Who Manipulates Data During Pandemics? Evidence From Newcomb-Benford Law

Vadim S. Balashov (vadim.balashov@rutgers.edu)
Rutgers, The State University of New Jersey
Yuxing Yan
SUNY Geneseo
Xiaodi Zhu
New Jersey City University

Research Article

Keywords: Newcomb-Benford Law, COVID-19, Democracy Index (EIU), Gross Domestic Product (GDP), data manipulation

Posted Date: June 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-555372/v1

Who Manipulates Data During Pandemics? Evidence from Newcomb-Benford Law

Vadim S. Balashov*, Yuxing Yan**, Xiaodi Zhu***

*Corresponding author

Associate Professor, Rutgers School of Business-Camden, Rutgers University-Camden

227 Penn Street, Camden, NJ 08102

Work (856) 225-6706, Cell (225) 202-8432, E-mail: vadim.balashov@rutgers.edu

** Assistant Professor, SUNY at Geneseo

1 College Circle, Geneseo, NY 14454

(585) 245-5260, E-mail: pyan@geneseo.edu

*** Assistant Professor, New Jersey City University

200 Hudson Street, Jersey City, NJ 07305

(201)-200-2408, E-mail: xzhu@njcu.edu

Abstract

The COVID-19 pandemic has spurred controversies related to whether countries manipulate reported data for political gains. We study the association between accuracy of reported COVID-19 data and developmental indicators. We use the Newcomb-Benford law (NBL) to gauge data accuracy. We run an OLS regression of developmental indicators (EIU index, GDP per capita, healthcare expenditures, and universal healthcare coverage index) on goodness-of-fit measures to the NBL. We find that democratic countries, countries with the higher gross domestic product (GDP) per capita, higher healthcare expenditures, and better universal healthcare coverage are less likely to deviate from the Newcomb-Benford law. The relationship holds for the cumulative number of reported deaths and total cases but is more pronounced for the death toll. The findings are robust for second-digit tests, for a sub-sample of countries with regional data, and in relation to the previous swine flu (H1N1) 2009–2010 pandemic. The NBL provides a first screening for potential data manipulation during pandemics. Our study indicates that data from autocratic regimes and less developed countries should be treated with more caution. The paper further highlights the importance of independent surveillance data verification projects.

JEL classification: F5, I10, I18, O1, O57, P52.

Keywords: Newcomb-Benford Law; COVID-19; Democracy Index (EIU); Gross Domestic Product (GDP); data manipulation.

Acknowledgement: The authors report no conflicts of interest. The project received no funding. We thank Junyi Zhang, David Pedersen, Vsevolod Strekalov, and Arkady Kurnosov for useful comments that helped improve the paper. We thank Aleksandr Sabirov for swine flu (H1N1) 2009-2010 pandemic data collection.

1 Introduction

On March 11, 2020, the World Health Organization (WHO) declared the novel coronavirus disease 2019 (COVID-19) a pandemic. With tens of millions of confirmed cases and over three million of deaths, this pandemic has spurred a great number of controversies, including many related to the accuracy of the data countries report. Mass media organizations around the globe argue that many countries have continued to manipulate the data for political or other gains (Alwine and Goodrum Sterling 2020; Cambell and Gunia 2020; Economist 2020; Meyer 2020; Romaniuk and Burgers 2020; Sassoon 2020; Speak 2020; Wood 2020).

In this paper, we study the association between the accuracy of COVID-19 data reported by countries and their macroeconomic and political indicators. Our results show that countries that are more functional democracies, have higher income, and stronger healthcare systems report more accurate data. The relationship exists for the cumulative number of confirmed cases and for the cumulative number of reported deaths; however, the results are more pronounced for the number of deaths.

To gauge data accuracy, we use compliance with the Newcomb-Benford law (NBL), which is an observation that in many naturally occurring collections of numbers the first digit is not uniformly distributed. The numeral "1" will be the leading digit around one-third of the time; the numeral "2" will be the leading digit 18% of the time; and each subsequent numeral, "3" through "9," will be the leading digit with decreasing frequency. One property of the NBL is that manipulated or fraudulent data deviate significantly from the theoretical NBL distribution. Due to the ease of its application and straightforward approach, the law has been extensively used to detect fraud and data manipulations. It has been applied to accounting, finance, macroeconomic, and forensic data to test for data manipulation and fraud. We apply the NBL to COVID-19 data for 185 countries affected by the pandemic. For each country, we first identify the period of exponential growth when the data are expected to obey the NBL. After the country's data reach a plateau, the number of cases stabilizes, and the data are not expected to obey the NBL. During the growth period for each country, we calculate four goodness-of-fit measures to estimate compliance with the NBL and use these measures as proxies for data manipulation.

We then study the relationship between our proxies for accuracy of data and indicators of the strength of the economy, democratic institutions, and healthcare systems. Specifically, we use the regression analysis to find the association between goodness-of-fit measures and the Economist Intelligence Unit Democracy Index, the gross domestic product (GDP) per capita, healthcare expenditures as percentage of GDP, and the Universal Health Coverage Index (UHC). Previous studies have shown in other settings that countries with weaker democracies and less economic development are more likely to manipulate data and have lower transparency.For example, Adsera et al. (2003), Egorov et al. (2009), Magee and Doces (2015), Gehlbach and Sonin (2014), and Guriev and Treisman (2019) focus on data manipulation, and Mitchell (1998), Broz (2002), Djankov et al. (2003), Islam (2006), Lebovic (2006), and Fearon (2011) study data transparency. Judge and Schechter (2009) analyze the quality of survey data using the NBL and find that survey data in developing countries is of poor quality while data from developed countries is of better quality. The governments of such countries fabricate data for political gains and to consolidate power. Autocratic governments control most mass media outlets and often censor inconvenient facts that undermine the ruling regime. COVID-19 presents such a case because the wide spread of the pandemic and high death tolls would send a negative signal to citizens and indicate that the government is incompetent. Autocratic governments would try to downplay the scale of the pandemic for the sake of appearances.

Our main hypothesis is that countries with weaker democracies, and weaker economic and healthcare systems will have lower data accuracy as measured by the NBL goodness-of-fit statistic. Our results in many tests support the hypothesis. We find that our four goodness-of-fit measures (which measure deviations from the theoretical distribution as given by the NBL) are negatively correlated with macroeconomic indicators (Democracy Index, GDP, healthcare expenditures, and UHC). The results are true for the cumulative number of cases and the cumulative number of reported deaths. We also find the result is more pronounced for the reported number of deaths than for the number of confirmed cases. This indicates that, on average, autocratic regimes and poorer countries are more prone to manipulate death tolls than the total number of citizens infected.

We conduct a series of robustness tests and find that our results are not driven by the specific period in which we calculate the goodness-of-fit measures, by small countries, by countries with a small number of cases or deaths, or by countries with extreme deviations from the NBL. We also show that the same relationship between proxies for accuracy of data and economic indicators is observed when we apply the NBL to second digits. One concern of our study is that the proxies for data accuracy are calculated based on limited sample sizes for individual countries. To resolve this potential problem, we confirm our findings for the sub-sample of 50 countries that provide regional data (at a state or province level). Regional data increase the sample size from which we calculate our statistics substantially and heighten the precision of our accuracy measures.

We find a similar relationship for the previous swine flu (H1N1) pandemic of 2009–2010. We repeat our analysis for 35 countries that reported weekly data of the number of confirmed cases and the number of deaths to Pan American Health Organization (PAHO). We discover support for the negative relationship between deviations from the NBL and the selected developmental indicators.

There is substantial body of literature assessing the tendency of misreporting COVID-19 surveillance pandemic data by countries using different statistical techniques, such as case fatality rates, excess mortality rates, the variance of reported data, the clustering of data, and even trends in search engines (Aron and Muellbauer 2020; Dragan 2020; Goutte and Damette 2020; Polson 2020; Roukema 2020). The inherent problem with these methods is their reliance on uniform approaches to measure confirmed cases and COVID-19-related deaths across countries. Even though countries are expected to follow the same guidelines provided by the WHO when reporting cases, many variations exist (and sometimes appear in states and regions within a country) in how they collect and report data. Any comparisons of raw numbers—like the total number of confirmed cases, the number of deaths, and mortality rates—among countries may be driven by the difference in the number of tests conducted, the strength of the healthcare system, demographic composition, and reporting standards. Correct comparisons would require controlling for all those hard-to-observe variables. One helpful property of the NBL is its tolerance to different data generating processes between countries and its sensitivity to human intervention and manipulation of data in otherwise naturally occurring processes. This means that we can apply the test even if countries differ in how they measure COVID-19 cases and related deaths. The test is also free of country-specific differences, including public policies used to stop the pandemic, like quarantines, social distancing, testing, and availability of treatment.

There are some caveats with using the NBL. The NBL is an empirical observation. Departures from something only empirically observed should be treated with caution and do not by default mean causation. In applications, the NBL is usually used as a first filter for detecting fraud.

