

Replication of a Research Claim from Hossain (2020), from  
medRxiv

Replication Team: Carolin Nast and Esteban Méndez-Chacón

Utrecht University & Central Bank of Costa Rica

Research Scientist: Nick Fox

Action Editor: Kevin Esterling

Final Report

November 6, 2020

## **Replication of a Research Claim from Hossain (2020), from medRxiv**

### **Claim Summary**

The claim selected for replication from Hossain (2020) is that a cohesive society where people are involved in more festivals and religious gatherings is more likely to have more cases of COVID-19. This reflects the following statements from the paper's abstract: "The estimation results indicate that the number of confirmed cases of Coronavirus infection is higher in countries with lower yearly average temperature, higher economic openness, and stronger political democracy." The description of the analysis is as follows: "By estimating a regression equation where the dependent variable is the total number of cases confirmed infection per one million people in a country on a day (03 April 2020) and the predictor variables are the democracy index of the country (the proxy variable for social cohesiveness), the yearly average temperature of the country, and the openness of the country (measured by international trade as a percentage of GDP)...We apply Least Squares method on model (1) and find that precipitation and population density have no significant effect on the number of infection cases per one million people (Y). Those variables are then excluded, and the model is re-estimated. In the re-estimated model, the variables average temperature, openness and democracy appear as highly significant." The focal finding is: "The positive sign of democracy index indicates that more democratic countries are affected more by the disease."

**Focal hypothesis H\*: At the country level, the democracy index will be positively associated with the total number of confirmed infections per one million people.**

### **Replication Criteria**

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\*).

### **Replication Results**

#### ***1. Using data from the original study***

Table R.1 contains the results of the Ordinary least squares (OLS) regression using the data from the original study. This is, cases of infection per one million people on April 3, 2020. **The**

coefficient for democracy is equal to 78.96, suggesting a positive association between the democracy index and the total number of confirmed infections per one million people. The coefficient is statistically significant at the 5% level ( $p = 0.000$ ). Thus, this replication of the claim was successful according to the SCORE criteria. The analytic sample included 150 countries, which did meet the minimum threshold of 38 countries defined by the power analysis. The Cohen's  $f^2$  corresponds to 0.56.

Table R.1. Estimation Results - Data Used in the Original Study  
(Dependent Variable: Cases of Infection per Million on April 3, 2020)

<b>Democracy</b>	<b>78.96</b> <b>(0.000)</b>
Temperature	-10.96 (0.035)
Openness	7.81 (0.000)
Constant	-300.65 (0.097)
<b>Adjusted R<sup>2</sup></b>	<b>0.35</b>
<b>Number of countries</b>	<b>150</b>

Figures in the parenthesis below the coefficients are the p-values.

## *2. Using data after the original study was conducted*

Table R.2 contains the results of the Ordinary least squares (OLS) regression using data after the original study was conducted. This is, cases of infection per one million people after April 3, 2020. **The coefficient for democracy is equal to 46.35, suggesting a positive association between the**

democracy index and the total number of confirmed infections per one million people. However, the coefficient is not statistically significant at the 5% level ( $p = 0.826$ ). Thus, this replication of the claim was not successful according to the SCORE criteria. The analytic sample included 150 countries, which did meet the minimum threshold of 38 countries defined by the power analysis. The Cohen's  $f^2$  corresponds to 0.01.

Table R.2. Estimation Results - Data Used in the Original Study  
(Dependent Variable: Cases of Infection per Million After April  
3, 2020)

<b>Democracy</b>	<b>46.35</b> <b>(0.826)</b>
Temperature	-4.00 (0.941)
Openness	16.71 (0.317)
Constant	2171.69 (0.253)
<b>Adjusted R<sup>2</sup></b>	<b>-0.01</b>
<b>Number of countries</b>	<b>150</b>

Figures in the parenthesis below the coefficients are the p-values.

