COVID

MDPI

Article

Forensic Analysis of COVID-19 Data from 198 Countries Two Years after the Pandemic Outbreak

Noah Farhadi [1] and Hooshang Lahooti [2,*]

[1] Department of Business, Design and Technology, Macromedia University of Applied Sciences, 20095 Hamburg, Germany; n.farhadi@macromedia.de
[2] Sydney Medical School, Nepean Clinical School, University of Sydney, Sydney, NSW 2747, Australia
* Correspondence: hooshang.lahooti@sydney.edu.au

Abstract: The availability of accurate information has proved fundamental to managing health crises. This research examined pandemic data provided by 198 countries worldwide two years after the outbreak of the deadly Coronavirus in Wuhan, China. We compiled and reevaluated the consistency of daily COVID-19 infections with Benford's Law. It is commonly accepted that the distribution of the leading digits of pandemic data should conform to Newcomb-Benford's expected frequencies. Consistency with the law of leading digits might be an indicator of data reliability. Our analysis shows that most countries have disseminated partially reliable data over 24 months. The United States, Israel, and Spain spread the most consistent COVID-19 data with the law. In line with previous findings, Belarus, Iraq, Iran, Russia, Pakistan, and Chile published questionable epidemic data. Against this trend, 45 percent of countries worldwide appeared to demonstrate significant BL conformity. Our measures of Benfordness were moderately correlated with the Johns Hopkins Global Health Security Index, suggesting that the conformity to Benford's law may also depend on national health care policies and practices. Our findings might be of particular importance to policymakers and researchers around the world.

Keywords: COVID-19; forensic; Benford's law; pandemic; public health

Citation: Farhadi, N.; Lahooti, H. Statistical Analysis of COVID-19 Data from 198 Countries Two Years after the Pandemic Outbreak. **2022**, *1*, 472–485. https://doi.org/10.3390/covid2040034

Academic Editor: Guglielmo Campus

Received: 31 January 2022
Accepted: 22 March 2022
Published: 30 March 2022

1. Introduction

Over two years ago, the World Health Organization (WHO) reported the emergence of the new RNA virus—known as Coronavirus, COVID-19, or SARS-CoV-2—in Wuhan, China [1]. Since its outbreak, the world has faced millions of cases of infection. The continuing trends are still alarming. Researchers have identified several mutations of the deadly virus. The world has mobilized an army of scientists, doctors, pharmaceutical companies, and health experts to date to combat the pandemic's rapid growth. These efforts have led to a unique global ecosystem [2] of COVID-19 research and business communities nascent.

In this context, a tidal wave of data has been emerging. More than 200 territories globally have been reporting two primary metrics for the outbreak's unfolding, namely "new cases"—people who have tested positive for the virus—and "new deaths"—the incidence of deaths due to infection with the virus [3]. After introducing COVID-19 testing, countries began to report the number of Coronavirus tests conducted as "new daily tests." The emergence of vaccines in a bumper crop allowed effective preventive measures to control the highly infectious virus. Countries commenced reporting "new daily vaccinations," or the number of people vaccinated each day.

There is no doubt that reliable data are vital to tackle the pace of the global pandemic. Researchers base their studies on publicly available data. Stringent measures, known colloquially as lockdowns, travel bans, and flight restrictions, have been justified and imposed in numerous countries based on such data. For this reason, public faith in COVID-19 data is paramount to the effective implementation of each of these interventions. Public misgivings about the reliability of Coronavirus data can lead to a sluggish response of societies or, in the worst case, a lack of public support [3].

Various scientists performed forensic studies on the COVID-19 data. Unanimously, extant research provided empirical proof that COVID-19 data are to some extent reliable. The basic idea of these studies goes back to applying the mathematical theory, Benford's law, which was pioneered initially as the so-called "logarithmic law" by the Canadian American astronomer and mathematician, Simon Newcomb, in the American Journal of Mathematics in 1881 [4]. Later in 1938, Frank Benford introduced the thorough proof of Newcomb's notion and provided empirical evidence for the widespread existence of the theory [5]. We will explain Benford's law (BL) in detail later in the Methods section of this paper.

