, the most recent non-NA value was included as the country’s value for ‘gini_val’ in the replication dataset. Note that New Zealand, Qatar, Saudi Arabia, and Singapore have no valid Gini values for any years in the World Bank data, accounting for the 885 NA values in the replication dataset. See notes below about the ‘alternative_gini’ column, which supplies data for some of these countries.

Deviation: The countries Qatar, Saudi Arabia, and Singapore did not have Gini coefficients reported in the World Bank data source. The original study used alternative sources of data for these countries. The replication data used the same alternative sources for these countries, though these are reported in a separate variable. The resulting, combined data produced values very close, but not identical to the original study. The differences should not impact the ability of the replication study to provide a fair and reasonable test of the focal claim from the original study.

Control Variable3(original study)

Median age

- “We retrieved data on nations’ median ages from the 2018 CIA World Factbook, the most recent release where we could locate this information” (p. 5).

Control Variable3(replication)

Median age

The value for ‘median_age’ in the replication dataset was found in ‘rawdata_343.txt’ from the 2018 CIA World Factbook.

Deviation: Despite coming from the same original data source, the Data Finder notes that there were small discrepancies between the original study and the replication. However, this difference should not substantially impact the ability of the replication to provide a fair and reasonable test of the focal claim in the original study.

Data filter1 (original study)

- From page 3: “We focused on the rate of cases after the number of cases exceeded 1 per million people, a CDC metric intended to track growth rates as the disease posed a risk to increasing numbers of people.”
- The data finder (A. Tyner) interprets this to mean that, for each country, the first day included in the analysis is after its cumulative COVID-19 case count per million people is greater than 1.

Data filter1 (replication)

- The replication dataset contains a variable -- total_covid_per_million -- that tracks the number of cumulative COVID-19 cases per million in a country, in case the analyst wants to apply a similar filter in the replication.

Deviation: The data analyst will exclude observations with cases < 1 per million to ensure comparability with the original study.

Data filter2 (original study+replication)

Countries

The data finder reports that the countries in the replication dataset include all the countries in the original study plus Belgium, France, New Zealand, Norway, Pakistan, and Venezuela. As these countries were not included in the original study, the Data Analyst will drop these.

Deviation: The countries are identical in the original study and replication.

,

Additional variables

Weight

- “...we weighted cases in our second set of regressions by number of observations across nations, so that nations with high numbers of observations (and more reliable estimates) would be weighted over and above nations with low numbers of observations (and less reliable estimates)” (p. 4).

NOTE: There is no reason to weight the different estimated coefficients in the replication because the number of observations per country is the same.

Sample size (original study)

Total observations = 1202

Sample size (replication)

Total observations = 1710

Deviation: The replication dataset is 42% larger than the original. It includes much of the original study’s data. It mostly differs by adding additional days of data to the original study, which averaged about 22 days of day. In contrast, the replication study allows 30 days of observations for each country. The number 30 was chosen because an informal inspection of the literature indicated that this was a common number of days for estimating exponential growth rate regressions.

The replication differs in a few cases by using less data. The original study had eight countries with more than 30 days of data. The maximum was 65 days. This not only affected the total number of observations, but since the original study weighted by number of observations, it gave greater weight to countries with more observations. Thus, from a practical perspective, the replication study is more than 42% different than the original.

NOTE#1: If reviewers/the editor so advise, there is ample scope to increase the number of days included in the replication analysis, as the data are available. The issue centers on what is the appropriate number of days for estimating exponential growth rate regressions.

NOTE#2: Because the replication dataset uses a substantial number of observations from the original study, this is a hybrid replication. Note that it makes no sense to estimate the exponential growth rate regressions without the original data.

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?*

NOT APPLICABLE.

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?

This statement confirms that only 5% of the data have been randomly sampled in developing the analysis plan and code contained in this preregistration.^[1]

NOTE: Due to problems with multicollinearity, I randomly sample 20% of the data.

The analysis consisted of two stages. The first stage estimated country-specific exponential growth rate equations:

```
reg ltotalcases time
```

where $ltotalcases = \log$ of total number of cases per million, and time = days after total number of cases per million > 1 (following the original study).