Deviations from the NBL simply mean that further investigation is needed to identify causes of such anomalies. Our results indicate that data from autocracies and poorer countries should be trusted less. It should also be noted that the NBL test is not directional. However, it is unreasonable to believe that the government would willingly manipulate data to inflate the number of cases or deaths. Neither does the divergence from the NBL provide us specifics on how the data are being manipulated. We cannot ex ante predict which first digits will be over- or underinflated.For example, if the country's true number of cases is in the 2,000s and the government tries to falsify data to look smaller and reports high 1,900s, the first digit "one" will be over-represented in this country's statistics. However, if the country's numbers are in the 1,000s and the governments falsifies data into the 900s, then the first digit "one" will be under-represented.

We stress that our paper is not aimed to answer questions whether a particular government manipulates data. We indicate that such tests are problematic because they largely depend on the sample size. We calculate goodness-of-fit measures for all countries and then compare countries cross-sectionally based on developmental indicators.

Our paper contributes to the literature in several ways. First, our paper helps to resolve the controversy about different countries' data manipulation during pandemics and provides estimates of how widespread it is around the globe. Using the NBL and COVID-19 data, we document that about one-third of the 185 countries affected by the pandemic deviate from the NBL. Second, our study shows which data, if any, countries are more likely to misreport. We document that governments tend to downplay the news with the highest negative impact, i.e., the death toll, to the highest degree. To a slightly lower degree, countries tend to manipulate the total number of confirmed cases. We find no indication, on average, of systematic data manipulation for the number of conducted COVID-19 tests and the number of cured cases. Third, we are the first study to use the NBL to show that the strength of healthcare systems, as measured by healthcare expenditures and the Universal Healthcare Index, are linked to the government's ability to provide reliable data during pandemics. Finally, we are the first to document the cross-sectional link between macroeconomic and political regime indicators and the tendency to misreport data during pandemics. We show that authoritarian regimes and countries with low GDP per capita are more likely to falsify data. Thus, this study provides additional evidence of the link between democracy and transparency that is often taken for granted.

Our study has broad implications. First, we provide evidence that the data supplied during pandemics may be of low quality, especially from autocracies and poorer countries, and we suggest that caution should be used when interpreting and using the data. Second, the study highlights the importance of initiatives to externally verify data provided by governments, including independent surveillance data verification projects.An example of such a project for economic data would be the Billion Prices Project (BPP) by Alberto Cavallo and Roberto Rigobon at MIT Sloan and Harvard Business School. Finally, we provide new evidence on the applicability of the NBL to detect data fabrication during pandemics.

2 Literature Review and Hypotheses Development

Studies have long posed questions about whether democratic regimes provide more reliable data to the public than autocracies in both theoretical and empirical settings. Most studies show that democracies indeed are more transparent (Bueno de Mesquita et al. 2003; Gehlbach and Sonin 2014; Hollyer et al. 2011; Rozenas and Stukal 2019). They find that it is authoritarian regimes that are more vulnerable to negative information and have more incentives to distort and manipulate information that undermines their image. In addition, such regimes usually have control over mass media organizations and therefore have more capabilities to exercise control. Autocrats use data manipulation to improve their public image and prolong their stay in office. Rozenas and Stukal (2019) propose that autocrats are more likely to manipulate data for which it is more difficult for citizens to obtain hard external information benchmarks. COVID-19 provides a unique setting to test a related hypothesis. Pandemic surveillance data are hard to acquire independently by citizens because they lack access to the necessary large-scale data collection and medical facilities. At the same time, the news that the disease is raging and is widespread under authoritarian rule would be an indicator of the inefficiency or failure of the government. The death toll is even more damaging to the image of the autocrat, who sees such news as a threat and tries to downplay the scale of the problem. We therefore formulate the following two hypotheses:

Hypothesis 1: Democratic regimes are less likely to manipulate pandemic surveillance data.

Hypothesis 2: The link between democratic regimes and data manipulation is more pronounced *for the reported death toll.*

We also adopt other measures that may explain countries' tendencies to manipulate data. Hollyer et al. (2011) maintain that GDP per capita is a measure of the "*ability* of the governments to collect and disseminate high-quality statistical data." We therefore include the GDP per capita in our tests. Because our setup has been created during the pandemic and the testing is done on surveillance data, we use two other proxies for each country's ability to collect and report reliable health-related data: health expenditures as a percentage of GDP, and the Universal Health Coverage Index. We thus formulate our third hypothesis as follows:

Hypothesis 3: Countries with higher GDP per capita levels, higher levels of health expenditures as percentage of GDP, and higher levels of the Universal Health Coverage Index are less likely to manipulate pandemic surveillance data.

To gauge data manipulation, we use compliance with the NBL. We are not the only paper to use the NBL to test the validity of reported data during COVID-19 (Idrovo and Manrique-Hernández 2020; Koch and Okamura 2020; Peng and Nagata 2020; Sambridge and Jackson 2020; Zhang 2020). However, many papers usually select one or a few countries and apply the NBL to test if there is any evidence of manipulation in a given country's data. The authors use the cutoff values from the chi-squared distribution (or similar distributions) and give a "yes-or-no" type of answer to their binary research question. These test statistics and inference results greatly depend on the sample size. With large enough sample sizes, the null hypothesis of compliance with the NBL will be rejected in almost every case. Some studies estimate their test statistic at the country level, some studies estimate it at a regional or state level, and some studies use county-level data. This leads to contradictory findings among these studies even when looking at the same country. Sambridge and Jackson (2020) analyze reports from all countries affected by COVID-19 and conclude that the numbers tend to obey the NBL for infections and deaths prior to countries leveling out. Like Sambridge and Jackson (2020), we use all countries affected by COVID-19. We employ the same approach for all countries to calculate the test statistic and study the link between compliance with the NBL and economic indicators. We make any inferences from the NBL test only in comparison. To our best knowledge, this is the first paper to examine the cross-section of all countries and compare them based on developmental indicators when analyzing data manipulation during pandemics.

Not many studies apply the NBL in an international setting, though there are several notable exceptions. Nye and Moul (2007) indicate that international macroeconomic data generally conforms the NBL. They find, however, that for non-OECD (African) countries, the data do not conform with the law, which raises questions about data quality and manipulation in these countries. Gonzalez-Garcia (2009) uses a similar approach to test the annual IMF data, but finds no connection between independent assessments of data quality and adherence to the first-digit NBL in different country groups. Michalski and Stoltz (2013) provide a theoretical model and empirical findings that some countries strategically misreport their economic data for short-term government gains. We contribute to this body of literature by applying the NBL to the pandemic data in the international setting and providing additional evidence that some types of countries are more likely to falsify not only macroeconomic data but also surveillance data during pandemics.

3 Newcomb-Benford Law of Anomalous Numbers

In many naturally occurring processes, the resulting data have the leading significant digit that is not uniformly distributed. The distribution is monotonically decreasing, with "1" being the most common first digit, and "9" being the least common. The law was formally stated by Newcomb (1881) and Benford (1938). A set of numbers is said to follow the NBL if the first digit d occurs with probability $P(d) = log_{10}(1+\frac{1}{d})$. The law can be extended to digits beyond the first. In general, for the nth digit, $n \geq 2$, the probability is given by $P(d) = \sum_{k=10^{n-2}}^{10^{n-1}-1} log_{10}(1 + \frac{1}{10k+d})$. This gives the following probabilities for observing the first and second digits:

Digit	0	1	2	3	4	5	6	7	8	9
First	-	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%
Second	12.0%	11.4%	10.9%	10.4%	10.0%	9.7%	9.3%	9.0%	8.8%	8.5%

The NBL accurately describes many real-life sets of numerical data, including lengths of rivers, stock prices, street addresses, accounting data, populations, physical constants, and regression coefficients (Diekmann 2007). Data generated from many distributions and integer sequences have been shown to closely obey the NBL, including Fibonacci numbers, powers of numbers, exponential growth, many ratio distributions, and the F-distribution with low degrees of freedom (For more examples, see Formann (2010), Hill et al. (1995), Hill (1998), Leemis et al. (2000), and Morrow (2020)).

Not all distributions generate data that follow the law. For example, uniform distribution, normal distribution, and square roots of numbers do not obey it. For the data to obey the NBL, several criteria should be satisfied (Cho and Gaines 2007; Diekmann 2007; Durtschi et al. 2004): a) data span several orders of magnitude and are relatively uniform over such orders; b) the mean is greater than the median, with a positive skewness; c) naturally occurring processes, the data which are the result of multiplicative fluctuations, and data that is not influenced by human intervention.

The last requirement, i.e., the fact that human intervention usually generates data that violates the NBL, has led to its usefulness in detecting fraud and data manipulation. Studies have shown that when humans intervene with the data generating process that is expected to comply with the NBL, compliance stops. For example, Diekmann (2007) and Horton et al. (2018) show that when scientific data are fabricated, they do not conform with the NBL. Cantu and Saiegh (2010) and Breunig and Goerres (2011) reveal the same effect for electoral data. Kaiser (2019) uncovers how discrepancies from the target NBL distribution can be used to test reliability among survey data sets.