Since the onset of the pandemic, many papers emerged, which in part introduced contradictory results. Some articles even included misleading claims and discussions on statistical tests. As one of the first extensive studies on COVID-19, Sambridge and Jackson operationalized Newcomb Benford's law and analyzed pandemic data from 51 countries from 16 January 2020 to 9 April 2020; they offered supporting evidence for the application of the law in the context of pandemics. They showed partial integrity of global Coronavirus data in broad terms [6]. Around the same time, Farhadi examined over 100,000 integers from 154 countries and applied three goodness-of-fit criteria [3]. Approximately 28% of countries adhered well to the frequency of distribution expected, while six countries contained completely unreliable data. Farhadi recommended replicating the same study and incorporating additional goodness of fit tests based on a larger sample of observations. Extending the time frame and improving statistical measurement in the future will enhance the body of knowledge.

Other researchers conducted Benford's law focusing on smaller samples. Koch and Okamura [7] verified the reliability of the official data of COVID-19 hotspots, the United States, China, and Italy. Idrovo and Manrique-Hernández also proved China's compliance with the law [8]. Lee et al. noted that Japan did not supply robust data; they also argued that when the epidemic growth curve follows an exponential distribution, the number of infections and deaths will obey BL [9]. According to Isea, China, Germany, Brazil, Venezuela, Norway, South Africa, Singapore, Ecuador, Egypt, Ireland, France, Australia, Colombia, India, Russia, and Croatia comply with the law. However, Italy, Portugal, the Netherlands, the United Kingdom, Denmark, Belgium, and Chile failed three compliance tests [10].

Farhadi and Lahooti investigated the progress of Benfordness in 182 countries from 1 January 2020 to 6 June 2021 to explain previously reported results in COVID-19 reliability by expanding both the time frame and statistical tests [11]. The researchers compiled data on the inter-reliability of over 200,000 natural numbers and compared it with previous evidence in forensic statistics. They proved that about 32% of the nations achieved measurable Benfordness enhancements, while 68.2% showed no improvement. The selfsame study detected a moderate correlation between the goodness of fit tests for assessing Benfordness and the 2019 Johns Hopkins Global Health Risk Index, implying a good relationship between national health systems and policies and the trustworthiness of the epidemic data.

In another study, Morillas-Jurado et al. examined the COVID-19 epidemic specifically during the first wave of the pandemic between February and August 2020 in Spain [12]. The authors documented anomalies in the data that occurred in six Spanish regions. The study conducted is a prime example of the poor application of BL in the context of pandemics. Their report did not provide precise figures on the observed frequencies. In addition, the regional data sets were small, in fact too small to assess the conformity of COVID-19 data to BL. Furthermore, the researchers ignored that certain epidemic management limitations at the beginning of the outbreak may have affected the distribution of leading digits.

The body of knowledge includes some discussion to investigate the reasons for non-compliance of pandemic data with the law. In an extensive examination of pandemic growth, Farhadi and Lahooti empirically provided the notion that the greater the mean growth rates of new cases at the beginning of the epidemic, the larger the distance from Benford's distribution [13]. The same authors collected data from 176 countries and explored the association between pandemic growth and the reliability of the global data. Their findings suggested that the initial exponential growth factors within the first nine months of the worldwide pandemic underlie the overall divergence from Benford's Law later. These outcomes were coherent with the notion provided in earlier studies. Lee et al. [9] suggested that Benfordness decreases as the growth factors of daily cases flatten. Farhadi and Lahooti reasoned that assessment of BL compliance should focus on new cases, which are less influenced by the countries' health care processes and policy boundaries. Peaks and troughs may occur when daily testing or vaccinations are administered in some countries.

