FACTA UNIVERSITATIS
Series: **Economics and Organization** Vol. 18, Nº 2, 2021, pp. 135 - 156
https://doi.org/10.22190/FUEO210326008J

Original Scientific Paper

ASSESSING THE QUALITY OF COVID-19 DATA: EVIDENCE FROM NEWCOMB-BENFORD LAW

UDC 616.98:578.834
519.213

Hrvoje Jošić, Berislav Žmuk

University of Zagreb, Faculty of Economics and Business, Croatia

Abstract. *The COVID-19 infection started in Wuhan, China, spreading all over the world, creating global healthcare and economic crisis.* ***Countries all over the world are fighting hard against this pandemic; however, there are doubts on the reported number of cases. In this paper Newcomb-Benford Law is used for the detection of possible false number of reported COVID-19 cases. The analysis, when all countries have been observed together, showed that there is a doubt that countries potentially falsify their data of new COVID-19 cases of infection intentionally. When the analysis was lowered on the individual country level, it was shown that most countries do not diminish their numbers of new COVID-19 cases deliberately. It was found that distributions of COVID-19 data for 15% to 19% of countries for the first digit analysis and 30% to 39% of countries for the last digit analysis do not conform with the Newcomb-Benford Law distribution. Further investigation should be made in this field in order to validate the results of this research. The results obtained from this paper can be important for economic and health policy makers in order to guide COVID-19 surveillance and implement public health policy measures.***

Key words: *COVID-19, misreporting, Newcomb-Benford Law, Kolmogorov-Smirnov Z test, chi-square test*

JEL Classification: C12, I10

1. INTRODUCTION

The COVID-19 has been initially identified in Wuhan, China, spreading all over the world causing global healthcare and economic crisis. There has been a slowdown in all economic

Received March 26, 2021 / Revised April 15, 2021 / Accepted May 14, 2021
Corresponding author: Hrvoje Jošić
University of Zagreb, Faculty of Economics and Business Administration, Trg J. F. Kennedyja 6, HR-10000 Zagreb, Croatia
E-mail: hjosic@efzg.hr

sectors worldwide, namely tourism, oil industry, aviation, financial and healthcare sector, Shohini (2020). The spread of the virus benefited from the underlying interconnectedness due to globalization, catapulting a global health crisis into a global economic shock, hitting the most vulnerable the hardest, United Nations (2020:6). World Health Organization declared the outbreak of the COVID-19 infection to be a public health emergency of international concern, Zhang (2020). Countries reported their first cases of infection transparently; however, there were doubts about the reported number of cases. There also appears to be a doubt regarding the reported number of cases in the early stages of the epidemic. There were ongoing concerns about the level of transparency around the data from China. The manipulation of pandemic numbers by underreporting for the interest of politics risks lives, Cambell and Gunia (2020). Accurate pandemic numbers are essential for shaping an ongoing response and in making informed decisions on easing restrictions. Reporting accurate numbers is hard because many countries have struggled with adequate testing, which skews the official numbers of those infected, Alwine and Goodrum Sterling (2020). The politics continue to obfuscate the inconvenient truths about the true numbers of COVID-19 cases and deaths. This was encouraged in order to create a false sense of security but the COVID-19 data must be collected and released independently of politics, Alwine and Goodrum Sterling (2020).

In this paper the interaction between new daily cases of COVID-19 disease and the conformance with the Newcomb-Benford Law (NBL) or Benford's Law, Newcomb (1881) and Benford (1938) was investigated. The aim of the analysis is primarily not to report whether a particular country misreports or manipulates the COVID-19 data. The purpose is to assess the quality of COVID-19 data by using Newcomb-Benford Law as a tool. Newcomb-Benford Law is a natural occurrence of digits which are not uniformly distributed. The property of the Newcomb-Benford Law is that the fraudulent or misreported data deviate significantly from the NBL distribution, Balashov et al. (2020). The analysis was made for the early stages of the epidemic for which the numbers of new cases rose exponentially and the Benford's Law should hold, Kennedy and Yam (2020). Benford's Law (BL), or Newcomb-Benford Law (NBL), has many applications in economics. The most important one is as a forensic accounting tool in auditing and fraud detection, Nigrini (2012). This paper follows on previous investigation in this field (Balashov et al. (2020), Kennedy and Yam (2020), Kilani and Georgiou (2020), Zhang (2020)) by analysing the conformance of COVID-19 data with the NBL distribution. In order to detect possible misreported numbers of infection, the distribution of first and last digits of the new cases of COVID-19 infection for 206 countries and self-government territories worldwide will be analysed. The compliance with the Newcomb-Benford Law will be inspected by using chi-square and Kolmogorov-Smirnov Z tests. The expected result is that the distribution of first digits of new COVID-19 cases of infection would follow the Newcomb-Benford Law distribution, meaning that countries do not falsify or diminish their COVID-19 data intentionally. It is also expected that the distribution of last digits in new cases of infection would follow the uniform distribution or equal probability of occurrence. Main contribution of this paper is comprehensive analysis of conformity between new cases of infection and NBL distribution for almost all countries and self-government dependencies in the world in the beginning period of the COVID-19 epidemic, from December 31st, 2019 to April 23rd, 2020.

Paper is organised in six sections. After the introduction, in literature review the history of Newcomb-Benford Law is explained with main applications in the field of economics and epidemiology. In the methodology and data section the Newcomb-Benford Law distribution is derived, the methodology for conducting the chi-square and Kolmogorov-

Smirnov Z tests explained as well as descriptive statistics of data. In the results and discussion section the main results of the analysis are displayed, both for the first and last digits of COVID-19 cases by using Newcomb-Benford Law and uniform distribution as a tool. Final chapter presents concluding remarks.

2. LITERATURE REVIEW

Benford's Law or Newcomb-Benford Law is a natural observation in many occurring selections of numbers for which the first digit is not uniformly distributed. The history of Newcomb-Benford Law originates in 1881 when Simon Newcomb (Newcomb, 1881) noticed that the first pages of logarithmic tables were more worn out than the rest. That implies there are more digits starting with the digit one than that is expected under the uniform distribution. Newcomb described this phenomenon in his paper "*Note on the Frequency of Use of the Different Digits in Natural Numbers*". Unaware of Newcomb's findings, Frank Benford came to the similar conclusion almost 60 years later in his paper "*The law of anomalous numbers*", Benford (1938). Therefore, Newcomb-Benford Law was named according to both deserving economists. Newcomb-Benford Law has applications in various fields of economics but the most important one is as a tool for forensic accounting and fraud detection, Nigrini (1996). Other applications of Newcomb-Benford Law are for campaign fraud detection, Cho and Gaines (2007), governmental statistics inspection, Hindls and Hronová (2015), fraudulent scientific data, Diekmann (2007) and for inspection whether countries falsify their economic data strategically, Michalski and Stoltz (2013).

Jošić and Žmuk (2018) used Benford's Law for psychological pricing detection. Seminal paper in this field was published by El Sehity et al. (2005) which analyses consumer price digits before and after the euro introduction. Another piece of empirical evidence on psychological pricing was related to Austrian retailers, Wagner and Jamsawang (2012). Zhang (2020) proposed a test for checking the reported number of COVID-19 cases in China using the Newcomb-Benford Law. The obtained p-value of 92.8% indicated that the distribution of COVID-19 cumulative cases abide by the Newcomb-Benford Law. The author stated that the reported number of cases could be lower than the real number of infected people due to the lack of medical equipment and resources. Balashov et al. (2020) used Newcomb-Benford Law to test whether countries manipulate their COVID-19 data during the pandemics. The most important finding of the paper was that democratic countries with higher values of gross domestic product per capita, higher healthcare expenditures and universal healthcare coverage are the ones less likely to deviate from the Newcomb-Benford Law. It was found that roughly one third out of the 185 countries in the world affected by the pandemics seem to misreport their data. Kennedy and Yam (2020) studied the applicability of Benford's Law to national COVID-19 case figures. The aim was to establish guidelines for methods of fraud detection in epidemiology. Benford's Law largely held across countries in the early stages of the epidemic for which the number of infected people is relatively small in regards to the population. This argument also held for the second digit analysis.

Kilani and Georgiou (2020) collected a database of potential data misreports by 171 countries regarding their COVID-19 daily reported cases. They employ three different tests (chi-square, Kuiper and Mean Absolute Deviation (MAD)) in order to determine if data for each observed country fit the Benford's Law. For most of the countries the results showed the conformity of COVID-19 data with the Benford's Law. Koch and

Okamura (2020) emphasized the importance of veracity of reported contagious diseases data in real time. The authors found that the Chinese, United States and Italian data matched the distribution expected by Benford's Law. If the numbers were taken from the exponential distribution, it could be demonstrated that they automatically follow the Benford's Law distribution, Lee et al. (2020). The number of cases of infections and/or deaths will not obey the Benford's Law if the current control interventions are successful in flattening the epidemic curve. It is the case when the epidemic growth rate is below the exponential growth rate. Investigating whether countries misreport or diminish their numbers of COVID-19 cases in the early stages of infection can be therefore considered as valid.

Moreno-Montoya (2020) propose a new test in evaluating compliance with the Benford's Law distribution in the case of small data samples because conventional statistical methods for evaluation of small data samples are controversial. According to Peng and Nagata (2020), China's empirical distribution of new cases of infection appears to be particularly different from other countries. Despite being the first country affected by the disease, there was a linear trend present in the early stages of infection. Silva and Figueiredo Filho (2020) employed Newcomb–Benford Law to evaluate the reliability of COVID-19 figures in Brazil in the period from February 25th to September 15th . They found strong evidence that Brazilian reports do not conform with the Newcomb-Benford Law theoretical predictions showing that the Brazilian epidemiological surveillance system failed to provide trustful data on the COVID-19 epidemic.