Overall, the NBL has been used to detect fraud in: a) scientific studies (Diekmann 2007; Geyer and Williamson 2004; Horton et al. 2018; Judge and Schechter 2009), b) accounting (Durtschi et al. 2004; Horton et al. 2018; Nigrini 2012; Stambaugh et al. 2012; Suh et al. 2011; Varian 1972), c) election data (Breunig and Goerres 2011; Cantu and Saiegh 2010; Cho and Gaines 2007; Deckert et al. 2011; Mebaner 2006), d) macroeconomic data (Gonzalez-Garcia 2009; Hussain 2010; Kalaichelvan and Shawn 2012; Michalski and Stoltz 2013; Nye and Moul 2007; O'Keefe and Yom 2017; Rauch, Goettsche, et al. 2013; Rauch, Göttsche, et al. 2011), e) forensic analysis (Pinilla et al. 2018), f) tax evasion (Demir and Javorcik 2018; Nigrini 1996), g) toxic release inventory (Marchi and Hamilton 2006), h) reported data during pandemics (Gómez-Camponovo et al. 2016; Idrovo, Fernández-Niño, et al. 2011; Idrovo and Manrique-Hernández 2020).

Another useful property of the data obeying the NBL is that it is scale invariant, i.e., it is independent of the measurement units. This makes it a powerful tool when testing data from different sources (i.e., countries, companies). The NBL is also not the same as the imprecision (or variance) of the data. The data may well be very noisy but still is expected to conform with the law, as long as there is no deliberate falsification of data. For example, if a country's data are collected with error or irregularly but there are no data manipulations, the first digit should still adhere to the NBL. In our application, it means that countries may differ in the way they count COVID-19 cases or deaths, but as long as the data for each country is expected to obey the NBL, we can test the data for the goodness-of-fit to the NBL.

4 Data and Variables

4.1 Goodness-of-Fit Measures

To measure how well the data comply with the NBL, we use several goodness-of-fit measures. The most intuitive and commonly used is the chi-squared statistic:

$$Chi\text{-}sq. = \sum_{d=1}^{9} \frac{(O_d - E_d)^2}{E_d}, \tag{1}$$

where O_d and E_d are observed and expected by the NBL frequencies for digit d, respectively. Chi-squared, however, has several problems: it has low statistical power when used with small sample sizes and enormous power with large sample sizes. Therefore, we use alternative measures of goodness-of-fit proposed in extant studies. We use a modified version of the Kuiper (1960) test proposed by Stephens (1970) and Giles (2007) that is less dependent on the sample size N:

$$Kuiper = (D_N^+ + D_N^-)\left[\sqrt{N} + 0.155 + \frac{0.24}{\sqrt{N}}\right], \text{ and} \tag{2}$$

$$D_N^+ = \sup_{-\infty<x<+\infty} |F_N(x) - F_0(x)| \text{ and } D_N^- = \sup_{-\infty<x<+\infty} |F_0(x) - F_N(x)|, \tag{3}$$

where $F_N(x)$ and $F_0(x)$ are the observed cumulative distribution functions (*cdf*) of leading digits and the *cdf* of the data that comply with the NBL. In addition, we calculate the M-statistic proposed by Leemis et al. (2000):

$$M = \max_{d=1}^{9} |o_d - e_d| \sqrt{N} \tag{4}$$

and the D-statistic proposed by Cho and Gaines (2007):

$$D = \sqrt{N \sum_{d=1}^{9} (o_d - e_d)^2} \tag{5}$$

where o_d and e_d are the proportions of observations with d as the first digit and proportions expected by the NBL, respectively. The latter two measures are also insensitive to sample size.

4.2 Sample Description and Developmental Indicators

We first collect daily data from John Hopkins University for the cumulative number of confirmed cases, the cumulative number of cured cases, and the cumulative number of deaths between January 22, 2020 and June 10, 2020. We also obtain the number of conducted tests from Our World in Data (https://ourworldindata.org/coronavirus-testing). Studies have shown that naturally occurring processes comply well with the NBL when the data grow exponentially or close to it (Formann 2010; Leemis et al. 2000). Once the data reach the plateau, they are no longer expected to obey the NBL. Hence, for the data to comply with the NBL, we select the growth part using the following approach. Because data show weekly seasonality, we first compute seven-day moving averages (MA) for the new daily number of confirmed cases. Then, for each country, we identify the date with the highest MA number of new daily confirmed cases. For our main analyses, we use data before the obtained cutoff for each country. In unreported tests, we also use modified approaches. We find the maximum ratio MA(number of new daily cases)/(Days since the first case for the country) and MA(number of new daily cases)/(Days since the latest nonzero case for the country). The results are robust to alternative definitions of the cutoff date.

For developmental indicators, we select the following four proxies for democratic and economic development widely used in the literature: the *Economist Intelligence Unit* Democracy Index, GDP per capita, healthcare expenditures as a percentage of GDP, and the Universal Health Coverage Index. We use GDP per capita as a proxy for the country's ability to provide precise data. We also take the country's healthcare spending as a percentage of GDP and its Universal Health Coverage Index as proxies for the strength of each country's healthcare system. We download countries' democracy indices from the *Economist Intelligence Unit* for 2019. We collect the Gross Domestic Product (GDP) per capita, healthcare expenditures as percentage of GDP (HE_GDP), and Universal Health Coverage Index (UHC) for 2017 from the World Bank (https://data.worldbank.org/). At the time we collected the data, many countries still did not have the World Bank data available for 2018 or 2019. 2017 is the latest year for which the data are available for all countries. We also acquire 2019 population data for each country from Worldometer. A total of 185 countries with available data were affected by COVID-19. Variable definitions can be found in Table 6.

We find that we cannot reject the NBL distribution for the entire world population for the cumulative number of confirmed cases using the 1% significance level. The 1% threshold for all four measures are: 20.09 for Chi-squared, 2.00 for Kuiper, 1.21 for M, and 1.57 for D. For the cumulative number of reported deaths, however, we reject the hypothesis of compliance with the NBL for the total world numbers. Using country-level data, we also find that between 37 and 62 countries (depending on the goodness-of-fit measure used) out of 185 deviate from the NBL when reporting the confirmed cases. Between 50 and 71 countries deviate from the NBL when reporting the number of deaths. The values go up to between 62 and 103 and between 72 and 99 countries for the number of cases and deaths, respectively, when the 10% level of significance is used. Note also that switching from country-level data to state data or county-level data increases the statistics significantly. In our analyses, however, we avoid country-specific conclusions based on goodness-of-fit measures that are largely dependent on a sample size, we aim to compare countries cross-sectionally.

Table 1 provides descriptive statistics for the major variables in our analyses. Observe that the corresponding mean goodness-of-fit measures for the number of deaths are higher than for the number of confirmed cases and are lagrely not consistent with the NBL at 1%. This is consistent with countries, ceteris paribus, being more prone to manipulate data on death rates.

[Insert Table 1 around here]

Figure 1 provides intuition for the main findings in our paper and plots values of the Kuiper goodness-of-fit measure for the four quartiles of each of our independent variables: EIU, $\ln(GDP)$,

HE_GDP, and UHC. The quartiles for the EIU Democracy Index roughly correlate with the definitions of the four regime types: full democracy, flawed democracy, hybrid regime, and authoritarian regime. The figure shows a general trend for the data to deviate less from the Newcomb-Benford distribution as we move from the smallest quartile to the largest. The trend is more pronounced for the death toll: for the smallest quartile, the deviation from the NBL is significant at 1%, whereas for the largest quartile it is not significant. In a univariate setting, this is consistent with our three hypotheses.

[Insert Figure 1 around here]

We also find that the four major economic indicators are highly correlated, with correlation coefficients ranging between 0.37 and 0.85 (all values are statistically significant, untabulated). These variables are most likely proxies for the same indicator, the development level of a country, and therefore—to avoid multicollinearity—we include only one indicator at a time in our analysis.In unreported tests we put all economic indicators together in one equation on the right-hand-side and test for collinearity. The result shows the Condition Number is over 46, indicative of serious collinearity. The four goodness-of-fit measures are also highly correlated with each other, as are the total number of confirmed cases and the country's population.

5 Results

5.1 Goodness-of-fit and Economic Indicators

We start with the simple ordinary least squares (OLS) regression model where our goodness-of-fit measures appear on the left-hand-side and economic indicators are on the right-hand-side:

$$\begin{aligned} Goodness\text{-}of\text{-}fit_i &= \beta_0 + \beta_1 \mathbf{Indicator}_i + \beta_2 \ln(Population)_i + \\ &\quad + \beta_3 Number\_of\_Days_i + \varepsilon_i, \end{aligned} \tag{6}$$

where $Indicator_i$ denotes one of the four economic indicators: EIU, $\ln(GDP)$, HE_GDP, or UHC. Higher values of the goodness-of-fit measures indicate greater deviation from the NBL. If more developed countries are less likely to manipulate data, we expect the coefficient β_1 to be negative.

How well the data for each country are *expected* to obey the NBL depends on the span. For example, countries with higher populations and more confirmed cases or deaths are expected to follow the NBL more closely. To control for that, we include the natural logarithm of the country's total population. Alternatively, we include the very correlated number of confirmed cases (or deaths). We find that the results are qualitatively and quantitatively the same when we used alternative control variables (untabulated). Even though the Kupier, M, and D-statistics are more independent of the sample size, goodness-of-fit measures may still be affected by the sizes of the samples used to estimate them. To control for the sample size effect, we include $Number\_of\_Days_i$, which is the number of days with nonzero confirmed cases (or the number of days with nonzero deaths) between January 22, 2020 and the cutoff date for the growth part for each country.