### *3. Using the whole-time frame*

Table R.3 contains the results of the Ordinary least squares (OLS) regression using the whole-time frame. This is, cases of infection per one million people between December 31, 2019 and August 11, 2020. The coefficient for democracy is equal to 125.31, suggesting a positive association between the democracy index and the total number of confirmed infections per one million people. However, the coefficient is not statistically significant at the 5% level ( $p = 0.558$ ).

**Thus, this replication of the claim was not successful according to the SCORE criteria.** The analytic sample included 150 countries, which did meet the minimum threshold of 38 countries defined by the power analysis. The Cohen's  $f^2$  corresponds to 0.03.

Table R.3. Estimation Results - Data Used in the Original Study  
(Dependent Variable: Cases of Infection per Million between  
December 31, 2019 and August 11, 2020)

<b>Democracy</b>	<b>125.31</b> <b>(0.558)</b>
Temperature	-14.96 (0.785)
Openness	24.52 (0.147)
Constant	1871.04 (0.330)
<b>Adjusted R<sup>2</sup></b>	<b>0.01</b>
<b>Number of countries</b>	<b>150</b>

Figures in the parenthesis below the coefficients are the p-values.

#### ***4. Using time frames different to the ones in the original study***

Table R.4 contains the results of the Ordinary least squares (OLS) regression using time frames different from the original study. This is, cases of infection per one million people between December 31, 2019 and August 11, 2020; and cases of infection per one million people after April 3, 2020. **The coefficient for democracy is equal to 85.83, suggesting a positive association between the democracy index and the total number of confirmed infections per one million people. However, the coefficient is not statistically significant at the 5% level ( $p = 0.565$ ).** **Thus, this replication of the claim was not successful according to the SCORE criteria.** The

analytic sample included 150 countries, which did meet the minimum threshold of 38 countries defined by the power analysis. Note that although the same 150 countries are used, they have duplicate observations in the analysis, leading to a total number of 300 country-date observations. The Cohen's  $f^2$  corresponds to 0.02.

Table R.4. Estimation Results – Time frames not used in the original study (Dependent Variable: Cases of Infection per Million After April 3, 2020; and Cases of Infection per Million Between December 31, 2019 and August 11, 2020)

<b>Democracy</b>	<b>85.83</b> <b>(0.565)</b>
Temperature	-9.48 (0.804)
Openness	20.61 (0.081)
Constant	2021.37 (0.132)
<b>Adjusted R<sup>2</sup></b>	<b>0.01</b>
<b>Number of countries</b>	<b>150</b>

Figures in the parenthesis below the coefficients are the p-values.

### *5. Using all available time frames*

Table R.5 contains the results of the Ordinary least squares (OLS) regression all the available time frames. **The coefficient for democracy is equal to 83.54, suggesting a positive association between the democracy index and the total number of confirmed infections per one million people. However, the coefficient is not statistically significant at the 5% level ( $p = 0.426$ ). Thus, this replication of the claim was not successful according to the SCORE criteria.** The

analytic sample included 150 countries, which did meet the minimum threshold of 38 countries defined by the power analysis. Note that although the same 150 countries are used, they have triplicate observations in the analysis, leading to a total number of 450 country-date observations. The Cohen's  $f^2$  corresponds to 0.02.

Table R.5. Estimation Results – All Available Time Frames  
(Dependent Variable: Cases of Infection per Million on April 3, 2020; Cases of Infection per Million After April 3, 2020; and Cases of Infection per Million Between December 31, 2019 and August 11, 2020)

<b>Democracy</b>	<b>83.54</b> <b>(0.426)</b>
Temperature	-9.97 (0.711)
Openness	16.35 (0.049)
Constant	1247.36 (0.186)
<b>Adjusted R<sup>2</sup></b>	<b>0.01</b>
<b>Number of countries</b>	<b>150</b>

Figures in the parenthesis below the coefficients are the p-values.