Balashov et al. studied COVID-19 data from 185 countries to gauge pandemic data accuracy [14]. They used cumulated daily data, applied multiple goodness-of-fit tests, and claimed to be the pioneers of large-scale BL research into COVID-19 worldwide data while neglecting similar previous COVID-19 studies focusing on global data [3,6,11,12]. Balashov et al. did not provide details of their BL findings, partially neglected the extant body of knowledge, and suggested a regressive relationship between democracy and epidemic data accuracy. The authors assumed that BL divergence is caused by data manipulation without considering the effect of exponential growth rates on conformity to the law. Furthermore, they reported that undemocratic countries showed the empirical distance to BL distribution. However, these findings were nothing new. Kilani, in July 2021, initiated and assessed Benfordness of COVID-19 daily incidents and investigated the relationship between democratic measures and conformity to the law [15]. By correlating the results with four democracy and freedom indices using ordinary least squares, Kilani provided evidence for the notion that countries with high freedom scores do show consistency with the logarithmic law. Notably, the World Press Freedom index showed the most substantial relationship with the expected frequencies. Table 1 summarizes the notable studies on COVID-19 data, emphasizing the Newcomb–Benford law.

Table 1. Prior research into COVID-19 data reliability.

Researcher	Variables	Deadline	Number of Countries
Idrovo and Manrique-Hernández	Confirmed cases, suspected cases, and deaths cumulated confirmed cases and cumulated actual deaths	21 January 2020–15 March 2020	1
Koch and Okamura	Daily Cases, Deaths	20 January 2020–28 April 2020	3
Lee, Han and Jeong	Daily Deaths	22 January 2020–6 April 2020	10
Wei and Vellwock	Daily Cases, Deaths	1 January 2020–1 September 2020	20
Isea	Daily Cases, Deaths	29 December 2019–30 April 2020	23
Jackson and Sambridge	Cumulated confirmed cases and deaths	16 January 2020–4 September 2020	51
Farhadi	Daily Cases, Deaths, Tests	31 December 2019–24 September 2020	182
Farhadi and Lahooti	Daily Cases, Deaths, Tests, Vaccination	31 December 2019–6 June 2021	176
Farhadi and Lahooti	Periodic growth ratios, Daily Cases	31 December 2019–6 June 2021	176
Morillas-Jurado et al.	Daily death cases	1 March 2020–30 June 2020	1

There are three major underlying concerns with the studies to date: (a) inconsistent statistical results due to small sample size, (b) misleading application of goodness of fit tests to assess Benfordness, and (c) disingenuous use of variables to examine Benfordness. These limitations may have led to deceptively conservative statistical results.

Our primary concern is to go beyond the previous limitations and comprehensively assess the most recent data. Therefore, we aim to evaluate Benfordness by improving the statistical tests and the quality of raw data from over 200 countries within 24 months since the onset of the coronavirus pandemic in 2020. This research examines the central question: "After two years, are countries still telling the truth about the spread of the virus in their territories?"

2. Method

2.1. Benford's Law and Goodness-of-Fit Tests

Benford's Law (BL), also known as the "Law of First Digits," is a widely known technique that refers to the frequency of first digits in naturally generated data sets. The idea was initiated and grounded on the specification of the probabilities of the first digits with Equation (1). According to BL, the leading digits of numerals adhere to a particular logarithmic pattern: 30.1% for one, 17.6% for two, 12.5% for three, 9.7% for four, 7.9% for five, 6.7% for six, 5.8% for seven, 5.1% for eight, and 4.6% for nine [4,5]. See Table 2.

$$P(d) = log_{10}\left(1 + \frac{1}{d}\right)\ d\ \in \{1, 2, 3, \dots, 9\} \quad (1)$$

Table 2. Benford's Law Distribution of First Digit.