The results of estimating Equation 6 are presented in Table 2. Panel A provides estimates for the cumulative number of confirmed cases. All but one coefficient in front of economic indicators are negative. Coefficients for $\ln(GDP)$ and UHC are always significant. The coefficient for EIU is significant only when the chi-squared goodness-of-fit measure is used, and the coefficient for HE_GDP lacks significance in all tests. Panel B provides estimates for the cumulative number of deaths. All coefficients are negative, and all are significant. The magnitude of coefficients for the number of deaths is also much higher than that for the number of confirmed cases. We find that the coefficients for corresponding economic indicators are statistically different from each other between Panels A and B in each case in Table 2.

[Insert Table 2 around here]

We interpret the data as being consistent with the argument that more democratic and more highly developed countries are less likely to deviate from the NBL when reporting pandemic data. Specifically, countries with higher GDP per capita and universal health coverage are less likely to manipulate their data during COVID-19. For the democracy index (EIU) and health expenditures as percentage of GDP (HE_GDP), we find convincing evidence only for the number of deaths. We conclude that the relationship is more pronounced for the total number of deaths than for the number of confirmed cases. As predicted, the control variable $\ln(Population)_i$ is negative and significant in all regressions.

The results are also economically significant: an increase of one standard deviation in the economic indicators, on average, results in a 0.25 increase of the standard deviation in the goodness-of-fit measures. This value is roughly the same for the number of confirmed cases and for the number of deaths, across all economic indicators. Using the predicted values from the model, for the top quartiles, we cannot reject the hypothesis that countries manipulate confirmed cases or death data at the 1% level. For the bottom quartile, however, we reject that hypothesis in almost all specifications.

Finally, we conduct the same tests for the cumulative number of cured cases and the number of COVID-19 tests conducted. On average, we cannot reject the null hypothesis that countries manipulated data on cured cases or the number of conducted tests. The regression results are also not significant, indicating systematic cross-sectional difference between countries. We conclude that countries are most prone to falsify mortality data, slightly less so the number of confirmed cases, and that there is no evidence of systematic data falsification of cured cases or the number of tests. The cross-sectional difference between countries is also the strongest for the death toll, weaker for the total number of confirmed cases, and is insignificant for the number of cured cases and tests.

5.2 Robustness Analyses

One limitation of our analysis above is that it depends on the cutoff date for the growth part of the data. We try several different approaches. First, we use the same, "global," cutoff value for all countries, which is 80 days since January 22, 2020 (or April 11, 2020). We pick 80 days because it corresponds to the second tercile of cutoff dates in our sample. Anything much sooner will result in too small sample sizes for many countries, especially for the ones that were affected by the pandemic later than others. Anything much later, and too many countries will have already reached their plateaus, and are not expected to obey the Newcomb-Benford law anymore. Alternatively, we try the same cutoff number of days since the first case for each country. Again, we pick the second tercile, which is 45 days since the first case for the country. The unreported results are largely consistent with our previous findings.

Another concern is that our data might be driven by countries with few cases or few data points. To test for that, we exclude countries with lower than 200 (500 and 1,000) total number of confirmed cases. We then exclude countries with fewer than 30 (40) days of nonzero cases. We also tried excluding countries with the highest 1% (5%) goodness-of-fit measures. In reported tests we find that the results are robust in all cases. We conclude that our results are not driven by the specific pick of the cutoff date, by small countries, countries with a small number of cases, or by extreme deviations from the NBL.

5.3 Regional Data

Testing for compliance with the NBL requires sufficient data. For many countries in our analysis, the goodness-of-fit measures are calculated based on relatively small samples sizes between 40 and 140 days. Even though we control for the sample size and conduct robustness checks, making inferences from results based on such small sample sizes might be problematic. The sample size may increase significantly for a country if it reports data at a regional (state, territory, or provinces) level. Each reported value at the regional level can then be used to estimate goodness-of-fit measures, instead of using the country-level data. The method has the upside that the goodness-of-fit measures are estimated with a greater precision, though the downside is the lack of countries that collect regional data.

Fifty out of 185 countries in our sample collect regional-level data. Regional data are from the COVID-19 Coronavirus Map (https://covid19.health/). We check for data consistency between the two data sources and find the high degree of agreement. For these countries, we re-estimate the goodness-of-fit measures and re-run Equation 6. The results are reported in Table 3. For the second-digit test to work, the data should be at least over ten. Panel A depicts the confirmed number of cases; Panel B the number of deaths. For the cumulative number of cases, 12 out of 16 coefficients are negative, with eight demonstrating significance. For the cumulative number of deaths, all coefficients are negative and significant, even with the much smaller number of countries for these tests. The results are consistent with our earlier finding: countries with higher democracy indices, GPD per capita, health expenditures, and universal healthcare coverage are less likely to manipulate pandemic data, especially the number of deaths. We further conclude that our findings are not driven by the errors in goodness-of-fit measures.

[Insert Table 3 around here]

5.4 Second Digit Tests

The NBL can be extended to digits beyond the first (O'Keefe and Yom 2017 and Hussain 2010). Beyond the second digit, the theoretical distribution quickly converges to uniform. Diekmann (2007) notes that, when fabricating data, test subjects also naturally lean toward smaller first digits, resulting in Benford-like distributions of fabricated data. He suggests that in some cases the second-digit test may provide a clearer assessment of data manipulation. Therefore, we repeat our tests but use the second-digit goodness-of-fit measures instead of the leading digit. Our sample size drops somewhat, especially for the number of deaths, because the test requires values higher than ten. The results are presented in Table 4, again, with two panels: one for the confirmed number of cases, and one for the number of deaths. In Panel A, all coefficients are negative and nine out of 16 coefficients are significant. In Panel B, all coefficients are negative, and, except for two coefficients for UHC, all are significant. We conclude that second-digit test results accord with our main findings.

[Insert Table 4 around here]

5.5 Swine Flu Pandemic of 2009-2010

A natural extension to our study is to see if the negative relationship between goodness-of-fit measures and economic indicators holds for other pandemics. Pandemics that engulf many countries and for which data are available are rare in modern history. One natural candidate is the recent swine flu (H1N1) pandemic of 2009–2010. Swine flu (H1N1) 2009–2010 was a pandemic that lasted over 19 months between January of 2009 and August 2010. The pandemic affected 58 countries, with tens of thousands (in some estimates, millions and even hundreds of millions) of people infected and tens of thousands (in some estimates, hundreds of thousands) of deaths.

Even though the pandemic happened in relatively recent times and after the advent of the Internet, surveillance data availability and reporting was much more limited in 2009 than during COVID-19. Many countries did not collect daily or even weekly data, reporting was limited, and there was little public availability of data. As a result, only a very small number of studies directly test the accuracy or manipulation of data during the swine flu (H1N1) 2009–2010 pandemic (a notable exception is the study by Idrovo, Fernández-Niño, et al. 2011). The WHO, Pan American

Health Organization (PAHO), and the Center for Disease Control and Prevention (CDC) provide many *estimates* for the total number of cases and deaths, but these cannot be used with the Newcomb-Benford test because the test gauges human intervention in *actual* reported data.

To apply the NBL test, we collect data for 35 countries in the Americas that provided weekly reports of the number of confirmed cases and deaths to the PAHO. We obtain the data for the weekly number of confirmed cases and the distribution of first digits from Idrovo, Fernández-Niño, et al. (2011). The data for the weekly number of deaths is downloaded from the PAHO website (https://www.paho.org/hq/images/atlas/en/atlas.html?detectflash=false). We then repeat the analyses and re-estimate regression 6 for swine flu (H1N1) 2009-2010 data. For the economic indicators, we use 2009 values. The results are reported in Table 5. Panel A depicts the results for the number of confirmed cases. Out of 16, 12 coefficients in front of macroeconomic indicators are negative. It should be noted that the sample size for this test is extremely small, with at most 35 countries. Obtaining significant results with such small sample sizes is challenging. Yet, we are able to obtain significant coefficients for five coefficients, and two more just barely lack significance. Panel B illustrates the results for the number of deaths. The sample size for Panel B is even smaller: 14 out of 16 coefficients are negative, and seven are significant. We conclude that the swine flu (H1N1) 2009–2010 results are largely consistent with our findings for the COVID-19 pandemic.

[Insert Table 5 around here]

6 Discussion and Conclusion

In this paper, we investigate the relationship between the accuracy of reported data and macroeconomic indicators for a set of 185 countries affected by the COVID-19 pandemic. We use the deviation from the Newcomb-Benford law of anomalous numbers as a proxy for data manipulation. For approximately one-third of countries, we document some evidence of data manipulation, especially for the death toll. We find the negative relationship between the four NBL goodness-of-fit measures and four economic indicators. We also find that the relationship is stronger for the number of deaths than for the number of confirmed cases. Overall, we conclude that democratic regimes and more economically developed countries provide more accurate data during pandemics. We also show that the relationship holds in alternative specifications, for 50 countries that report regional data, for second-digit tests, and for the previous swine flu (H1N1) pandemic.