3. DATA AND METHODOLOGY

Newcomb-Benford Law (NBL) is empirical wellknown pattern for frequency of first digit occurrence in various datasets. The first digit is not uniformly distributed: the number one appears as a leading digit in 30.1% of cases, the number two appears as a leading digit in 17.6% of cases while the number nine occurs as the first digit in 4.5% of the time. Checking for conformance with the NBL would be the best approach in a forensic analysis looking at potential manipulations of the number of cases since the distribution of first digits that deviates from the expected distribution may indicate frauds, Lee et al. (2020:4). In this paper it is analysed whether distribution of new cases of COVID-19 disease conform with the Newcomb-Benford Law distribution for the first leading digit and whether distribution of new cases of the COVID-19 conform with the uniform distribution for the last digit. A reasonable assumption will be that COVID-19 new case numbers should follow the Newcomb-Benford Law distribution. It seems the infection grows exponentially, particularly at the beginning in the early stage, Zhang (2020). It is hard to fabricate data closely following the Newcomb-Benford Law distribution. That implies if the distribution of first digits for new daily cases of COVID-19 follows the NBL distribution then there is no misreporting or possible diminishing of the number of new daily cases. Also, it is expected that the distribution of last digits of new daily cases would follow the uniform distribution, meaning the same frequency of number occurrence, leading again to the conclusion that there are no frauds or misreports of data detected.

The probabilities of first digit occurrence in the Newcomb-Benford Law are derived using the following Equation 1:

$$p(d) = \log_{10}\left(1 + \frac{1}{d}\right), \text{ where } d \in \{1,2,3,...,8,9\}. \quad (1)$$

The probabilities for the second digit occurrence in the NBL are derived from the Equation 2:

$$P(d) = \sum_{k=1}^{9} \log_{10}\left(1 + \frac{1}{10k + d}\right), \text{ where } d = 0,1,2,...,9. \quad (2)$$

In Equation 3 the probabilities of occurrence for the higher-order digits up to the last digit with equal probability of 0.1 which is identical to uniform distribution are derived.

$$P(d_k) = \sum_{d_1=1}^{9} \sum_{d_2=0}^{9} ... \sum_{d_{k-1}=0}^{9} \log_{10}\left(1 + \frac{1}{\sum_{i=1}^{k} d_i \cdot 10^{k-i}}\right), \text{where } d_k = 0,1,2,...,9. \quad (3)$$

The calculated probabilities of occurrence for the first digit, second digit, higher-order and the last digit are presented in Table 1.

Table 1 Expected frequencies of digit occurrence in NBL distribution

Number	1st digit	2nd digit	3rd digit	4th digit	5th digit
0	-	0.11968	0.10178	0.10018	0.10
1	0.30103	0.11389	0.10138	0.10014	0.10
2	0.17609	0.10882	0.10097	0.10010	0.10
3	0.12494	0.10433	0.10057	0.10006	0.10
4	0.09691	0.10031	0.10018	0.10002	0.10
5	0.07918	0.09668	0.09979	0.09998	0.10
6	0.06695	0.09337	0.09940	0.09994	0.10
7	0.05799	0.09035	0.09902	0.09990	0.10
8	0.05115	0.08757	0.09864	0.09986	0.10
9	0.04576	0.08500	0.09827	0.09982	0.10

Source: Nigrini (1996), Jošić and Žmuk (2018)

Epidemics such as COVID-19, which we are experiencing at the moment, are classic examples of exponential growth function. The number of infected people tomorrow, I_1 , is equal to constant α times the amount of infected people today, I_0, that is $I_1 = \alpha \cdot I_0$. This expression could be generalized for t days as $I_1 = \alpha^t \cdot I_0$. This exponential growth could obey Newcomb-Benford Law, Peng and Nagata (2020). Kennedy and Yam (2020) provided a justification for the emergence of Benford's Law during the early stages of epidemic. Let $S(t)$ denote the number of susceptible individuals. In the early stages of the epidemic the upper constraint of population size is negligible. Under the assumptions of fixed infectiousness $\theta > 0$, fixed recovery rate $\delta > 0$ and $\delta < \theta$, the evolution of $I(t)$ can be described by:

$$I(t + 1) = I(t) + (\theta + \varepsilon^{I}_{t+1})I(t) - (\delta + \varepsilon^{R}_{t+1})I(t) \quad (4)$$

for $t = 1,...,T-1$, ε_t^I are independent and identically distributed (i.i.d.) random noise terms, as are ε_t^R. The evolution of $S(t)$ is analogously defined as:

$$S(t+1) = S(t) - (\theta + \varepsilon_{t+1}^I)I(t) + (\delta + \varepsilon_{t+1}^R)I(t) \tag{5}$$

The epidemic growth of $I(t)$ can be further expressed as:

$$I(t+1) = A_{t+1} \times A_t \times \ldots \times A_1 \tag{6}$$

where

$$A_t \triangleq 1 + \theta - \delta + \varepsilon_t^I - \varepsilon_t^R \tag{7}$$

The Equation 6 suggests that Newcomb-Benford Law should emerge naturally during the early stages of an epidemic.

Data about new cases of COVID-19 infection are taken from the EU Open Data Portal database, EU Open Data Portal (2020). The number of new cases is observed from the start of the infection, from December 31st, 2019 up to April 23rd, 2020. The days in which there were no new cases of COVID-19 infection were omitted from the analysis. The data for overall 206 countries and self-government dependencies in the world were collected. Firstly, the analysis will be conducted by taking into account all observed countries together. After that the analysis will be conducted for each country separately. In order to inspect whether the distributions of the first and the last digits follow NBL or uniform distribution, chi-square and Kolmogorov-Smirnov Z tests will be used. The chi-square test values will be calculated by using the following Equation 8:

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i} \tag{8}$$

where f_i are the actual values for the i-th first digit or the i-th last digit, e_i are actual values of the i-the first digit or the i-th last digit under the assumption that the distribution of the first digits is distributed according to the NBL distribution and the distribution of the last digits is distributed according to the uniform distribution. Similarly, the values for Kolmogorov-Smirnov Z test will be calculated as follows:

$$K - S = \frac{\sqrt{-\frac{1}{2} ln\left(\frac{\alpha}{2}\right)}}{\sqrt{n}} \tag{9}$$

where α is statistical significance level (here 0.05) and n is the total number of new daily values. For both statistical tests the null hypothesis contains an assumption that the observed daily new cases of COVID-19 will follow the certain distribution (here NBL or uniform distribution). On the other hand, the alternative hypothesis assumes that the observed data will not follow the certain data distribution. Before conducting the chi-square and Kolmogorov-Smirnov Z tests, basic descriptive statistics analysis was done. In Table 2 basic descriptive statistics results for the new cases of COVID-19 infection, first digit and the last digit of the new cases by taking into account all countries together are presented.

Table 2 Descriptive statistics for the new cases, first and last digit, all countries together, daily values from December 31st, 2019 to April 23rd, 2020

Statistics	New cases	First digit	Last digit
Sample size	6,787	6,787	6,787
Mean	381.36	3.17	3.95
Standard deviation	1,998.73	2.34	2.76
Coeff. of variation	524%	74%	70%
Skewness	12	0.98	0.35
Kurtosis	176	-0.09	-1.11
Mode	1	1	1
Minimum	1	1	0
1st quartile	4	1	1
Median	19	2	4
3rd quartile	106	5	6
Maximum	37,289	9	9
Range	37,288	8	9
Interquartile range	102	4	5

Source: EU Open Data Portal (2020), authors.

According to the results from Table 2 there were overall 6,787 daily data about new cases of COVID-19 infection. The total number of days in the observed period was 12,596, but there were 5,809 days without new cases of infection which were excluded from the analysis. On average there were 381 new cases of infection daily with an average deviation of 1,999 new cases or 524%. The very high variability level is obvious if just minimum and maximum values are compared. From the new cases values their first and last digits are taken and basic descriptive statistics analysis is conducted as well. The results are shown in the last two columns and they are quite similar.

4. Results and Discussion

In addition to the numeric analysis for the first digits, their distributions and comparison with the Newcomb-Benford Law distribution are graphically shown in Figure 1. According to the Figure 1, the most common first digit is one. It appeared in 2,279 cases or 33.58% of total cases. On the other hand, the number eight the lowest appearance had; it appeared in 244 cases or 3.60% of total cases. From the graphical analysis it can be seen that the daily distribution of new cases of infection and Newcomb-Benford Law distribution are close to each other indicating that the distribution of the first digits for new cases of COVID-19 infection is conforming with the Newcomb-Benford Law distribution. However, in order to be sure, statistical tests (chi-square and Kolmogorov-Smirnov Z test) are going to be applied. The results of the chi-square and the Kolmogorov-Smirnov Z tests for the first digit of new cases of COVID-19 infection on the overall sample of countries are presented in Tables 2 and 3.