In the second stage, the country-specific regression coefficients are regressed on government efficiency, cultural tightness, the interaction of these two variables ("eff_tight"), and some control variables:

```
regress coeffs1 eff_tight efficiency tightness gdp gini median_age
```

The focal claim being investigated stated in Section 3 above:

"The claim selected for replication from Gelfand et al. is that cultural tightness and government efficiency should combine to predict the infection rate associated with COVID-19, such that the nations that fare the best may have both culturally tight norms and efficient governments. This reflects the following statement from the paper's abstract: "Nations with efficient governments and tight cultures have been most effective at limiting COVID-19's infection rate and mortality likelihood." The authors predicted that cultural tightness and government efficiency would predict slower growth rates of COVID-19 and lower mortality likelihoods, and that nations with high cultural tightness and high government efficiency would show especially slow growth rate and mortality likelihood. The authors captured infection rate by fitting regression equations for each nation, log-transforming the outcome variable (cases per million people) and the predictor variable (days) to account for the exponential growth rate of the virus. Log-transformation converts exponential growth rates into linear growth rates, which can be predicted in a general linear model. This model found a significant interaction between tightness and efficiency, $b = -.17$, $SE = .07$, $t(41) = -2.23$, $p = .031$, such that cultural tightness significantly predicted slower rates of COVID-19 cases amongst governments with relatively high (1 SD above the mean) efficiency."

Here is the corresponding result from the Gelfand study:

We next tested the interaction of cultural tightness and government efficiency on growth rates of COVID-19. This model found a significant interaction between tightness and efficiency, $b = -.17$, $SE = .07$, $t(41) = -2.23$, $p = .031$, such that cultural tightness significantly predicted

NOTE: It is apparent that the original study created a set of dummy variables for High, Medium, and Low government efficiency and cultural tightness, respectively. However, as it was not clear what omitted category was used as the benchmark, the Data Analyst chose to interact the continuous variables and have the associated estimated coefficient be the focal estimate.

A test of the focal claim is then given by the following:

H*: The interaction between cultural tightness and government efficiency will be negative in its association with the COVID-19 infection rate.

I test the focal claim using the regression described above.

Below is a screen shot from the focal replication analysis on a random sample of 20% of the observations:

. regress coeffs1 eff_tight gdp gini median_age efficiency tightness						
Source	SS	df	MS	Number of obs = 10		
Model	.045606164	6	.007601027	F(6, 3) = 10.94		
Residual	.002084615	3	.000694872	Prob > F = 0.0379		
Total	.047690779	9	.005298975	R-squared = 0.9563		
				Adj R-squared = 0.8689		
				Root MSE = .02636		
coeffs1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
eff_tight	-.1081035	.1138253	-0.95	0.412	-.4703463	.2541393

The script to implement these tests is given in the do file “Analysis_script” and can be found here: <https://osf.io/wk7n3/>

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?

Transformation of the dependent variable (“ltotalcases”)

The dependent variable was logged to get the natural log of total cases:
`gen ltotalcases = log(total_covid_per_million)`

The focal variable (“eff_tight”) was created as the interaction of the efficiency and tightness variables:

`gen eff_tight = efficiency*tightness`

The control variable “gini” was created as an amalgam of two other gini coefficient variables provided in the replication dataset:

`gen gini = gini_val
replace gini = alternative_gini if gini_val == .`

Country-specific exponential growth rate regressions were estimated to obtain the growth rate coefficients for the second stage:

```
// This estimates country-specific exponential growth regressions
matrix coeffs = J(57,1,.)
matrix names = J(57,1,.)
forvalues i = 1/57 {
    reg ltotalcases time if countryid`i' == 1
    matrix coeffs[`i',1] = _b[time]
    matrix names[`i',1] = `i'
}

```

The code used to make these transformations is included in the file “Analysis_script” which can be found here: The script to implement these tests is given in the do file “Analysis_script” and can be found here: <https://osf.io/wk7n3/>

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 negative and statistically significant coefficient on the variable “eff_tight”.

A screen shot of the output is provided in Section 19.

The location of the analysis script is also provided in Section 19.

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?)*

The replication analysis implemented the following restrictions:

- Only used the countries that were included in the original study
- Counted days of infection starting with the first day where the total cases per million exceeded 1 (to be consistent with the original study)
- Limited the days of analysis for the individual exponential growth rate regressions to 30.

The code used to implement these data exclusion criteria is included in the file “Analysis_script” which can be found here: The script to implement these tests is given in the do file “Analysis_script” and can be found here: <https://osf.io/wk7n3/>

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?*

Listwise deletion was used to handle missing data.

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.*

NOT APPLICABLE.

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?*

NOT APPLICABLE.

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.