The interpretations of our findings assume that deviations from the NBL are indicative of data manipulation. Indeed, many studies in macroeconomic, accounting, finance, and forensic analysis demonstrate that human intervention and data manipulation create data sets that violate the NBL. However, several limitations to our study should be mentioned. Caution should be used when applying the NBL. Deviations from the NBL should be only treated as first filters. The aim of this paper is not to provide evidence whether a particular country manipulates data. Such claims require further investigations. We should also note that there is still some chance that the divergence from the expected NBL distribution is due to the low quality or structural breaks in the data.

Our paper highlights the importance of independent projects to verify data supplied by the governments. The paper also leads to a question about whether falsifying data during pandemics is a short- or long-lived strategy for governments. Further research is needed that would combine different methods that test for data manipulation, including the Newcomb-Benford law, biostatistics, moments of distributions, excess mortality rates, and social media data. Even more important is research related to methods that can prevent data manipulation and fraud during pandemics.

References

Adsera, A., Boix, C., & Payne, M. (2003). Are you being served? Political accountability and quality of government. *The Journal of Law, Economics, and Organization*, *19*(2), 445–490.

Alwine, J., & Goodrum Sterling, F. (2020). Manipulation of pandemic numbers for politics risks lives [Accessed: 2020-09-16].

Aron, J., & Muellbauer, J. (2020). A pandemic primer on excess mortality statistics and their comparability across countries. [Accessed: 2020-09-16].

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American philosophical society*, 551–572.

Breunig, C., & Goerres, A. (2011). Searching for electoral irregularities in an established democracy: Applying Benford's law tests to Bundestag elections in Unified Germany. *Electoral studies*, *30*(3), 534–545.

Broz, L. J. (2002). Political system transparency and monetary commitment regimes. *International Organization*, 861–887.

Bueno de Mesquita, B., Smith, A., Siverson, R. M., & Morrow, J. D. (2003). *The logic of political survival*. Cambridge, MA: The MIT Press.

Cambell, C., & Gunia, A. (2020). China says it's beating coronavirus. But can we believe its numbers? [Accessed: 2020-09-16].

Cantu, F., & Saiegh, S. M. (2010). A supervised machine learning procedure to detect electoral fraud using digital analysis. *Available at SSRN 1594406*.

Cho, W. K. T., & Gaines, B. J. (2007). Breaking the (Benford) law: Statistical fraud detection in campaign finance. *The american statistician*, *61*(3), 218–223.

Deckert, J., Myagkov, M., & Ordeshook, P. C. (2011). Benford's law and the detection of election fraud. *Political Analysis*, *19*(3), 245–268.

Demir, B., & Javorcik, B. K. S. (2018). *Forensics, elasticities and benford's law: Detecting tax fraud in international trade*. Centre for Economic Policy Research.

Diekmann, A. (2007). Not the first digit! Using benford's law to detect fraudulent scientific data. *Journal of Applied Statistics*, *34*(3), 321–329.

Djankov, S., McLiesh, C., Nenova, T., & Shleifer, A. (2003). Who owns the media? *The Journal of Law and Economics*, *46*(2), 341–382.

Dragan, A. (2020). Kak uvidet jepidemiju, esli ejo staratelno prjachut. Opyt pjati rossijskih regionov. (in Russian). [Accessed: 2020-09-16].

Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of forensic accounting*, *5*(1), 17–34.

Economist, T. (2020). Tracking covid-19 excess deaths across countries [Accessed: 2020-09-16].

Egorov, G., Guriev, S., & Sonin, K. (2009). Why resource-poor dictators allow freer media: A theory and evidence from panel data. *American political science Review*, 645–668.

Fearon, J. D. (2011). Self-enforcing democracy. *The Quarterly Journal of Economics*, *126*(4), 1661–1708.

Formann, A. K. (2010). The Newcomb-Benford law in its relation to some common distributions. *PloS one*, *5*(5), e10541.

Gehlbach, S., & Sonin, K. (2014). Government control of the media. *Journal of Public Economics*, *118*, 163–171.

Geyer, C. L., & Williamson, P. P. (2004). Detecting fraud in data sets using Benford's law. *Communications in Statistics-Simulation and Computation*, *33*(1), 229–246.

Giles, D. E. (2007). Benford's law and naturally occurring prices in certain ebay auctions. *Applied Economics Letters*, *14*(3), 157–161.

Gómez-Camponovo, M., Moreno, J., Idrovo, Á. J., Páez, M., & Achkar, M. (2016). Monitoring the Paraguayan epidemiological dengue surveillance system (2009-2011) using Benford's law. *Biomédica*, *36*(4), 583–592.

Gonzalez-Garcia, J. (2009). *Benford's law and macroeconomic data quality*. International Monetary Fund.

Goutte, S., & Damette, O. (2020). The macroeconomic determinants of COVID19 mortality rate and the role of post subprime crisis decisions. *Available at SSRN 3610417*.

Guriev, S., & Treisman, D. (2019). Informational autocrats. *Journal of Economic Perspectives*, *33*(4), 100–127.

Hill, T. P. et al. (1995). A statistical derivation of the significant-digit law. *Statistical science*, *10*(4), 354–363.

Hill, T. P. (1998). The first digit phenomenon: A century-old observation about an unexpected pattern in many numerical tables applies to the stock market, census statistics and accounting data. *American Scientist*, *86*(4), 358–363.

Hollyer, J. R., Rosendorff, P. B., & Vreeland, J. R. (2011). Democracy and transparancy. *The Journal of Politics*, *73*(4), 1191–1205.

Horton, J., Krishnakumar, D., & Wood, A. (2018). Detecting academic fraud in accounting research: The case of Professor James Hunton. *Available at SSRN 3164961*.

Hussain, S. A. (2010). The application of Benford's law in forensic accounting: An analysis of credit bureau data. *Available at SSRN 1626696*.

Idrovo, A. J., Fernández-Niño, A., Bojórquez-Chapela, I., & Moreno-Montoya, A. (2011). Performance of public health surveillance systems during the influenza A (H1N1) pandemic in the Americas: Testing a new method based on Benford's law. *Epidemiology and Infection*, *139*(12), 1827–1834.

Idrovo, A. J., & Manrique-Hernández, E. F. (2020). Data quality of Chinese surveillance of COVID-19: Objective analysis based on WHO's situation reports. *Asia-Pacific Journal of Public Health*.

Islam, R. (2006). Does more transparency go along with better governance? *Economics and Politics*, *18*(2), 121–167.

Judge, G., & Schechter, L. (2009). Detecting problems in survey data using Benford's law. *Journal of Human Resources*, *44*(1), 1–24.

Kaiser, M. (2019). Benford's law as an indicator of survey reliability—Can we trust our data? *Journal of Economic Surveys*, *33*(5), 1602–1618.

Kalaichelvan, M., & Shawn, L. K. J. (2012). A critical evaluation of the significance of round numbers in major European stock indices in light of the predictions from Benford's law. *International Research Journal of Finance and Economics*, (95), 196–210.

Koch, C., & Okamura, K. (2020). Benford's law and COVID-19 reporting. *Available at SSRN 3586413*.

Kuiper, N. H. (1960). Tests concerning random points on a circle. *Nederl. Akad. Wetensch. Proc. Ser. A*, *63*(1), 38–47.

Lebovic, J. H. (2006). Democracies and transparency: Country reports to the UN Register of Conventional Arms, 1992-2001. *Journal of Peace Research*, *43*(5), 543–562.

Leemis, L. M., Schmeiser, B. W., & Evans, D. L. (2000). Survival distributions satisfying Benford's law. *The American Statistician*, *54*(4), 236–241.

Magee, C. S. P., & Doces, J. A. (2015). Reconsidering regime type and growth: Lies, dictatorships, and statistics. *International Studies Quarterly*, *59*(2), 223–237.

Marchi, S., & Hamilton, J. T. (2006). Assessing the accuracy of self-reported data: An evaluation of the toxics release inventory. *Journal of Risk and Uncertainty*, *32*, 57–76.

Mebaner, W. R. J. (2006). Election forensics: Vote counts and Benford's law. *Summer Meeting of the Political Methodology Society, UC-Davis, July*, 20–22.

Meyer, H. (2020). Experts question Russian data on Covid-19 death toll [Accessed: 2020-09-16].

Michalski, T., & Stoltz, G. (2013). Do countries falsify economic data strategically? Some evidence that they might. *Review of Economics and Statistics*, *95*(2), 591–616.

Mitchell, R. B. (1998). Sources of transparency: Information systems in international regimes. *International Studies Quarterly*, *42*(1), 109–130.

Morrow, J. (2020). Benford's law, families of distributions and a test basis. *Working Paper*.

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, *4*(1), 39–40.

Nigrini, M. J. (1996). A taxpayer compliance application of Benford's law. *The Journal of the American Taxation Association*, *18*(1), 72.

Nigrini, M. J. (2012). *Benford's law: Applications for forensic accounting, auditing, and fraud detection* (Vol. 586). John Wiley; Sons.