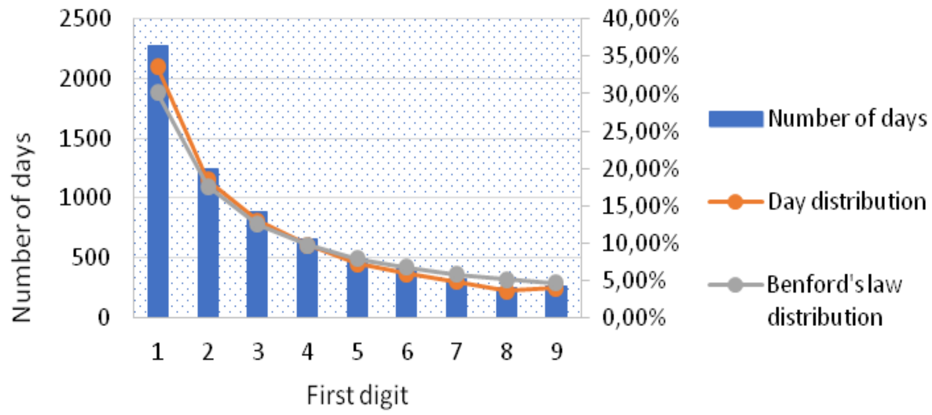

Fig. 1 Distribution of first digits of the new cases and comparison with the NBL distribution

It is examined whether distribution of first digits in the sample follows the distribution defined by the Newcomb-Benford Law. The hypotheses are as follows:

H_0... *The distribution of first digits for the number of new cases of COVID-19 follows the distribution defined by the Newcomb-Benford Law.*

H_1... *The distribution of first digits for the number of new cases of COVID-19 does not follow the distribution defined by the Newcomb-Benford Law.*

Table 3 Chi-square test for the first digit of new cases of COVID-19 infection

First digit	Number of days	Percentage of days	Benford rate	f_i	e_i	$(f_i-e_i)^2/e_i$
1	2,279	33.58%	30.10%	2,279	2,043	27
2	1,250	18.42%	17.61%	1,250	1,195	3
3	882	13.00%	12.49%	882	848	1
4	657	9.68%	9.69%	657	658	0
5	486	7.16%	7.92%	486	537	5
6	400	5.89%	6.69%	400	454	7
7	324	4.77%	5.80%	324	394	12
8	244	3.60%	5.12%	244	347	31
9	265	3.90%	4.58%	265	311	7
Total obs.	6,787	100.00%	100.00%	6,787	6,787	92

Source: Authors' calculations.

According to the chi-square test results presented in Table 3 (empirical chi-square value equal to 92.196, theoretical chi-square of 15.51 (α=0,05), p-value < 0.0001 and with 8 degrees of freedom) the null hypothesis of the chi-square test can be rejected at any commonly used statistical significance level. It can be concluded that the first digit distribution of new cases, when all countries are observed together, is not following the Newcomb-Benford Law distribution, meaning that countries are possibly misreporting the number of new COVID-19 cases of infection. Comparison of the first digit cumulative density distribution of COVID-19 new cases and the cumulative density of Newcomb-Benford Law distribution is presented in Figure A1 in Appendix.

Table 4 Kolmogorov-Smirnov Z test for the first digit of new cases of COVID-19 infection

First digit	Number of days	Percentage of days	Benford rate	Cum. density new cases distribution	Cum. density Benford's law distribution	Absolute difference
1	2,279	33.58%	30.10%	0.3358	0.3010	0.0348
2	1,250	18.42%	17.61%	0.5200	0.4771	0.0428
3	882	13.00%	12.49%	0.6499	0.6021	0.0479
4	657	9.68%	9.69%	0.7467	0.6990	0.0478
5	486	7.16%	7.92%	0.8183	0.7782	0.0402
6	400	5.89%	6.69%	0.8773	0.8451	0.0322
7	324	4.77%	5.80%	0.9250	0.9031	0.0219
8	244	3.60%	5.12%	0.9610	0.9542	0.0067
9	265	3.90%	4.58%	1.0000	1.0000	0.0000
Total	6,787	100.00%	100.00%	-	-	-

Source: Authors' calculations.

Again, at first, it could be said that the first digit distribution of new cases follows the Newcomb Benford Law distribution. However, the conducted Kolmogorov-Smirnov Z test (empirical test value equal to 0.0479, theoretical K-S value of 0.0015) indicates that the null hypothesis can be rejected at any commonly used statistically significant level. So, the conclusion is that the first digit distribution of new cases does not follow the Newcomb-Benford Law distribution. In Figure 2 distribution of last digits of the new cases of COVID-19 and comparison with the uniform distribution is presented.

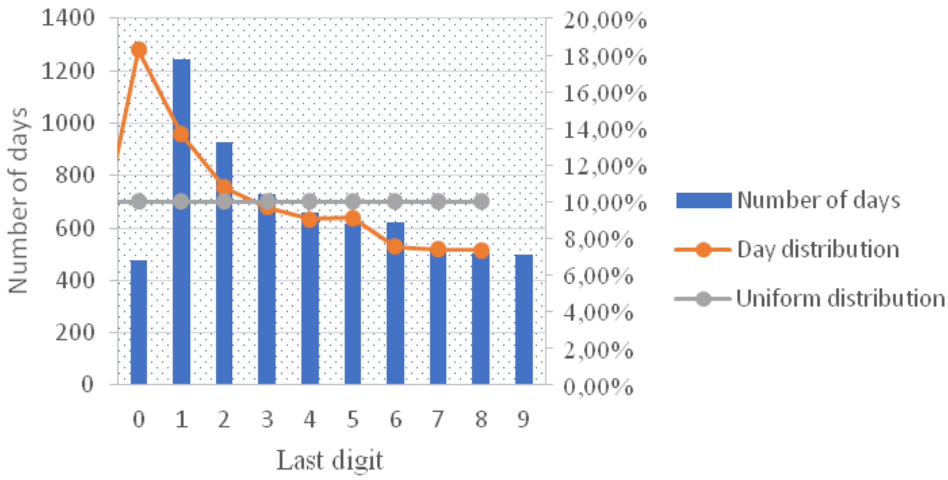

Fig. 2 Distribution of last digits of the cases and comparison with the uniform distribution

According to the Figure 2 the most common last digit is one (1,242 cases or 18.30% of total cases) and the least common last digit is zero (478 cases or 7.04% of total cases).

From the graphical representation it is obvious that the last digit distribution of new cases does not follow the uniform distribution. In the following hypotheses it is examined whether distribution of the last digits in the sample conforms with the uniform distribution.

H_0... The distribution of the last digits for the new cases of COVID-19 infection follows the uniform distribution.

H_1... The distribution of the last digits for the new cases of COVID-19 infection does not follow the uniform distribution.

Table 5 Chi-square test for the last digit of new cases of COVID-19 infection

Last digit	Number of days	Percentage of units	Uniform distribution	f_i	e_i	$(f_i-e_i)^2/e_i$
0	478	7.04%	10.00%	478	679	59
1	1,242	18.30%	10.00%	1,242	679	468
2	928	13.67%	10.00%	928	679	92
3	731	10.77%	10.00%	731	679	4
4	658	9.70%	10.00%	658	679	1
5	614	9.05%	10.00%	614	679	6
6	620	9.14%	10.00%	620	679	5
7	513	7.56%	10.00%	513	679	40
8	502	7.40%	10.00%	502	679	46
9	501	7.38%	10.00%	501	679	47
Total	6,787	100.00%	100.00%	6,787	6,787	767

Source: Authors' calculations.

The conducted chi-square test (empirical chi-square value equal to 767.33, theoretical chi square 16.92, p-value < 0.0001 with 9 degrees of freedom) confirmed that the null hypothesis of the test can be rejected at any usually used statistically significance level. In Figure A2 in Appendix the comparison between the last digit cumulative density distribution of new cases and the cumulative density uniform distribution is presented.

Table 6 Kolmogorov-Smirnov Z test for the last digit of new cases of COVID-19 infection

Last digit	Number of days	Percentage of units	Uniform distribution	Cumulative density new cases distribution	Cumulative density uniform distribution
0	478	7.04%	10.00%	0.0704	0.1000
1	1,242	18.30%	10.00%	0.2534	0.2000
2	928	13.67%	10.00%	0.3902	0.3000
3	731	10.77%	10.00%	0.4979	0.4000
4	658	9.70%	10.00%	0.5948	0.5000
5	614	9.05%	10.00%	0.6853	0.6000
6	620	9.14%	10.00%	0.7766	0.7000
7	513	7.56%	10.00%	0.8522	0.8000
8	502	7.40%	10.00%	0.9262	0.9000
9	501	7.38%	10.00%	1.0000	1.0000
Total	6,787	100.00%	100.00%	-	-

Source: Authors' calculations.

The Kolmogorov-Smirnov Z test (empirical test value equal to 0.0979 and theoretical Kolmogorov-Smirnov Z value of 0.0015) led to the same conclusion as the corresponding chi-square test. The conclusion is that the last digit distribution of new cases does not follow the uniform distribution. It can be concluded that when all countries in the world are observed together, there is a potential doubt that countries misreport their data of new cases of infection. The same analysis, as explained here for all countries together, is conducted for each country separately. The aggregated results are shown in Table 7.