First Digit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Benford's frequency	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

BL has been empirically and elusively proven in various distributions, such as stock market indexes [16], the numbers of cases by country of (almost any type of) infectious diseases reported to the World Health Organization (WHO) [17]. Since pandemic data follow an exponential logistic curve, BL is useful for studying Benfordness of infectious disease samples that show progressive growth over time, especially in the early stages of dissemination or the arrival of a new variant during pandemics [9,13].

It is generally accepted that BL applies to a data set exhibiting exponential growth. In some cases, this pattern is not proven [18]. BL is best applied to data sets that span multiple orders of magnitude (e.g., population counts of cities, income distributions). Not all data sets conform to this theory, especially when the underlying data do not exhibit a geometric trend characterized by the absence of minima and maxima. Or, to put it another way, when data contain numbers tied to a maximum or minimum, such as individuals' height, weight, and intelligence quotient. The theory is also not feasible when the data comprise numbers that span only a few orders of magnitude. Furthermore, artificially constructed data sets can contradict the expected BL distributions [19].

BL is a common practice in forensic studies within the social sciences and has been applied in various disciplines, e.g., finance and accounting [20–22], politics [23,24], medicine [25], and pandemics [6–18]. The body of knowledge is grounded upon various goodness-of-fit tests that evaluate the deviation between observed and expected frequencies. The methods frequently used in prior research are, *inter alia*, the *Kolmogorov–Smirnov*, *Kuiper* for continuous data as well as the *Chi-square test*, the *Euclidean Distance*, and *M-statistics* for discrete data [26–28].

In the context of epidemic distributions, the goodness of fit tests compares the leading digits' observed proportions to the expected BL frequencies. The *Chi*-square test is sensitive to the sample size and does not allow reliable inferences when a data set contains

5000 observations or more [3,13,26–32]. The Chi-square goodness-of-fit test can be applied to discrete distributions when there are no parameters that need to be estimated [28,29]. If the sample size is too large, the null hypothesis can be rejected with a high degree of probability (even if there is no significant difference between the actual and the expected subset). With a small sample size, χ^2 encounters difficulties in measurement too. It is calculated as follows:

$$\chi^2 = \sum_{i=1}^{9} \frac{\left(\tilde{p}_i - p_{i)}\right)^2}{p_i} \tag{2}$$

The Euclidean test (*d** or *d*-factor) ultimately measures the Euclidean distance between the measured and expected frequencies of leading digits in discrete data, where $\tilde{p}_i$ and p_i are the observed and expected frequencies [3,23,30]. According to Goodman, the *d*-factor quantifies the distance between the sample and the cumulative distribution function of the reference dataset after normalization by 1.03606, the maximum possible distance [30], which converts the Euclidean test into a value bounded by 0 and 1. A *d** equal to 0 suggests full conformity to BL, while the highest Euclidian distance, *d** = 1, signifies full non-conformity. As a rule of thumb, perhaps, *d** of 0.25 or higher may indicate non-conformity to the law [3,30,31]. The Euclidean distance is defined as follows [20,23]:

$$d^* = \frac{1}{1.03606} \sqrt{\sum_{i=1}^{9} [\tilde{p}_i - p_i]^2} \tag{3}$$

Another commonly applied (but a simple maximum norm) statistic is *Chebyshev Distance Test* or *M-statistic* [15,16]. The M-statistic concentrates on the maximum deviation of the observed proportions from the expected BL values, above or below the anticipated frequencies (see Equation (4)). According to Morrow [31], rejection regions for common test levels are 0.967 and 1.212 for $\alpha = 0.05$ and $\alpha = 0.01$, respectively:

$$M = \sqrt{n} \times \max_{i=1,2,\ldots,9} \{|O_i - E_i|\} \tag{4}$$

Consistent with Fairweather [20], we operationalize *Weighted Maximum Statistic*, the largest term in χ^2 and a modification of *Leemis M-statistic*; mathematically, it is delineated as follows; see Equation (5):