## Methods & Materials

The following materials are publicly available on the OSF site:

- The **preregistration** file: [Hossain covid b344 5g09 \(Nast Méndez-Chacón\) Preregistration.pdf](#)

- The **raw data**. No registration is needed to download the necessary data. At the time of the replication the data could be accessed from:

#### Coronavirus infection cases

European Centre for Disease Prevention and Control (2020). Data on geographic distribution of COVID-19 cases worldwide. Retrieved from

<https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>

#### Average yearly temperature

Derived from the Climate Research Unit (Mitchell et al, 2003). Retrieved from

<https://datacatalog.worldbank.org/dataset/climate-change-knowledge-portal-historical-data>

#### Openness

The World Bank. Data on Trade (% of GDP). Retrieved from

<https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS>

#### Democracy index 2019

EIU. Democracy index 2019. Retrieved from

<https://www.gapminder.org/data/documentation/democracy-index/>

Filenames:

- [EIU-Democracy Indices - Dataset - v3 \(1\).xlsx](#)
- [API\\_NE.EXP.GNFS.ZS\\_DS2\\_en\\_excel\\_v2\\_1218003.xls](#)
- [Average yearly temperature Wiki vs. original data.PNG](#)
- [worldbank\\_historical\\_data\\_0.xls](#)
- The **R code** to produce the replication dataset. Filename:
  - [Hossain script.R](#)



- The **replication dataset dictionary**. Filename:
  - [COVID Factors Data Dictionary.xlsx](#)
- The **replication dataset**. Filenames (same data but in different format):
  - [COVID replication.dta](#)
  - [COVID replication.rds](#)
- The **code for replication**. Along with the replication dataset, this is the only file required to replicate the original study. To replicate, just change the working directory to where the data is in your computer and run this file using Stata (the code was written using Stata 15.1). Filename:
  - [Hossain 2020 - Replication Analysis.do](#)
- The **output** from the Stata analyses, available in two formats: smcl (Stata output) and a pdf file. Filenames:
  - For the analysis using the data from the original study:
    - [Hossain\\_Replication\\_Time\\_Original\\_Study.pdf](#)
    - [Hossain\\_Replication\\_Time\\_Original\\_Study.smcl](#)
  - For the analysis using the data after the original study:
    - [Hossain\\_Replication\\_Time\\_After\\_Original\\_Study.pdf](#)
    - [Hossain\\_Replication\\_Time\\_After\\_Original\\_Study.smcl](#)
  - For the analysis using the whole-time frame:
    - [Hossain\\_Replication\\_Whole\\_Frame.pdf](#)
    - [Hossain\\_Replication\\_Whole\\_Frame.smcl](#)
  - For the analysis using the time frames not used in the original study:
    - [Hossain\\_Replication\\_Time\\_Not\\_Original\\_Study.pdf](#)
    - [Hossain\\_Replication\\_Time\\_Not\\_Original\\_Study.smcl](#)

- For the analysis using all available time frames:
  - [Hossain\\_Replication\\_All\\_Available\\_Time\\_Frames.pdf](#)
  - [Hossain\\_Replication\\_All\\_Available\\_Time\\_Frames.smcl](#)

### **Deviations from the Original Study**

1. The unit of analysis of the original study are 163 countries. It is not clear which countries were specifically included in the original analysis. Countries with any missing value for COVID cases, population, temperature, openness, or democracy index, are deleted from the replication dataset. After listwise deletion, the replication dataset includes 150 countries. This difference of 13 countries compared to the original study can potentially be explained by the use of a different data source for the variable “Temperature” (World bank instead of Wikipedia), and the years used for the imputation of the “Openness” variable.

### **Deviations from the Preregistration**

There were no deviations from the preregistration.

### **Citation**

Hossain, Mohammad Alamgir (2020) Is the spread of COVID-19 across countries influenced by environmental, economic and social factors? medRxiv 2020.04.08.20058164; doi: <https://doi.org/10.1101/2020.04.08.20058164>