Table 7 Summary results for individual countries, 206 countries, data are daily values of new cases in the period from December 31st, 2019 to April 23rd, 2020

Continent	Test conclusion at significance level 0.05	Null hypothesis: the distribution of the first digits of new cases is following the NBL distribution		Null hypothesis: the distribution of the last digits of new cases is following the uniform distribution	
		Chi-square test	Kolmogorov-Smirnov Z test	Chi-square test	Kolmogorov-Smirnov Z test
Overall	Do not reject null hypothesis	167	175	127	146
	Reject null hypothesis	39	31	79	60
Africa	Do not reject null hypothesis	50	47	28	32
	Reject null hypothesis	2	5	24	20
America	Do not reject null hypothesis	40	43	22	34
	Reject null hypothesis	9	6	27	15
Asia	Do not reject null hypothesis	32	34	27	27
	Reject null hypothesis	10	8	15	15
Europe	Do not reject null hypothesis	36	43	46	48
	Reject null hypothesis	18	11	8	6
Oceania	Do not reject null hypothesis	8	7	3	4
	Reject null hypothesis	0	1	5	4
Other	Do not reject null hypothesis	1	1	1	1
	Reject null hypothesis	0	0	0	0

Source: EU Open Data Portal (2020), authors.

When the analysis is lowered on the individual country level, different conclusions could bereached. Detailed results of conducted chi-square and Kolmogorov-Smirnov Z tests for the first and last digit for 206 countries and self-government dependencies are presented in Table A1 in Appendix. The chi-square tests have shown that for 167 countries (out of 206) the distribution of the first digits for new cases follows the Newcomb Benford's distribution meaning that countries do not misreport or diminish data of new cases of COVID-19. The distribution of the last digits of new cases is following the uniform distribution for 127 countries, leading to the similar conclusion. The Kolmogorov-Smirnov Z tests results are going even more in favour of not rejecting the null hypothesis. The difference between the results achieved in the analysis for all countries together and on the individual country level can be explained with heterogeneity in data or unique characteristics of each individual country.

The obtained results are in the line with previous investigation in this field of research, however, there is no general theory that the epidemics like COVID-19 should obey the Newcomb-Benford Law. Balashov et al. (2020) came to the conclusion that roughly one third out of 185 countries misreport their data intentionally, which are results similar to our findings. We found that 39 out of 206 countries for the chi-square test and 31 out of 206 countries for the Kolmogorov-Smirnov Z test for the first digit analysis which is result in the range of 15%-19% of countries, potentially misreport their data. On the other hand, we found that 79 out of 206 countries for the chi-square test and 60 out of 206 countries for the Kolmogorov-Smirnov Z test for the last digit analysis, which is in the range of 29%-38% of countries, potentially misreport their data. Lee at al. (2020) found that 9 out of 10 countries satisfy the Newcomb-Benford Law, indicating that the growth rates of COVID-19 in these 9 countries were close to an exponential trend. Kilani and Georgiou (2020) made ranges of tests (chi-square, Kuiper and MAD) for 171 countries regarding their COVID-19 daily reported cases. The results of chi-square and Kuiper tests mostly confirmed the conformity with the Benford's Law, in 78.4% and 65.50% respectively. On the other hand, the MAD test pointed out to different conclusion; 111 out of 171 countries or 64.91% showed the non-conformity with the Newcomb-Benford Law. The authors devised the conformity ranges with the NBL distributions dividing them into close conformity, acceptable conformity, marginable acceptable conformity and nonconformity.

Kennedy and Yam (2020) found empirical evidence that Benford's Law largely hold across countries while deviations could be easily explained, including constrained testing, poorly defined start dates or government intervention through social distancing measures in slowing down transmission of the disease. Zhang (2020) showed that Newcomb-Benford Law held for the cumulative case numbers of COVID-19 on data for 31 province-level divisions in China in the period from January 15th 2020 to February 10th 2020. There were overall 628 data points in the analysis which was not a big dataset compared to ours (6,787 data points). Miranda (2020) conducted test of frauds by examining the cumulative distribution of the Philipinnian COVID-19 data and the Newcomb-Benford Law distribution by employing the Kolmogorov-Smirnov test in order to analyse the differences between the distributions. The data were used for three months after the first case of COVID-19 in the country, that is in the beginning of the epidemic, similar as in this paper. There was no significant difference between the COVID-19 data's first digit distribution and the distribution set by NBL suggesting no evidence for data manipulation. Wong et al. (2020) focused the study on two Southeast Asian countries: Indonesia and Malaysia during the period between March and November 2020. A chi-square test was recruited to quantify the closeness of the data and Newcomb-Benford Law distribution. Distribution of daily infection and death cases in Indonesia followed the Newcomb-Benford Law while the opposite result was obtained for Malaysia.

Contribution of this paper to the existing theory and knowledge in this field of research is twofold. Firstly, in line with conducting the first digit analysis for the new cases of COVID-19 infection, the analysis was broadened to the last digit analysis using uniform distribution as a reference distribution. Secondly, the dataset included almost all countries in the world with consequent cases of infection at the beginning of the epidemics. According to the Kennedy and Yam (2020) there are some ambiguities in how the timeline of the epidemic should be defined; the beginning of the epidemic should be set on date when sustained community transmission firstly occur, as opposed to the emergence of the first case of infection.

This study has important implications for the government health care systems and overall community. Similar tests can be applied to epidemics other than COVID-19. Countries should report their numbers of COVID-19 cases correctly. However, the motivation for possible data misreporting or diminishing could be to avoid travel bans and decline in tourism.

That could lead to taking the disease not seriously so there is clear need to verify the data throughout rigorous statistical techniques, help detect fraudulent behavior and verify the authenticity of published figures. Without valid data it is almost impossible to correctly evaluate the government intervention measures. It can be concluded that falsifying epidemic data is a short-lived strategy for governments and is not sustainable over the long run, Balashov et al. (2020).

5. CONCLUSIONS

Main findings of the paper can be summarized as: (1) the results of the Kolmogorov-Smirnov Z test and chi-square test, when all countries in the world were observed together, pointed out to the conclusion that the distribution of the first digits of new COVID-19 cases was not following the NBL distribution meaning that countries are potentially misreporting their COVID-19 data, (2) the aforementioned tests confirmed that the distribution of the last digits of new cases did not follow the uniform distribution, (3) when the analysis was lowered on an individual country level, both tests, chi-square and Kolmogorov-Smirnov Z test, pointed out to the conclusion that the distribution of first digits in most cases (167 out of 206 and 175 out of 206) obey the NBL, indicating that most of the countries do not diminish their numbers of new COVID-19 cases deliberately, (4) when the distribution of the last digits of new cases of infection was observed, the similar conclusion could be reached.

It can be concluded that the quality of COVID-19 data in most of countries in the world at the beginning of the epidemic is on the satisfactory level of trust. The divergences from the expected distributions should not be attributed to the deliberate falsification of data from governments but possibly from the low quality or structural breaks in data. In addition, government measures intended to flatten the epidemic curve could influence the results even at the early stages of the epidemic, in its exponential phase of growth. When the main findings of this paper are compared with the previous research it can be said that they are in the line with the state of the art of economic theory. The COVID-19 data show exponential growth at the beginning of the epidemic with distribution conforming with the Newcomb-Benford Law distribution.

Limitations of the research are related to the uneven number of observed days of infection duration for the each observed country, meaning that the reported number of cases is actually lower than the real number of infected people. The data are incomplete due to the lack of medical equipment and resources with many suspected cases remaining to be confirmed. There is also a cyclic component of data reports on weekends, especially Sundays, for which the data of new cases of infection are usually lower due to less testing on these days.

There are several areas of future research that could be built upon this paper such as detailed analysis of individual countries, second and/or higher order digit analysis, observation of cumulative number of cases or the number of reported deaths. The spread of the disease come in waves, so similar analysis could be made for the start of the second and third wave of infection or any other successive wave. The methodology displayed in this paper could be additionally improved in order to include government measures for preventing the disease through limitation of social contacts and lockdowns, in testing the compliance of COVID-19 data distribution with the Newcomb-Benford Law distribution. The results obtained in this paper can be important for economic and health policy makers in order to guide the COVID-19 surveillance by evaluating the effectiveness and performance of COVID-19 control interventions and public health surveillance systems.

Acknowledgement: *This manuscript is an improved version of the paper „Do countries diminish the number of new COVID-19 cases? A test using Benford's Law and uniform distribution" authors Berislav Žmuk and Hrvoje Jošić presented online at the 3rd Conference titled "Economic System of the European Union and Accession of Bosnia and Herzegovina - Challenges and Policies Ahead", Mostar, Bosnia and Herzegovina, 24th October, 2020.*

REFERENCES

Alwine, J., & Goodrum Sterling, F. (2020). *Manipulation of pandemic numbers for politics risks lives*. Available at: https://thehill.com/opinion/healthcare/499535-manipulation-of-pandemic-numbers-for-politics-risks-lives [Accessed: 2021-03-10].

Balashov, V., S., Yan, Y., Zhu, X. (2020). Who Manipulates Data During Pandemics? Evidence from Newcomb-Benford Law, https://doi.org/10.2139/ssrn.3662462. Available at: https://ssrn.com/abstract=3662462 [Accessed: 2021-03-10].

Benford, F. (1938). The law of anomalous numbers, *Proceedings of the American Philosophical Society*, *78*(4), pp. 551-572.

Cambell, C., & Gunia, A. (2020). China says it's beating coronavirus. But can we believe its numbers? Available at: https://time.com/5813628/china-coronavirus-statistics-wuhan/ [Accessed: 2021-03-10].

Cho, W. K. T., & Gaines, B. J. (2007). Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance, *American Statistician*, *61*(3), 218-223.

Diekmann, A. (2007). Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data, *Journal of Applied Statistics*, *34*(3), 321-329.