Nye, J., & Moul, C. (2007). The political economy of numbers: On the application of Benford's law to international macroeconomic statistics. *The BE Journal of Macroeconomics*, *7*(1).

O'Keefe, J. P., & Yom, C. (2017). Offsite detection of insider abuse and bank fraud among US failed banks 1989-2015. *Available at SSRN 3013174*.

Peng, Y., & Nagata, M. H. (2020). Statistical analysis of the Chinese COVID-19 data with Benford's law and clustering [Accessed: 2020-09-16].

Pinilla, J., López-Valcárcel, B. G., González-Martel, C., & Peiro, S. (2018). Pinocchio testing in the forensic analysis of waiting lists: Using public waiting list data from Finland and Spain for testing Newcomb-Benford's law. *BMJ open*, *8*(5), e022079.

Polson, D. (2020). Manipulated, agenda-driven data [Accessed: 2020-09-16].

Rauch, B., Goettsche, M., & El Mouaaouy, F. (2013). LIBOR manipulation–empirical analysis of financial market benchmarks using Benford's law. *Available at SSRN 2363895*.

Rauch, B., Göttsche, M., Engel, S., & Brähler, G. (2011). Fact and fiction in EU-governmental economic data. *German Economic Review*, *12*(3), 243–255.

Romaniuk, S. N., & Burgers, T. (2020). Can China's COVID-19 statistics be trusted? [Accessed: 2020-09-16].

Roukema, B. (2020). *Anti-clustering in the national SARS-CoV-2 daily infection counts*.

Rozenas, A., & Stukal, D. (2019). How autocrats manipulate economic news: Evidence from Russia's state-controlled television. *Journal of Politics*, *81*(3), 982–996.

Sambridge, M., & Jackson, A. (2020). National COVID numbers-Benford's law looks for errors. *Nature*, *581*(7809), 384–384.

Sassoon, A. M. (2020). Florida's scientist was fired for refusing to 'manipulate' COVID-19 data [Accessed: 2020-09-16].

Speak, C. (2020). What's the problem with Italy's official coronavirus numbers? [Accessed: 2020-09-16].

Stambaugh, C., Tipgos, M. A., Carpenter, F., & Smith, M. (2012). Using Benford analysis to detect fraud. *Internal auditing*, *27*(3), 24–29.

Stephens, M. A. (1970). Use of the Kolmogorov–Smirnov, Cramer–Von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *32*(1), 115–122.

Suh, I., Headrick, T. C., & Minaburo, S. (2011). An effective and efficient analytic technique: A bootstrap regression procedure and Benford's law. *Journal of Forensic and Investigative Accounting*.

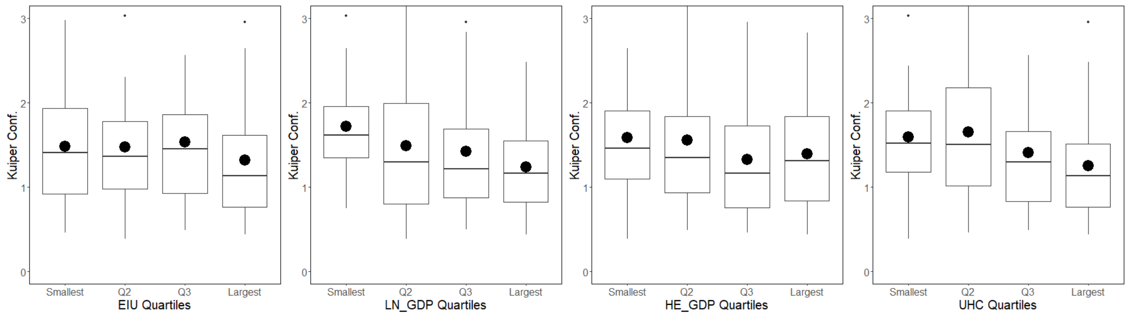
Varian, H. R. (1972). Benford's law. *American Statistician*, *26*(3), 65.

Wood, G. (2020). Iran has far more coronavirus cases than it is letting on [Accessed: 2020-09-16].

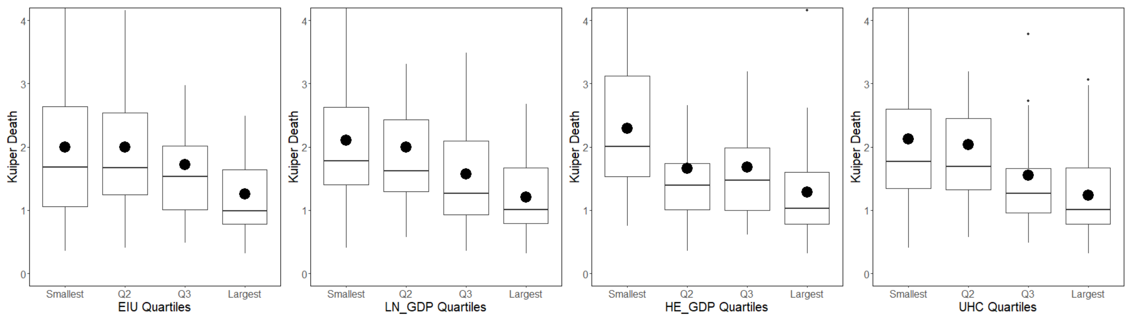
Zhang, J. (2020). Testing case number of coronavirus disease 2019 in China with Newcomb-Benford law [Available on: https://arxiv.org/pdf/2002.05695.pdf]. *arXiv preprint arXiv:2002.05695*.

Figure 1: Goodness-of-fit Measures and Developmental Indicators

Panel A: Confirmed Cases



Panel B: Death Cases



Comparison between goodness-of-fit measures by four quartiles of developmental indicators. The figure presents the boxplot of Kuiper goodness-of-fit measure for the cumulative number of confirmed and death cases by four quartiles fo developmental indicators (EIU, ln(GDP), HE_GDP, UHC). Smallest, Q2, Q3 and Largest represent the values partitioned by quartile 25%, 50%, and 75%. The boxplots represent 25%, 50%, and 75% quartiles. The dots represent the mean values.

Table 1: Descriptive Statistics

Variables	Obs	Mean	Min	Med	Max	Std.
Chi-sq. Conf.	185	19.55^{**}	1.40	13.60^{*}	129.38^{***}	20.22
Kuiper Conf.	185	1.48	0.39	1.37	4.70^{***}	0.73
M Conf.	185	1.08^{**}	0.29	0.90^{*}	4.54^{***}	0.69
D Conf.	185	1.44^{**}	0.42	1.31^{*}	5.04^{***}	0.74
Chi-sq. Death	160	$29.35^{***,3}$	1.71	17.22^{**}	261.49^{***}	37.35
Kuiper Death	160	$1.75^{**,3}$	0.32	1.54	5.81^{***}	1.07
M Death	160	$1.32^{***,3}$	0.19	1.00^{**}	5.68^{***}	0.99
D Death	160	$1.72^{***,3}$	0.39	1.46^{**}	6.10^{***}	1.07
EIU	163	54.84	13.20	56.50	98.70	21.98
ln(GDP)	178	8.68	5.68	8.65	12.03	1.45
HE_GDP	174	6.44	1.18	6.23	17.06	2.57
UHC	175	64.44	25.00	69.00	89.00	15.68
No. of Days Conf.	185	61.24	1.00	61.00	136.00	30.61
No. of Days Death	185	39.59	0.00	36.00	124.00	29.74
ln(Population)	182	15.84	10.43	16.08	21.09	2.02

The table presents the mean value of goodness-of-fit measures (Chi-square, Kuiper, M and D) for the cumulative number of confirmed and death cases, developmental indicators (EIU, ln(GDP), HE_GDP, UHC), and other variables. The number of observations vary due to missing values. ***, ** and * denote goodness-of-fit measures that correspond to significant differences from the theoretical NBL distribution at 1%, 5% and 10% level, respectively. We also analyze the difference between the Confirmed an the Death mean values for each goodness-of-fit measures using the t-test. 3, 2, and 1 indicate significant difference between the Confirmed and the Death cases at the 1%, 5% and 10% level, respectively. All variable definitions are in Table 6.