$$w = n \times \max_{i=1,2,\ldots,9} \frac{|O_i - E_i|^2}{E_i} \tag{5}$$

To align with previous studies, we posit the following null hypothesis for all countries in the scope, mainly, H_0: COVID-19 data from *Ji* adhere to BL, where *Ji* stands for individual jurisdiction in the range. In this study, we initialized the goodness-of-fit tests based on a significance level. Notably, several previous studies in the pandemic, financial, or forensic fields suggest that data manipulations cause a divergence from the BL. The hypothesis in our study did not assume that variation in BL is an indication of fraud or data manipulation. Deviation from the BL distribution may also be due to inconsistent policies or an indicator of national public health countermeasures against the spread of epidemics. Thus, the hypothesis here addresses the conformity to BL distribution only.

2.2. COVID-19 Data Sampling

As data sets grow over time, compliance with the BL can be improved by expanding the sample size [3,9,13]. Two years after the pandemic outbreak, the available data on new cases and new deaths have reached a sufficient level to conduct empirical tests. We sampled from the COVID-19 database at the *Centre for Systems Science and Engineering* at *Johns Hopkins University* and the *Our World in Data repository*. Nine cases were ruled out after consolidating all regions with small samples, including Vatican, Congo, and Tanzania.

We collected 404,489 integers, including 159,832 new cases, 142,336 new deaths, 66,218 new tests, and 36,103 new deaths from 201 countries and territories between 1 January 2020 and 18 February 2022. We intentionally excluded recent deaths, new tests, and new vaccinations as these values exhibit certain minima and maxima in the data due to the inherent limitations of national health policies and capabilities that may affect testing procedures and patient care services—the WHO addressed early that some nations do not have sufficient access to testing kits, for instance [27]. Statistical examination of Benfordness makes sense if the COVID-19 data set is of a substantial range. There is no explicit minimum threshold as guidance. If a data set is too small, BL will be ineffective in spotting abnormalities.

Daily incidents—new cases and new deaths—were less suitable for the BL assessment. Detailed analysis of the data revealed that 62 percent of countries had an order of magnitude of three or less. Previous research sought to alleviate the range of observed data by accumulating daily incidents. Over time, however, the cumulative data can become overly large and skew the observed frequencies of leading digits. As such, assessment of data highly likely leads to falsified outcomes. We overcame this issue by moving away from daily toward periodic data intervals. Data in a three-day, five-day, or seven-day interval led to obvious improvements. Our analysis discovered that the five-day interval spanned 50 percent of territories with an order of magnitude of four or larger. We therefore accumulated daily data based on a five-day interval leading to a final sample of 98 countries with an order of magnitude of at least four and an average sample size (n) of 146 sporadic cases for each territory.

3. Results

We evaluated the conformity of the observed leading digits to the expected BL frequencies. The goodness of fit test, Pearson χ^2, was operationalized. We furthermore calculated the $d^*$and M-statistic based on a significant level. These tests were frequently used for examining discrete pandemics data. Table 3 summarizes the detailed results for all countries investigated in this study. Figure 1 summarizes our results in a nutshell.

Seven countries failed to comply with BL based on all tests. These countries are Iraq, Belarus, Dominican Republic, Chile, Iran, Colombia, and Pakistan. We confidently rejected the null hypothesis for these jurisdictions. Fifty-one jurisdictions (52%) passed the χ^2, *Chebyshev Distance,* and *Euclidean Distance* tests. *Chebyshev Distance* test red-flagged Iraq, Chile, and Iran ($2.81 \leq m \leq 3.77$). Seventeen (83%) countries passed the test according to the Chi-square results. Belarus, Iraq, Pakistan, and Russia showed the largest distance to BL accordingly ($\chi2 > 55$; p <0.001). Our findings are overwhelmingly consistent with prior research, including equivalent conclusions for the stated jurisdictions [3,13].