El Sehity, T. J., Hoelzl, E., & Kirchler, E. (2005). Price Developments after a Nominal Shock: Benford's Law and Psychological Pricing after the Euro Introduction. *International Journal of Research in Marketing (IJRM)*, *4*(22), 471-480.

EU Open Data Portal (2020). COVID-19 Coronavirus data [online]. Available at: data.europa.eu/euodp/en/data/dataset/covid-19-coronavirusdata/resource/55e8f966-d5c8-438e-85bc-c7a5a26f4863 [Accessed: 2021-03-10].

Hindls, R., & Hronová, S. (2015). Benford's Law and Possibilities for Its Use in Governmental Statistics. *Statistika*, *95*(2), 54-64.

Jošić, H., & Žmuk B. (2018). The application of Benford's law in psychological pricing detection. *Zbornik radova Ekonomskog fakulteta Sveučilišta u Mostaru*, *24*, 37-57.

Kennedy, A. P., & Yam, S. C. P., (2020). On the authenticity of COVID-19 case figures, *PLoS ONE*, *15*(12), 1-22.

Kilani, A., & Georgiou, G. P. (2020). The Full Database of Countries with Potential COVID-19 Data Misreport based on Benford's Law. Available at: https://www.medrxiv.org/content/10.1101/2020.12.04.20243832v1 [Accessed: 2021-03-10].

Koch, C., Okamura, K. (2020). Benford's law and COVID-19 reporting. *Economics Letters*, *196* (2020), 109573.

Lee, K.-B., Han, S., & Jeong, Y. (2020). COVID-19, flattening the curve, and Benford's law. *Physica A*, 1-12.

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

Miranda, A. T. (2020). The Distribution of COVID-19 Cases in the Philippines and the Benford's Law. *Philippine e-Journal for Applied Research and Development 10*(2020*)*, 29-34.

Moreno-Montoya, J. (2020). Benford´s Law with small sample sizes: A new exact test useful in health sciences during epidemics. *Salud UIS*, *52*(2), 161-163. http://dx.doi.org/10.18273/revsal.v52n2-2020010

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. *Journal of the American Taxation Association*, *18*(1), 72-91.

Nigrini, M. J. (2012). *Benford's Law: Application for Forencis Accounting. Auditing, and Fraud detection*, 2012, John Wiley and Sons.

Peng, Y., & Nagata, M. H. (2020). Statistical analysis of the Chinese COVID-19 data with Benford's law and clustering. Available at: https://lamfo-unb.github.io/2020/04/21/COVID-China-EN/ [Accessed: 2021-03-10].

Shohini, R. (2020). Economic impact of COVID-19 pandemic. Technical report. Available at: https://www.researchgate.net/publication/343222400_ECONOMIC_IMPACT_OF_COVID-19_PANDEMIC [Accessed: 2021-03-10].

Silva, L., & Figueiredo Filho, D, B. F. (2020). Using the Benford's Law to Assess the Quality of COVID-19 Register Data in Brazil. *Journal of Public Health*, fdaa193. https://doi.org/10.1093/pubmed/fdaa193, http://dx.doi.org/10.17605/OSF.IO/74XJC

United Nations (2020). Impact of the COVID-19 Pandemic on Trade and Development, Transitioning to a new normal, United Nations Conference on Trade and Development. Available at: https://unctad.org/system/files/official-document/osg2020d1_en.pdf [Accessed: 2021-03-10].

Wagner, U., & Jamsawang, J. (2012). Several Aspects of Psychological Pricing: Empirical Evidence from some Austrian Retailers. In: Rudolph T., Foscht T., Morschett D., Schnedlitz P., Schramm-Klein H., Swoboda B. (eds) *European Retail Research*. European Retail Research. Gabler Verlag, Wiesbaden.

Wong, W. K, Juwono, F. H., Loh, W. N., & Ngu, I. Y. (2020). Newcomb-Benford Law Analysis on COVID-19 Daily Infection Cases and Deaths in Indonesia and Malaysia. Available at: https://unctad.org/system/files/official-document/osg2020d1en.pdf [Accessed: 2021-03-10].

Zhang, J. (2020). Testing Case Number of Coronavirus Disease 2019 in China with Newcomb-Benford Law. *Physics and Society*, 1-7.

Žmuk, B., & Jošić, H. (2020) *Do countries diminish the number of new COVID-19 cases? A test using Benford's law and uniform distribution*, 3rd Conference titled "Economic System of the European Union and Accession of Bosnia and Herzegovina - Challenges and Policies Ahead", Mostar, Bosnia and Herzegovina, online, 24 October, 2020.

PROCENA KVALITETA COVID-19 PODATAKA: PRIMENA NEWCOMB-BENFORDOVOG ZAKONA

Infekcija COVID-19 započela je u kineskom grad Wuhanu, šireći se celim svetom stvarajući globalnu zdravstveno-zaštitnu i ekonomsku krizu. Zemlje širomsveta se žestoko bore protiv ove pandemije, međutim, postoje sumnje u prijavljeni broj zaraženih ljudi. U ovom se radu Newcomb-Benfordov zakon koristi za otkrivanje lažnog broja prijavljenih slučajeva COVID-19. Analiza, kada su sve zemlje posmatrane zajedno, je pokazala da postoji potencijalna sumnja da zemlje prijavljuju lažne podatke o novim slučajevima zaraze. Kada je analiza spuštena na nivopojedinih zemalja, pokazala je da većina zemalja ne umanjuje broj novih slučajeva COVID-19 namerno. U analizi prvih cifara je utvrđeno da u 15-19 odsto slučajeva kao i u analizi zadnjihcifara u 30-39 odsto slučajeva da distribucija COVID-19 brojki ne odgovara distribuciji Newcomb-Benfordovog zakona. Međutim, na ovom polju je potrebno činiti daljnja istraživanja kako bi se potvrdili rezultati ovog rada. Rezultati dobijeni u ovom istraživanju mogu biti važni za kreatore ekonomskih i javno-zdravstvenih politika kako bi usmeravali nadzor nad COVID-19 sprovođenjem mera politike javnog zdravstva.

Ključne reči: *COVID-19, pogrešno prijavljivanje, Newcomb-Benfordov zakon, Kolmogorov-Smirnov Z test, hi-kvadrat test*

APPENDIX

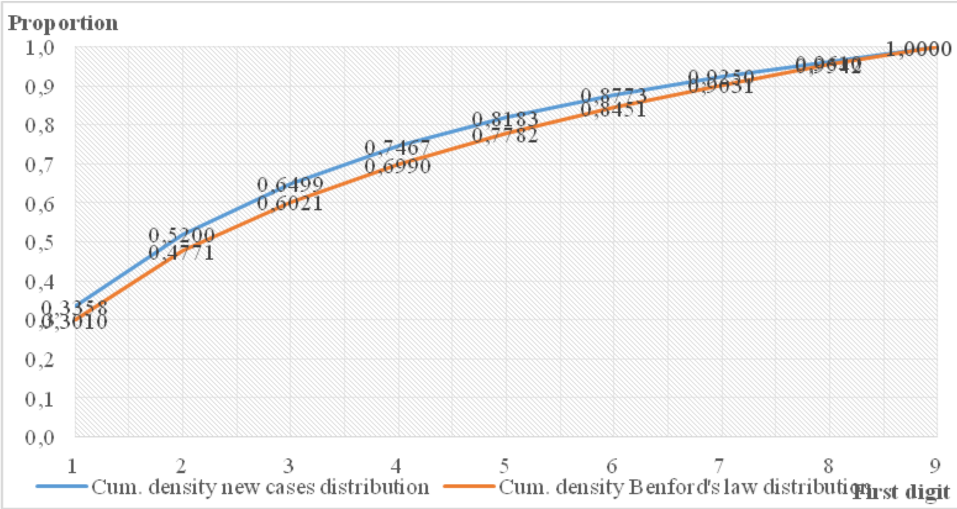

Fig. 1A Comparison of the first digit cumulative density distribution of COVID-19 new cases and the cumulative density Benford's Law distribution

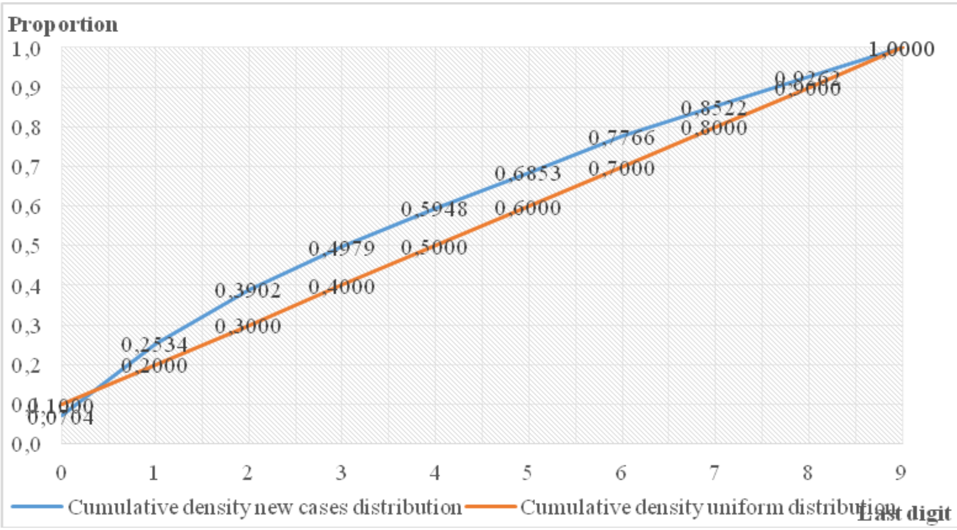

Fig. 2A Comparison of the last digit cumulative density distribution of COVID-19 new cases and the uniform distribution