Table 2: Main Results: OLS Regressions

	Panel A. Confirmed Cases				Panel B. Death Cases			
Variable	**Chi-squared**	**Kuiper**	**M**	**D**	**Chi-squared**	**Kuiper**	**M**	**D**
EIU	-9.54^{*}	-0.19	-0.02	-0.15	-33.61^{***}	-1.25^{***}	-1.09^{***}	-1.36^{***}
	(0.08)	(0.22)	(0.46)	(0.27)	(0.01)	(0.00)	(0.00)	(0.00)
ln(Population)	-3.41^{***}	-0.12^{***}	-0.08^{**}	-0.11^{***}	-6.39^{***}	-0.24^{***}	-0.24^{***}	-0.26^{***}
	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
No. of Days	20.78^{***}	0.93^{***}	0.75^{***}	0.87^{***}	37.45^{***}	1.11^{***}	0.69^{**}	0.96^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Sample Size	162	162	162	162	148	148	148	148
Adj. R^2	10.05%	11.57%	7.12%	9.39%	11.63%	18.08%	15.43%	19.00%
ln(GDP)	-3.41^{***}	-0.13^{***}	-0.07^{**}	-0.12^{***}	-6.53^{***}	-0.25^{***}	-0.22^{***}	-0.27^{***}
	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ln(Population)	-3.93^{***}	-0.15^{***}	-0.11^{***}	-0.14^{***}	-5.71^{***}	-0.22^{***}	-0.19^{***}	-0.21^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
No. of Days	23.13^{***}	0.99^{***}	0.83^{***}	0.95^{***}	36.00^{***}	1.07^{***}	0.66^{**}	0.92^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Sample Size	176	176	176	176	155	155	155	155
Adj. R^2	16.06%	20.39%	11.98%	17.05%	13.89%	22.99%	16.97%	22.21%
HE_GDP	-0.53	-0.01	0.00	-0.01	-3.21^{***}	-0.11^{***}	-0.09^{***}	-0.11^{***}
	(0.17)	(0.38)	(0.45)	(0.34)	(0.00)	(0.00)	(0.00)	(0.00)
ln(Population)	-3.16^{***}	-0.13^{***}	-0.10^{***}	-0.12^{***}	-4.47^{***}	-0.17^{***}	-0.15^{***}	-0.17^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
No. of Days	21.40^{***}	0.99^{***}	0.81^{***}	0.92^{***}	31.12^{***}	1.00^{***}	0.61^{**}	0.83^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.01)
Sample Size	173	173	173	173	151	151	151	151
Adj. R^2	10.94%	13.67%	9.17%	11.15%	11.24%	16.50%	11.56%	15.33%
UHC	-24.33^{***}	-0.81^{***}	-0.46^{*}	-0.81^{***}	-59.61^{***}	-2.09^{***}	-1.87^{***}	-2.33^{***}
	(0.01)	(0.01)	(0.08)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
ln(Population)	-3.97^{***}	-0.14^{***}	-0.10^{***}	-0.13^{***}	-5.49^{***}	-0.21^{***}	-0.19^{***}	-0.21^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
No. of Days	22.75^{***}	0.98^{***}	0.81^{***}	0.93^{***}	38.18^{***}	1.19^{***}	0.77^{***}	1.04^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Sample Size	174	174	174	174	155	155	155	155
Adj. R^2	14.01%	16.57%	10.27%	14.29%	13.73%	20.66%	16.63%	21.52%

The table presents the main results using COVID-19 pandemic data. We estimate equation 6 using OLS for first-digit goodness-of-fit measures. Panel A shows the results for the cumulative number of confirmed cases, while panel B shows the results for the cumulative number of deaths. To avoid small coefficients, we divide EIU, UHC, and No. of Days values by 100 for all models. Sample sizes vary due to missing values. P-values for a one-tailed t-test are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. All variable definitions are in Table 6.

Table 3: Results for 50 Countries with Regional Data

Variable	Panel A. Confirmed Cases				Panel B. Death Cases			
	Chi-squared	**Kuiper**	**M**	**D**	**Chi-squared**	**Kuiper**	**M**	**D**
EIU	27.37	0.09	0.29	0.34	-220.31^{***}	-4.74^{***}	-3.64^{***}	-4.51^{***}
	(0.37)	(0.46)	(0.37)	(0.38)	(0.01)	(0.00)	(0.00)	(0.00)
ln(Population)	1.60	0.01	0.03	0.02	13.16	0.23	0.02	0.13
	(0.44)	(0.49)	(0.41)	(0.44)	(0.14)	(0.13)	(0.44)	(0.25)
No. of Days	3.56^{**}	0.04^{**}	0.03^{**}	0.05^{**}	-1.77	-0.01	-0.02	-0.02
	(0.03)	(0.04)	(0.05)	(0.02)	(0.26)	(0.38)	(0.32)	(0.35)
Sample Size	50	50	50	50	30	30	30	30
Adj. R^2	5.09%	3.55%	2.64%	6.39%	19.78%	32.99%	23.23%	27.44%
ln(GDP)	-21.00^{*}	-0.36^{**}	-0.16	-0.25	-52.00^{***}	-0.78^{***}	-0.36^{*}	-0.70^{**}
	(0.09)	(0.04)	(0.17)	(0.13)	(0.00)	(0.01)	(0.08)	(0.01)
ln(Population)	-2.05	-0.04	0.00	-0.02	15.52^{*}	0.32^{*}	0.13	0.23
	(0.42)	(0.38)	(0.48)	(0.44)	(0.09)	(0.06)	(0.23)	(0.13)
No. of Days	2.42^{*}	0.03	0.02	0.04^{*}	-3.35	-0.03	-0.02	-0.03
	(0.10)	(0.15)	(0.12)	(0.07)	(0.10)	(0.26)	(0.33)	(0.25)
Sample Size	50	50	50	50	30	30	30	30
Adj. R^2	8.43%	9.61%	4.30%	8.85%	28.71%	25.12%	0.90%	16.75%
HE_GDP	-6.08	-0.15^{**}	-0.10^{**}	-0.11^{*}	-19.65^{***}	-0.34^{***}	-0.20^{**}	-0.31^{***}
	(0.13)	(0.02)	(0.05)	(0.06)	(0.00)	(0.00)	(0.02)	(0.00)
ln(Population)	1.36	0.01	0.02	0.02	19.45^{*}	0.38^{**}	0.15	0.28^{*}
	(0.45)	(0.46)	(0.43)	(0.46)	(0.05)	(0.03)	(0.19)	(0.08)
No. of Days	2.58^{*}	0.02	0.02	0.03^{*}	-4.09^{*}	-0.05	-0.03	-0.05
	(0.09)	(0.18)	(0.19)	(0.09)	(0.07)	(0.15)	(0.21)	(0.15)
Sample Size	49	49	49	49	29	29	29	29
Adj. R^2	7.15%	12.07%	7.79%	10.38%	31.47%	34.29%	11.57%	25.88%
UHC	-301.29^{**}	-4.29^{**}	-2.21	-3.37^{*}	-639.03^{***}	-9.79^{***}	-4.42^{**}	-8.42^{***}
	(0.03)	(0.02)	(0.11)	(0.06)	(0.00)	(0.00)	(0.04)	(0.00)
ln(Population)	-0.44	-0.01	0.01	0.00	17.61^{**}	0.35^{**}	0.14	0.26^{*}
	(0.48)	(0.46)	(0.47)	(0.49)	(0.04)	(0.03)	(0.20)	(0.09)
No. of Days	2.61^{*}	0.03^{*}	0.03^{*}	0.04^{**}	-2.15	-0.01	-0.01	-0.02
	(0.07)	(0.09)	(0.09)	(0.05)	(0.17)	(0.37)	(0.40)	(0.36)
Sample Size	50	50	50	50	30	30	30	30
Adj. R^2	11.70%	11.69%	5.67%	10.86%	45.25%	38.17%	5.52%	26.45%

The table presents the results using regional data from 50 selected countries. We estimate equation 6 using OLS for first-digit goodness-of-fit measures. Panel A shows the results for the cumulative number of confirmed cases, while panel B shows the results for the cumulative number of deaths. To avoid small coefficients, we divide EIU, UHC, and No. of Days values by 100 for all models. Sample sizes vary due to missing values. *P*-values for a one-tailed *t*-test are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. All variable definitions are in Table 6.

Table 4: Second Digit Tests

Variable	Panel A. Confirmed Cases: Chi-squared	Kuiper	M	D	Panel B. Death Cases: Chi-squared	Kuiper	M	D
EIU	−6.72	−0.30*	−0.20	−0.36**	−17.49*	−0.65**	−0.53**	−0.64***
	(0.22)	(0.09)	(0.17)	(0.05)	(0.06)	(0.01)	(0.03)	(0.01)
ln(Population)	−2.39**	−0.06**	−0.05*	−0.06**	−2.15	−0.08**	−0.06*	−0.07**
	(0.04)	(0.04)	(0.05)	(0.03)	(0.12)	(0.04)	(0.08)	(0.05)
No. of Days	−1.66	−0.27*	−0.18	−0.21	−6.39	−0.52**	−0.31	−0.31
	(0.42)	(0.09)	(0.17)	(0.13)	(0.27)	(0.02)	(0.11)	(0.12)
Sample Size	159	159	159	159	113	113	113	113
Adj. R^2	1.02%	3.79%	2.09%	4.29%	1.59%	10.24%	4.73%	6.88%
ln(GDP)	−2.41**	−0.08***	−0.07**	−0.09***	−2.48**	−0.11***	−0.09***	−0.10***
	(0.03)	(0.01)	(0.01)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
ln(Population)	−1.83**	−0.06**	−0.05**	−0.05**	−2.50**	−0.10***	−0.08***	−0.09***
	(0.05)	(0.02)	(0.03)	(0.03)	(0.02)	(0.00)	(0.00)	(0.00)
No. of Days	0.13	−0.23	−0.13	−0.14	−9.65*	−0.60***	−0.40**	−0.37**
	(0.49)	(0.12)	(0.22)	(0.23)	(0.10)	(0.00)	(0.03)	(0.03)
Sample Size	170	170	170	170	116	116	116	116
Adj. R^2	1.74%	5.82%	4.08%	5.51%	7.48%	20.22%	13.76%	15.66%
HE_GDP	−0.21	−0.02	−0.01	−0.01	−2.14***	−0.07***	−0.06***	−0.07***
	(0.39)	(0.10)	(0.22)	(0.26)	(0.01)	(0.00)	(0.00)	(0.00)
ln(Population)	−1.34	−0.04*	−0.03*	−0.03	−1.98	−0.08**	−0.07**	−0.07**
	(0.11)	(0.08)	(0.10)	(0.11)	(0.10)	(0.02)	(0.04)	(0.03)
No. of Days	−1.46	−0.29*	−0.19	−0.21	−8.13	−0.59**	−0.38*	−0.35*
	(0.42)	(0.07)	(0.14)	(0.13)	(0.21)	(0.01)	(0.07)	(0.08)
Sample Size	167	167	167	167	115	115	115	115
Adj. R^2	−0.30%	3.97%	1.94%	1.81%	5.19%	15.88%	9.16%	11.63%
UHC	−13.32	−0.57**	−0.45**	−0.66***	−5.42	−0.57*	−0.41	−0.51*
	(0.12)	(0.03)	(0.05)	(0.01)	(0.35)	(0.06)	(0.13)	(0.08)
ln(Population)	−1.93**	−0.06**	−0.05**	−0.05**	−2.47*	−0.10**	−0.09**	−0.09**
	(0.05)	(0.03)	(0.04)	(0.03)	(0.07)	(0.01)	(0.02)	(0.02)
No. of Days	−0.24	−0.22	−0.15	−0.15	−3.98	−0.45**	−0.25	−0.22
	(0.49)	(0.13)	(0.20)	(0.20)	(0.35)	(0.04)	(0.16)	(0.20)
Sample Size	169	169	169	169	117	117	117	117
Adj. R^2	1.01%	4.90%	3.37%	5.35%	0.38%	10.32%	5.29%	5.89%