Other notable countries that fully conformed to BL are Australia, Israel, the United Kingdom, the United States, and Spain. It is noteworthy to emphasize that the null hypothesis can be overwhelmingly confirmed for these countries. Spain showed the smallest distance to BL frequencies based on at least two measures and disclosed the highest BL conformity (see Figures 2 and 3).

Table 3. Results of the leading digit distribution analysis.

Territory	OM	GHSI	n	d^*	m	w	p-Value	χ^2
Albania	4	53	143	0.10	0.840	4.696	0.879	3.750
Algeria	4	24	145	0.09	0.923	6.821	0.710	5.435
Argentina	5	59	144	0.15	1.030	6.736	0.706	5.473
Armenia	4	50	143	0.06	0.362	1.657	0.947	2.786
Australia	5	76	148	0.12	0.745	3.894	0.168	11.632
Austria	5	59	145	0.11	1.026	3.495	0.316	9.317
Azerbaijan	4	34	144	0.07	0.499	1.995	0.998	1.024
Bahrain	4	39	145	0.09	0.624	4.924	0.647	6.003
Bangladesh	4	35	142	0.13	1.112	4.109	0.141	12.232
Belarus	4	35	144	0.29	2.612	40.873	0.000	59.001
Belgium	5	61	145	0.09	0.591	6.106	0.322	9.250
Bolivia	4	36	142	0.13	1.070	4.933	0.609	6.346
Bosnia & Herzegovina	4	43	142	0.06	0.462	3.190	0.991	1.616
Botswana	4	31	130	0.07	0.479	3.956	0.959	2.559
Brazil	6	60	144	0.12	1.054	6.304	0.410	8.243
Bulgaria	4	46	142	0.07	0.483	3.088	0.989	1.691
Cameroon	4	34	109	0.10	0.748	3.175	1.000	0.643
Canada	4	75	151	0.19	1.997	13.253	0.009	20.402
Chile	4	58	145	0.27	3.019	30.272	0.000	43.576
China	4	48	151	0.12	0.929	4.896	0.879	3.751
Colombia	4	44	143	0.26	2.596	22.392	0.000	44.757
Costa Rica	4	45	143	0.12	1.091	5.205	0.663	5.860
Croatia	4	53	145	0.11	0.958	5.210	0.825	4.341
Cuba	4	35	141	0.16	1.478	8.831	0.038	16.344
Cyprus	4	43	142	0.09	0.776	7.599	0.954	2.658
Czechia	4		144	0.15	1.470	12.275	0.487	7.466
Denmark	5	70	149	0.13	1.299	5.603	0.248	10.253
Dominica	4		282	0.13	1.733	9.982	0.071	14.420
Dominican Republic	4	38	144	0.28	2.362	32.039	0.000	46.674
Ecuador	4	50	143	0.18	1.844	11.292	0.006	21.481
El Salvador	4	44	140	0.12	0.900	4.885	0.091	13.679
Estonia	4	57	149	0.09	0.718	2.926	0.959	2.554
Finland	4	69	150	0.13	1.049	8.662	0.133	12.421
France	5	68	151	0.18	1.502	10.589	0.204	10.964
Georgia	4	52	144	0.09	0.696	4.309	0.161	11.780
Germany	5	66	150	0.13	1.237	5.086	0.440	7.938
Greece	4	54	145	0.08	0.718	2.872	0.942	2.877
Guatemala	4	33	146	0.18	1.899	11.984	0.026	17.380
Honduras	4	28	141	0.19	1.867	19.798	0.001	27.624
Hungary	4	54	143	0.07	0.645	7.163	0.972	2.248
India	5	47	150	0.10	1.049	3.733	0.345	8.972
Indonesia	5	57	143	0.08	0.589	2.722	0.989	1.693
Iran	5	38	146	0.26	2.810	26.226	0.000	37.730
Iraq	4	26	145	0.33	3.766	47.118	0.000	67.430
Ireland	4	59	143	0.07	0.454	3.076	0.997	1.155
Israel	5	47	146	0.06	0.375	2.427	0.887	3.648
Italy	5	56	150	0.12	1.237	5.086	0.504	7.307
Japan	5	60	151	0.09	0.721	4.166	0.551	6.865
Jordan	4	42	143	0.10	0.909	6.609	0.089	13.720