Table A1 Results of the chi-square and Kolmogorov-Smirnov Z tests for the first and last digit, 206 countries and self-government dependencies

Country	Continent	First digit chi-square test			First digit Kolmogorov-Smirnov Z test			Last digit chi-square test			Last digit Kolmogorov-Smirnov Z test		
		Empirical chi-square	p-value	Decision	Kolmogorov-Smirnov Z empirical (D)	Kolmogorov-Smirnov Z theoretical	Decision	Empirical chi-square	p-value	Decision	Kolmogorov-Smirnov Z empirical (D)	Kolmogorov-Smirnov Z theoretical	Decision
Afghanistan	Asia	15.90	0.04	H1	0.16	0.22	H0	5.16	0.82	H0	0.09	0.22	H0
Albania	Europe	16.18	0.04	H1	0.23	0.20	H1	4.56	0.87	H0	0.06	0.20	H0
Algeria	Africa	12.99	0.11	H0	0.13	0.20	H0	9.92	0.36	H0	0.09	0.20	H0
Andorra	Europe	7.91	0.44	H0	0.13	0.22	H0	12.00	0.21	H0	0.09	0.22	H0
Angola	Africa	4.91	0.77	H0	0.25	0.41	H0	20.82	0.01	H1	0.43	0.41	H1
Anguilla	America	2.50	0.96	H0	0.52	0.96	H0	8.00	0.53	H0	0.70	0.96	H0
Antigua and Barbuda	America	8.82	0.36	H0	0.19	0.45	H0	18.78	0.03	H1	0.39	0.45	H0
Argentina	America	23.08	0.00	H1	0.19	0.20	H0	6.78	0.66	H0	0.12	0.20	H0
Armenia	Europe	23.79	0.00	H1	0.18	0.21	H0	11.19	0.26	H0	0.14	0.21	H0
Aruba	America	8.28	0.41	H0	0.16	0.29	H0	18.00	0.04	H1	0.22	0.29	H0
Australia	Oceania	3.86	0.87	H0	0.08	0.16	H0	27.09	0.00	H1	0.22	0.16	H1
Austria	Europe	11.84	0.16	H0	0.13	0.18	H0	6.33	0.71	H0	0.09	0.18	H0
Azerbaijan	Europe	15.17	0.06	H0	0.15	0.21	H0	2.81	0.97	H0	0.04	0.21	H0
Bahamas	America	13.22	0.10	H0	0.23	0.26	H0	44.86	0.00	H1	0.43	0.26	H1
Bahrain	Asia	4.07	0.85	H0	0.06	0.18	H0	5.55	0.78	H0	0.07	0.18	H0
Bangladesh	Asia	11.34	0.18	H0	0.15	0.23	H0	10.71	0.30	H0	0.08	0.23	H0
Barbados	America	4.54	0.81	H0	0.13	0.28	H0	33.96	0.00	H1	0.30	0.28	H1
Belarus	Europe	11.01	0.20	H0	0.11	0.23	H0	8.71	0.46	H0	0.06	0.23	H0
Belgium	Europe	22.18	0.00	H1	0.25	0.18	H1	7.48	0.59	H0	0.07	0.18	H0
Belize	America	12.42	0.13	H0	0.45	0.39	H1	58.00	0.00	H1	0.55	0.39	H1
Benin	Africa	3.51	0.90	H0	0.15	0.38	H0	7.77	0.56	H0	0.25	0.38	H0
Bermuda	America	12.68	0.12	H0	0.19	0.32	H0	22.00	0.01	H1	0.23	0.32	H0
Bhutan	Asia	16.25	0.04	H1	0.70	0.51	H1	63.00	0.00	H1	0.80	0.51	H1
Bolivia	America	8.54	0.38	H0	0.07	0.21	H0	23.00	0.01	H1	0.13	0.21	H0
Bonaire, Saint Eustatius and Saba	America	6.77	0.56	H0	0.52	0.61	H0	29.00	0.00	H1	0.70	0.61	H1
Bosnia and Herzegovina	Europe	17.99	0.02	H1	0.19	0.22	H0	3.05	0.96	H0	0.05	0.22	H0
Botswana	Africa	4.50	0.81	H0	0.15	0.48	H0	12.00	0.21	H0	0.35	0.48	H0
Brazil	America	9.16	0.33	H0	0.18	0.19	H0	8.40	0.49	H0	0.08	0.19	H0
British Virgin Islands	America	4.89	0.77	H0	0.52	0.68	H0	21.00	0.01	H1	0.70	0.68	H1

Brunei Darussalam	Asia	9.52	0.30	H0	0.15	0.26	H0	20.57	0.01	H1	0.33	0.26	H1
Bulgaria	Europe	16.01	0.04	H1	0.26	0.20	H1	10.33	0.32	H0	0.07	0.20	H0
Burkina Faso	Africa	4.17	0.84	H0	0.15	0.23	H0	5.86	0.75	H0	0.06	0.23	H0
Burundi	Africa	4.73	0.79	H0	0.32	0.61	H0	13.00	0.16	H0	0.50	0.61	H0
Cambodia	Asia	12.34	0.14	H0	0.22	0.27	H0	44.77	0.00	H1	0.40	0.27	H1
Cameroon	Africa	6.54	0.59	H0	0.14	0.26	H0	18.56	0.03	H1	0.13	0.26	H0
Canada	America	41.35	0.00	H1	0.34	0.17	H1	13.39	0.15	H0	0.10	0.17	H0
Cape Verde	Africa	7.79	0.45	H0	0.26	0.35	H0	37.67	0.00	H1	0.43	0.35	H1
Cases on an international conveyance Japan	Other	4.16	0.84	H0	0.14	0.35	H0	4.33	0.89	H0	0.10	0.35	H0
Cayman Islands	America	11.43	0.18	H0	0.13	0.33	H0	13.00	0.16	H0	0.18	0.33	H0
Central African Republic	Africa	13.29	0.10	H0	0.52	0.39	H1	54.67	0.00	H1	0.70	0.39	H1
Chad	Africa	8.22	0.41	H0	0.27	0.36	H0	38.86	0.00	H1	0.41	0.36	H1
Chile	America	21.86	0.01	H1	0.16	0.19	H0	10.40	0.32	H0	0.14	0.19	H0
China	Asia	3.32	0.91	H0	0.03	0.14	H0	7.81	0.55	H0	0.05	0.14	H0
Colombia	America	15.20	0.06	H0	0.11	0.20	H0	14.64	0.10	H0	0.15	0.20	H0
Congo	Africa	6.83	0.56	H0	0.23	0.35	H0	5.67	0.77	H0	0.20	0.35	H0
Costa Rica	America	5.98	0.65	H0	0.15	0.20	H0	14.06	0.12	H0	0.20	0.20	H0
Cote dIvoire	Africa	18.78	0.02	H1	0.17	0.23	H0	5.11	0.82	H0	0.08	0.23	H0
Croatia	Europe	7.77	0.46	H0	0.14	0.18	H0	23.41	0.01	H1	0.17	0.18	H0
Cuba	America	27.60	0.00	H1	0.25	0.22	H1	6.38	0.70	H0	0.08	0.22	H0
Curacao	America	6.80	0.56	H0	0.41	0.45	H0	29.89	0.00	H1	0.60	0.45	H1
Cyprus	Europe	5.16	0.74	H0	0.10	0.21	H0	7.50	0.59	H0	0.13	0.21	H0
Czechia	Europe	8.16	0.42	H0	0.07	0.19	H0	3.38	0.95	H0	0.07	0.19	H0
Democratic Republic of the Congo	Africa	9.47	0.30	H0	0.19	0.25	H0	15.33	0.08	H0	0.17	0.25	H0
Denmark	Europe	17.41	0.03	H1	0.22	0.18	H1	14.36	0.11	H0	0.19	0.18	H1
Djibouti	Africa	3.67	0.89	H0	0.10	0.26	H0	6.70	0.67	H0	0.07	0.26	H0
Dominica	America	5.96	0.65	H0	0.32	0.48	H0	27.00	0.00	H1	0.45	0.48	H0
Dominican Republic	America	11.85	0.16	H0	0.11	0.22	H0	13.05	0.16	H0	0.11	0.22	H0
Ecuador	America	8.93	0.35	H0	0.08	0.19	H0	14.88	0.09	H0	0.10	0.19	H0
Egypt	Africa	10.66	0.22	H0	0.17	0.20	H0	5.98	0.74	H0	0.07	0.20	H0
El Salvador	America	7.74	0.46	H0	0.12	0.26	H0	5.96	0.74	H0	0.13	0.26	H0
Equatorial Guinea	Africa	10.02	0.26	H0	0.35	0.33	H1	27.12	0.00	H1	0.48	0.33	H1
Eritrea	Africa	4.42	0.82	H0	0.15	0.38	H0	12.38	0.19	H0	0.35	0.38	H0
Estonia	Europe	20.52	0.01	H1	0.10	0.20	H0	12.42	0.19	H0	0.10	0.20	H0
Eswatini	Africa	9.26	0.32	H0	0.32	0.38	H0	24.69	0.00	H1	0.52	0.38	H1
Ethiopia	Africa	6.88	0.55	H0	0.10	0.24	H0	20.50	0.02	H1	0.25	0.24	H1
Falkland Islands (Malvinas)	America	5.57	0.70	H0	0.30	0.61	H0	17.00	0.05	H1	0.40	0.61	H0
Faroe Islands	Europe	9.69	0.29	H0	0.10	0.30	H0	18.00	0.04	H1	0.35	0.30	H1