The table presents the results of equation 6 using OLS for second-digit goodness-of-fit measures. Panel A shows the results for the cumulative number of confirmed cases, while panel B shows the results for the cumulative number of deaths. To avoid small coefficients, we divide EIU, UHC, and No. of Days values by 100 for all models. Sample sizes vary due to missing values. All models are estimated using OLS regression. P-values for a one-tailed t-test are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. All variable definitions are in Table 6.

Table 5: Swine Flu Pandemic 2009-2010

Variable	Panel A. Confirmed Cases				Panel B. Death Cases			
	Chi-squared	**Kuiper**	**M**	**D**	**Chi-squared**	**Kuiper**	**M**	**D**
EIU	-0.21^{**}	0.00	-0.01	0.00	-0.25^{**}	-0.01	0.00	-0.01
	(0.03)	(0.43)	(0.11)	(0.46)	(0.04)	(0.10)	(0.35)	(0.17)
ln(Population)	-2.55^{*}	0.05	-0.08	0.02	-2.89^{**}	-0.23^{***}	-0.13^{**}	-0.13^{**}
	(0.07)	(0.35)	(0.18)	(0.37)	(0.02)	(0.00)	(0.02)	(0.02)
No. of Days	3.53^{***}	-0.02	0.11^{*}	-0.02	1.01^{***}	0.05^{***}	0.03^{***}	0.03^{***}
	(0.01)	(0.44)	(0.07)	(0.32)	(0.00)	(0.00)	(0.01)	(0.00)
Sample Size	26	26	26	26	23	23	23	23
Adj. R^2	21.51%	-11.47%	1.32%	-11.01%	47.95%	45.96%	19.61%	33.05%
ln(GDP)	-2.18^{**}	-0.05	-0.03	-0.02	-1.97	-0.16^{**}	-0.06	-0.12^{*}
	(0.02)	(0.25)	(0.32)	(0.32)	(0.14)	(0.03)	(0.19)	(0.05)
ln(Population)	-1.19	0.01	-0.04	-0.01	-1.56^{*}	-0.16^{***}	-0.08^{**}	-0.07^{**}
	(0.13)	(0.44)	(0.27)	(0.39)	(0.08)	(0.00)	(0.03)	(0.04)
No. of Days	2.47^{***}	-0.01	0.04	-0.02	0.93^{***}	0.05^{***}	0.03^{***}	0.03^{***}
	(0.01)	(0.45)	(0.27)	(0.34)	(0.00)	(0.00)	(0.00)	(0.00)
Sample Size	35	35	35	35	26	26	26	26
Adj. R^2	35.78%	-7.44%	-8.02%	3.65%	41.60%	43.23%	16.95%	34.09%
HE_GDP	-0.56	-0.05^{*}	-0.04^{*}	-0.02	0.04	-0.05^{*}	0.01	-0.01
	(0.11)	(0.08)	(0.09)	(0.20)	(0.48)	(0.09)	(0.38)	(0.41)
ln(Population)	-0.25	0.04	-0.02	0.00	-1.33	-0.13^{***}	-0.08^{**}	-0.06^{*}
	(0.40)	(0.28)	(0.37)	(0.49)	(0.12)	(0.01)	(0.04)	(0.10)
No. of Days	1.90^{**}	-0.01	0.04	-0.02	0.86^{***}	0.05^{***}	0.02^{**}	0.03^{***}
	(0.03)	(0.42)	(0.25)	(0.31)	(0.00)	(0.00)	(0.01)	(0.00)
Sample Size	35	35	35	35	26	26	26	26
Adj. R^2	29.82%	-2.05%	-2.79%	5.13%	38.29%	37.76%	14.38%	25.50%
UHC	-0.09	-0.01^{*}	0.00	0.00	-0.04	-0.03^{***}	-0.02^{**}	-0.02^{**}
	(0.27)	(0.10)	(0.30)	(0.24)	(0.45)	(0.01)	(0.03)	(0.04)
ln(Population)	-0.78	-0.02	-0.07	-0.03	-1.93^{*}	-0.20^{***}	-0.11^{**}	-0.10^{**}
	(0.26)	(0.40)	(0.18)	(0.26)	(0.07)	(0.00)	(0.01)	(0.02)
No. of Days	2.25^{**}	0.03	0.07	0.00	0.93^{***}	0.06^{***}	0.03^{***}	0.04^{***}
	(0.03)	(0.36)	(0.17)	(0.50)	(0.00)	(0.00)	(0.00)	(0.00)
Sample Size	33	33	33	33	25	25	25	25
Adj. R^2	28.35%	-3.93%	-6.77%	4.94%	38.94%	52.47%	29.49%	37.89%

The table presents the results of 2009-2010 Swine Flu Pandemic analysis for 35 PAHO countries. We estimate equation 6 using OLS for first-digit goodness-of-fit measures. Panel A shows the results for the cumulative number of confirmed cases, while panel B shows the results for the cumulative number of deaths. To avoid small coefficients, we divide EIU, UHC, and No. of Days values by 100 for all models. Sample sizes vary due to missing values. P-values for a one-tailed t-test are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. All variable definitions are in Table 6.

Table 6: Variable Definitions

Variable Names	**Definitions**
Chi-sq. Conf.	Chi-squared statistic based on the cumulative number of confirmed cases, calculated as in formula 1.
Kuipier Conf.	Kuiper statistic based on the cumulative number of confirmed cases, calculated as in formulae 2 and 3.
M Conf.	M-statistic based on the cumulative number of confirmed cases, calculated as in formula 4.
D Conf.	D-statistic on the cumulative number of confirmed cases, calculated as in formula 5.
Chi-sq. Death	Chi-squared statistic based on the cumulative number of deaths, calculated as in formula 1.
Kuiper Death	Kuiper statistics based on the cumulative number of deaths, calculated as in formulae 2 and 3.
M Death	L-statistics based on the cumulative number of deaths, calculated as in formula 4.
D Death	D-statistics based on the cumulative number of deaths, calculated as in formula 5.
EIU	*The Economist Intelligence Unit* Democracy Index. It is presented on a scale of 0 to 100. The index consists of five components.
GDP per capita	Gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars.
HE_GDP	Level of current health expenditure (% of GDP). Level of current health expenditure expressed as a percentage of GDP. Estimates of current health expenditures include healthcare goods and services consumed during each year. This indicator does not include capital health expenditures such as buildings, machinery, IT and stocks of vaccines for emergency or outbreaks.
UHC	Coverage index for essential health services. UHC is the coverage index for essential health services (based on tracer interventions that include reproductive, maternal, newborn and child health, infectious diseases, noncommunicable diseases and service capacity and access). It is presented on a scale of 0 to 100.
No of Days Conf.	Number of non-zero days of daily new confirmed cases.
No of Days Death.	Number of non-zero days for daily new deaths.
Population	Population of a country.
Cutoff value	The earliest date with the maximum 7-day moving average number of new confirmed cases for the country.

The first eight variables are goodness-of-fit measures: Chi-sq. Conf., Kuipier Conf., M Conf., D Conf., Chi-sq. Death, Kuiper Death, M Death and D Death. They are calculated with 3 cutoff points: using the cutoff for the growth part, using 80 days since January 22, 2020, and using 45 calendar days since the first nonzero case for individual countries.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- NBL800020210621supplementarymaterials.pdf