Kazakhstan	4	41	140	0.12	0.816	6.530	0.135	12.389
Kuwait	4	46	145	0.22	1.964	17.655	0.000	29.546
Kyrgyzstan	4	49	140	0.23	2.270	17.113	0.000	37.253
Latvia	4	63	144	0.15	1.554	8.025	0.156	11.888
Lebanon	4	43	145	0.14	1.124	15.978	0.008	20.827
Libya	4	26	143	0.21	1.760	11.234	0.000	33.983
Lithuania	4	55	144	0.11	0.946	2.971	0.622	6.227
Malaysia	4	62	151	0.11	0.929	5.295	0.639	6.073
Mexico	5	58	156	0.12	1.031	10.977	0.030	17.024
Mongolia	4	50	142	0.05	0.481	4.530	0.954	2.658
Morocco	4	44	148	0.11	0.909	4.695	0.069	14.520
Myanmar	4	43	138	0.10	0.983	3.207	0.632	6.139
Nepal	4	35	151	0.07	0.607	1.368	0.987	1.801
Netherlands	5	76	144	0.05	0.304	1.595	1.000	0.057
Norway	4	65	145	0.16	1.275	8.692	0.436	7.972
Pakistan	4	36	145	0.25	1.856	23.045	0.000	65.686
Palestine	4	22	143	0.11	0.840	4.783	0.091	13.649
Panama	4	44	142	0.17	1.573	9.909	0.032	16.859
Paraguay	4	36	142	0.17	1.489	11.460	0.115	12.918
Peru	5	49	143	0.15	1.585	8.344	0.153	11.969
Philippines	4	48	150	0.14	1.001	8.142	0.007	21.077
Poland	5	55	143	0.12	1.007	4.669	0.619	6.253
Portugal	4	60	144	0.09	0.780	4.195	0.553	6.847
Romania	4	46	144	0.09	0.971	3.132	0.809	4.509
Russia	5	44	150	0.21	1.482	26.752	0.000	55.319
Saudi Arabia	4	49	143	0.14	1.426	7.163	0.079	14.125
Serbia	4	52	144	0.17	1.638	8.909	0.008	20.830
Singapore	4	59	151	0.06	0.455	1.612	1.000	0.026
Slovakia	4	48	143	0.10	0.832	3.001	0.839	4.195
Slovenia	4	67	149	0.11	0.889	5.792	0.372	8.661
South Africa	4	55	148	0.15	1.310	9.748	0.332	9.125
South Korea	5	70	151	0.13	1.014	7.805	0.565	6.740
Spain	5	66	149	0.03	0.234	0.485	0.995	1.359
Sri Lanka	4	34	150	0.09	0.593	2.874	0.453	7.806
Sweden	5	72	147	0.06	0.456	3.581	1.000	0.276
Switzerland	5	67	145	0.08	0.635	3.435	0.983	1.922
Thailand	4	73	155	0.13	1.181	7.923	0.184	11.328
Tunisia	4	34	143	0.08	0.673	1.932	0.962	2.486
Turkey	5	52	142	0.18	1.511	12.965	0.042	16.042
Uganda	4	44	138	0.06	0.347	1.531	1.000	0.477
Ukraine	5	38	142	0.10	0.650	4.651	0.924	3.160
United Arab Emirates	4	47	148	0.09	0.909	4.695	0.826	4.331
United Kingdom	5	78	150	0.09	0.804	2.538	0.721	5.338
United States	6	84	150	0.08	0.838	5.616	0.997	1.122
Uruguay	4	41	141	0.07	0.552	4.560	0.999	0.801
Venezuela	4	23	141	0.16	1.132	10.516	0.239	10.378
Vietnam	5	49	144	0.06	0.446	1.824	1.000	0.421
Zambia	4	29	139	0.06	0.437	3.385	0.999	0.940
Zimbabwe	4	38	139	0.09	0.804	3.669	0.858	3.994