Fiji	Oceania	13.37	0.10	H0	0.45	0.39	H1	59.67	0.00	H1	0.62	0.39	H1
Finland	Europe	20.63	0.01	H1	0.15	0.19	H0	9.20	0.42	H0	0.09	0.19	H0
France	Europe	13.45	0.10	H0	0.14	0.17	H0	11.31	0.25	H0	0.08	0.17	H0
French Polynesia	Oceania	6.29	0.61	H0	0.22	0.30	H0	35.00	0.00	H1	0.40	0.30	H1
Gabon	Africa	11.66	0.17	H0	0.21	0.30	H0	17.57	0.04	H1	0.31	0.30	H1
Gambia	Africa	9.95	0.27	H0	0.53	0.55	H0	37.33	0.00	H1	0.63	0.55	H1
Georgia	Europe	16.10	0.04	H1	0.16	0.20	H0	16.17	0.06	H0	0.13	0.20	H0
Germany	Europe	12.55	0.13	H0	0.13	0.16	H0	9.94	0.36	H0	0.13	0.16	H0
Ghana	Africa	4.46	0.81	H0	0.10	0.26	H0	9.86	0.36	H0	0.19	0.26	H0
Gibraltar	Europe	4.54	0.81	H0	0.10	0.30	H0	4.00	0.91	H0	0.10	0.30	H0
Greece	Europe	23.41	0.00	H1	0.13	0.19	H0	11.85	0.22	H0	0.12	0.19	H0
Greenland	America	4.98	0.76	H0	0.40	0.51	H0	23.00	0.01	H1	0.60	0.51	H1
Grenada	America	6.71	0.57	H0	0.38	0.51	H0	20.14	0.02	H1	0.56	0.51	H1
Guam	Oceania	4.68	0.79	H0	0.06	0.25	H0	16.17	0.06	H0	0.19	0.25	H0
Guatemala	America	3.28	0.92	H0	0.13	0.22	H0	22.73	0.01	H1	0.19	0.22	H0
Guernsey	Europe	7.66	0.47	H0	0.14	0.26	H0	6.70	0.67	H0	0.10	0.26	H0
Guinea	Africa	3.70	0.88	H0	0.06	0.27	H0	5.00	0.83	H0	0.14	0.27	H0
Guinea Bissau	Africa	8.08	0.43	H0	0.17	0.38	H0	13.92	0.13	H0	0.30	0.38	H0
Guyana	America	3.50	0.90	H0	0.10	0.29	H0	18.00	0.04	H1	0.27	0.29	H0
Haiti	America	9.66	0.29	H0	0.26	0.29	H0	30.73	0.00	H1	0.46	0.29	H1
Holy See	Europe	9.56	0.30	H0	0.53	0.55	H0	37.33	0.00	H1	0.63	0.55	H1
Honduras	America	8.42	0.39	H0	0.15	0.21	H0	7.50	0.59	H0	0.10	0.21	H0
Hungary	Europe	8.08	0.43	H0	0.10	0.20	H0	10.33	0.32	H0	0.12	0.20	H0
Iceland	Europe	7.69	0.46	H0	0.16	0.19	H0	10.58	0.31	H0	0.10	0.19	H0
India	Asia	7.00	0.54	H0	0.09	0.18	H0	7.11	0.63	H0	0.14	0.18	H0
Indonesia	Asia	17.15	0.03	H1	0.29	0.20	H1	6.78	0.66	H0	0.08	0.20	H0
Iran	Asia	25.15	0.00	H1	0.25	0.17	H1	11.13	0.27	H0	0.13	0.17	H0
Iraq	Asia	5.98	0.65	H0	0.10	0.20	H0	5.75	0.76	H0	0.07	0.20	H0
Ireland	Europe	7.13	0.52	H0	0.10	0.19	H0	11.94	0.22	H0	0.09	0.19	H0
Isle of Man	Europe	10.03	0.26	H0	0.22	0.25	H0	18.24	0.03	H1	0.36	0.25	H1
Israel	Asia	18.00	0.02	H1	0.14	0.18	H0	9.33	0.41	H0	0.13	0.18	H0
Italy	Europe	37.73	0.00	H1	0.24	0.17	H1	13.98	0.12	H0	0.09	0.17	H0
Jamaica	America	16.32	0.04	H1	0.21	0.23	H0	20.71	0.01	H1	0.32	0.23	H1
Japan	Asia	11.74	0.16	H0	0.13	0.15	H0	3.54	0.94	H0	0.04	0.15	H0
Jersey	Europe	2.32	0.97	H0	0.07	0.28	H0	5.26	0.81	H0	0.14	0.28	H0
Jordan	Asia	9.59	0.30	H0	0.08	0.23	H0	14.00	0.12	H0	0.11	0.23	H0
Kazakhstan	Asia	8.79	0.36	H0	0.17	0.22	H0	8.44	0.49	H0	0.09	0.22	H0
Kenya	Africa	10.48	0.23	H0	0.15	0.24	H0	10.94	0.28	H0	0.10	0.24	H0
Kosovo	Europe	13.53	0.09	H0	0.12	0.24	H0	9.12	0.43	H0	0.05	0.24	H0
Kuwait	Asia	9.86	0.28	H0	0.12	0.18	H0	5.91	0.75	H0	0.13	0.18	H0

Kyrgyzstan	Asia	7.61	0.47	H0	0.20	0.25	H0	14.00	0.12	H0	0.13	0.25	H0
Laos	Asia	4.81	0.78	H0	0.30	0.43	H0	24.00	0.00	H1	0.50	0.43	H1
Latvia	Europe	8.64	0.37	H0	0.12	0.20	H0	10.23	0.33	H0	0.10	0.20	H0
Lebanon	Asia	4.80	0.78	H0	0.11	0.19	H0	14.47	0.11	H0	0.09	0.19	H0
Liberia	Africa	6.73	0.57	H0	0.19	0.31	H0	14.16	0.12	H0	0.28	0.31	H0
Libya	Africa	7.15	0.52	H0	0.19	0.35	H0	21.67	0.01	H1	0.33	0.35	H0
Liechtenstein	Europe	9.37	0.31	H0	0.22	0.28	H0	34.33	0.00	H1	0.38	0.28	H1
Lithuania	Europe	19.20	0.01	H1	0.23	0.21	H1	9.50	0.39	H0	0.13	0.21	H0
Luxembourg	Europe	9.44	0.31	H0	0.17	0.20	H0	4.11	0.90	H0	0.07	0.20	H0
Madagascar	Africa	8.61	0.38	H0	0.23	0.28	H0	8.50	0.48	H0	0.10	0.28	H0
Malawi	Africa	12.66	0.12	H0	0.31	0.41	H0	37.18	0.00	H1	0.51	0.41	H1
Malaysia	Asia	29.03	0.00	H1	0.29	0.17	H1	16.42	0.06	H0	0.20	0.17	H1
Maldives	Asia	11.63	0.17	H0	0.31	0.31	H1	28.89	0.00	H1	0.33	0.31	H1
Mali	Africa	9.07	0.34	H0	0.10	0.27	H0	9.00	0.44	H0	0.08	0.27	H0
Malta	Europe	5.41	0.71	H0	0.10	0.20	H0	6.33	0.71	H0	0.11	0.20	H0
Mauritania	Africa	8.79	0.36	H0	0.53	0.55	H0	37.33	0.00	H1	0.70	0.55	H1
Mauritius	Africa	6.95	0.54	H0	0.13	0.26	H0	9.14	0.42	H0	0.08	0.26	H0
Mexico	America	6.66	0.57	H0	0.12	0.20	H0	13.21	0.15	H0	0.15	0.20	H0
Moldova	Europe	13.12	0.11	H0	0.21	0.20	H1	22.33	0.01	H1	0.19	0.20	H0
Monaco	Europe	8.91	0.35	H0	0.13	0.25	H0	23.33	0.01	H1	0.30	0.25	H1
Mongolia	Asia	14.53	0.07	H0	0.43	0.35	H1	57.67	0.00	H1	0.50	0.35	H1
Montenegro	Europe	7.48	0.49	H0	0.13	0.23	H0	3.57	0.94	H0	0.06	0.23	H0
Montserrat	America	2.46	0.96	H0	0.30	0.61	H0	9.00	0.44	H0	0.50	0.61	H0
Morocco	Africa	24.34	0.00	H1	0.20	0.20	H0	21.45	0.01	H1	0.21	0.20	H1
Mozambique	Africa	3.75	0.88	H0	0.18	0.34	H0	22.75	0.01	H1	0.38	0.34	H1
Myanmar	Asia	3.53	0.90	H0	0.13	0.28	H0	20.91	0.01	H1	0.30	0.28	H1
Namibia	Africa	4.42	0.82	H0	0.30	0.48	H0	17.00	0.05	H1	0.50	0.48	H1
Nepal	Asia	8.99	0.34	H0	0.40	0.38	H1	29.31	0.00	H1	0.52	0.38	H1
Netherlands	Europe	21.73	0.01	H1	0.15	0.18	H0	7.93	0.54	H0	0.09	0.18	H0
New Caledonia	Oceania	11.50	0.17	H0	0.41	0.45	H0	29.89	0.00	H1	0.59	0.45	H1
New Zealand	Oceania	7.77	0.46	H0	0.09	0.21	H0	8.00	0.53	H0	0.09	0.21	H0
Nicaragua	America	9.79	0.28	H0	0.52	0.48	H1	42.00	0.00	H1	0.70	0.48	H1
Niger	Africa	8.71	0.37	H0	0.10	0.27	H0	7.40	0.60	H0	0.08	0.27	H0
Nigeria	Africa	2.73	0.95	H0	0.09	0.23	H0	6.78	0.66	H0	0.13	0.23	H0
North Macedonia	Europe	13.13	0.11	H0	0.11	0.21	H0	17.78	0.04	H1	0.16	0.21	H0
Northern Mariana Islands	Oceania	6.13	0.63	H0	0.30	0.55	H0	14.00	0.12	H0	0.50	0.55	H0
Norway	Europe	12.11	0.15	H0	0.14	0.18	H0	7.73	0.56	H0	0.13	0.18	H0
Oman	Asia	11.13	0.19	H0	0.21	0.21	H0	18.76	0.03	H1	0.19	0.21	H0
Pakistan	Asia	7.33	0.50	H0	0.14	0.21	H0	4.67	0.86	H0	0.10	0.21	H0
Palestine	Asia	17.40	0.03	H1	0.14	0.22	H0	19.72	0.02	H1	0.11	0.22	H0
Panama	America	16.16	0.04	H1	0.20	0.20	H0	7.82	0.55	H0	0.11	0.20	H0

Papua New Guinea	Oceania	6.63	0.58	H0	0.45	0.68	H0	21.00	0.01	H1	0.55	0.68	H0
Paraguay	America	12.91	0.11	H0	0.11	0.23	H0	20.11	0.02	H1	0.23	0.23	H1
Peru	America	16.02	0.04	H1	0.08	0.20	H0	24.00	0.00	H1	0.11	0.20	H0
Philippines	Asia	13.12	0.11	H0	0.18	0.20	H0	10.96	0.28	H0	0.18	0.20	H0
Poland	Europe	11.43	0.18	H0	0.10	0.19	H0	6.71	0.67	H0	0.11	0.19	H0
Portugal	Europe	18.73	0.02	H1	0.26	0.19	H1	6.08	0.73	H0	0.07	0.19	H0
Puerto Rico	America	32.69	0.00	H1	0.28	0.27	H1	4.00	0.91	H0	0.05	0.27	H0
Qatar	Asia	8.89	0.35	H0	0.10	0.19	H0	7.94	0.54	H0	0.06	0.19	H0
Romania	Europe	17.75	0.02	H1	0.20	0.19	H1	9.92	0.36	H0	0.16	0.19	H0
Russia	Europe	12.45	0.13	H0	0.11	0.20	H0	7.48	0.59	H0	0.10	0.20	H0
Rwanda	Africa	4.08	0.85	H0	0.07	0.24	H0	10.94	0.28	H0	0.17	0.24	H0
Saint Kitts and Nevis	America	8.21	0.41	H0	0.41	0.45	H0	36.56	0.00	H1	0.59	0.45	H1
Saint Lucia	America	4.68	0.79	H0	0.27	0.51	H0	20.14	0.02	H1	0.41	0.51	H0
Saint Vincent and the Grenadines	America	6.68	0.57	H0	0.37	0.55	H0	24.00	0.00	H1	0.47	0.55	H0
San Marino	Europe	5.92	0.66	H0	0.10	0.20	H0	13.21	0.15	H0	0.11	0.20	H0
Sao Tome and Principe	Africa	7.16	0.52	H0	0.48	0.96	H0	8.00	0.53	H0	0.50	0.96	H0
Saudi Arabia	Asia	7.20	0.52	H0	0.15	0.20	H0	14.87	0.09	H0	0.11	0.20	H0
Senegal	Africa	9.51	0.30	H0	0.20	0.20	H0	15.55	0.08	H0	0.20	0.20	H0
Serbia	Europe	10.56	0.23	H0	0.12	0.20	H0	13.27	0.15	H0	0.13	0.20	H0
Seychelles	Africa	8.77	0.36	H0	0.52	0.48	H1	34.50	0.00	H1	0.70	0.48	H1
Sierra Leone	Africa	6.78	0.56	H0	0.27	0.34	H0	32.75	0.00	H1	0.45	0.34	H1
Singapore	Asia	6.89	0.55	H0	0.09	0.15	H0	27.19	0.00	H1	0.18	0.15	H1
Sint Maarten	America	7.47	0.49	H0	0.17	0.35	H0	12.33	0.20	H0	0.27	0.35	H0
Slovakia	Europe	12.81	0.12	H0	0.19	0.20	H0	9.89	0.36	H0	0.12	0.20	H0
Slovenia	Europe	31.43	0.00	H1	0.18	0.19	H0	6.71	0.67	H0	0.10	0.19	H0
Somalia	Africa	5.12	0.74	H0	0.18	0.34	H0	11.50	0.24	H0	0.15	0.34	H0
South Africa	Africa	6.49	0.59	H0	0.08	0.20	H0	5.74	0.77	H0	0.07	0.20	H0
South Korea	Asia	7.43	0.49	H0	0.07	0.15	H0	11.75	0.23	H0	0.08	0.15	H0
South Sudan	Africa	9.29	0.32	H0	0.70	0.68	H1	36.00	0.00	H1	0.80	0.68	H1
Spain	Europe	20.29	0.01	H1	0.25	0.17	H1	6.70	0.67	H0	0.08	0.17	H0
Sri Lanka	Asia	5.63	0.69	H0	0.08	0.21	H0	8.50	0.48	H0	0.13	0.21	H0
Sudan	Africa	8.99	0.34	H0	0.34	0.33	H1	18.88	0.03	H1	0.38	0.33	H1
Suriname	America	4.26	0.83	H0	0.32	0.61	H0	13.00	0.16	H0	0.50	0.61	H0
Sweden	Europe	8.27	0.41	H0	0.08	0.18	H0	10.62	0.30	H0	0.08	0.18	H0
Switzerland	Europe	4.72	0.79	H0	0.13	0.18	H0	11.50	0.24	H0	0.10	0.18	H0
Syria	Asia	12.95	0.11	H0	0.23	0.39	H0	16.33	0.06	H0	0.30	0.39	H0
Taiwan	Asia	21.34	0.01	H1	0.26	0.18	H1	51.33	0.00	H1	0.32	0.18	H1
Thailand	Asia	10.94	0.21	H0	0.18	0.18	H0	19.73	0.02	H1	0.20	0.18	H1
Timor Leste	Asia	6.22	0.62	H0	0.40	0.48	H0	24.50	0.00	H1	0.60	0.48	H1

Togo	Africa	4.24	0.83	H0	0.13	0.27	H0	22.46	0.01	H1	0.31	0.27	H1
Trinidad and Tobago	America	4.54	0.81	H0	0.16	0.25	H0	38.24	0.00	H1	0.39	0.25	H1
Tunisia	Africa	14.33	0.07	H0	0.15	0.21	H0	11.44	0.25	H0	0.20	0.21	H0
Turkey	Asia	24.98	0.00	H1	0.20	0.21	H0	12.41	0.19	H0	0.08	0.21	H0
Turks and Caicos islands	America	9.44	0.31	H0	0.41	0.51	H0	34.43	0.00	H1	0.60	0.51	H1
Uganda	Africa	7.19	0.52	H0	0.23	0.35	H0	27.00	0.00	H1	0.33	0.35	H0
Ukraine	Europe	9.07	0.34	H0	0.07	0.22	H0	4.33	0.89	H0	0.05	0.22	H0
United Arab Emirates	Asia	12.30	0.14	H0	0.15	0.20	H0	14.49	0.11	H0	0.24	0.20	H1
United Kingdom	Europe	36.71	0.00	H1	0.18	0.17	H1	7.35	0.60	H0	0.11	0.17	H0
United Republic of Tanzania	Africa	4.86	0.77	H0	0.11	0.29	H0	28.00	0.00	H1	0.25	0.29	H0
United States of America	America	20.24	0.01	H1	0.21	0.16	H1	28.01	0.00	H1	0.14	0.16	H0
United States Virgin Islands	America	4.87	0.77	H0	0.15	0.39	H0	11.33	0.25	H0	0.27	0.39	H0
Uruguay	America	14.34	0.07	H0	0.11	0.22	H0	12.54	0.18	H0	0.16	0.22	H0
Uzbekistan	Asia	5.08	0.75	H0	0.05	0.22	H0	3.27	0.95	H0	0.04	0.22	H0
Venezuela	America	9.47	0.30	H0	0.08	0.26	H0	5.96	0.74	H0	0.08	0.26	H0
Vietnam	Asia	16.64	0.03	H1	0.12	0.19	H0	60.59	0.00	H1	0.28	0.19	H1
Yemen	Asia	2.32	0.97	H0	0.70	1.36	H0	9.00	0.44	H0	0.80	1.36	H0
Zambia	Africa	4.88	0.77	H0	0.09	0.31	H0	15.21	0.09	H0	0.29	0.31	H0
Zimbabwe	Africa	10.25	0.25	H0	0.33	0.35	H0	48.33	0.00	H1	0.53	0.35	H1

Source: Authors' calculations

* Cases on an international conveyance Japan – related to COVID-19 cases on the cruise ship Princess Diamond

* Guernsey, Isle of Man, Jersey are self-governing dependencies of the United Kingdom