OM: The oder of magnitude; GHSI: Global Health Security Index; n: Sample size; d^*: Euclidean test (d^* or d-factor); m: Chebyshev Distance Test or M-statistic; w: Weighted Maximum Statistic; χ^2: Chi-square test.

The territorial score of BL conformity is constructed as the logarithmic sum of all goodness of fit results. The logarithmic sum made the scores more comparable and removed the effect of outliers in ranking (see Equation (6)). This approach allows plotting the BL results via a global map, cf. Figure 1. Higher scores intensify the color of the countries. Territories that are not highlighted were intentionally excluded from the analysis.

$$S = \ln(\sum_{i=1}^{n}(M_i, w_i, \chi_i, d_i^*) \quad (6)$$

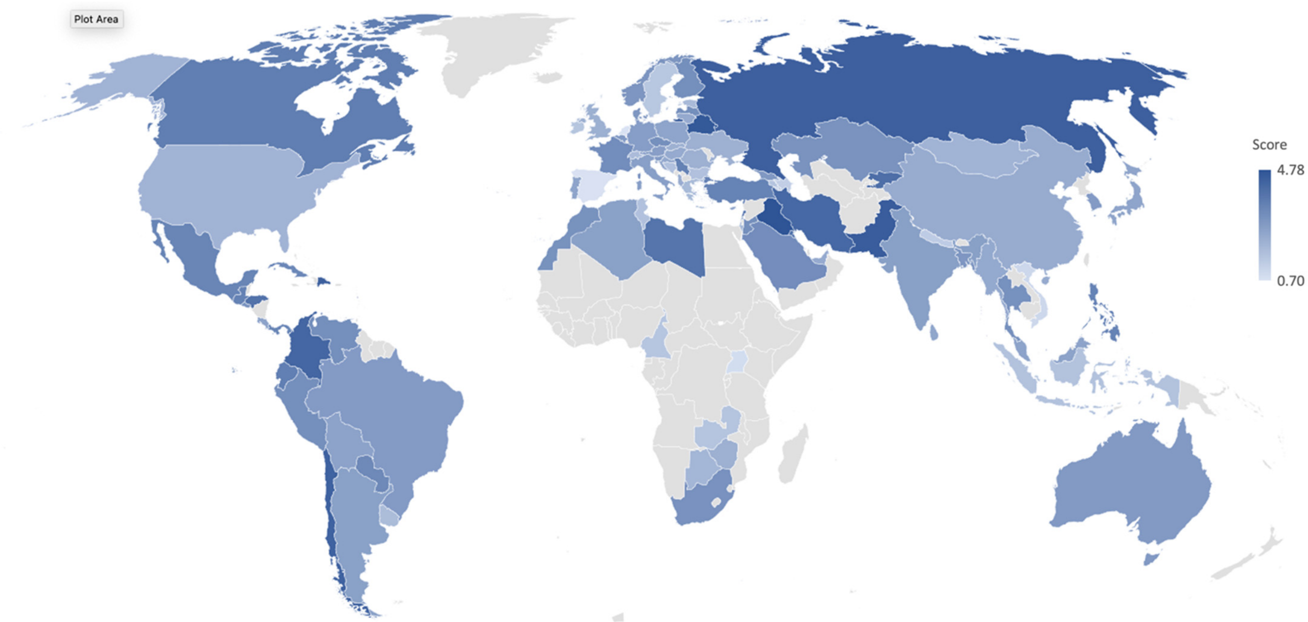

Figure 1. COVID-19 data conformity to BL worldwide.

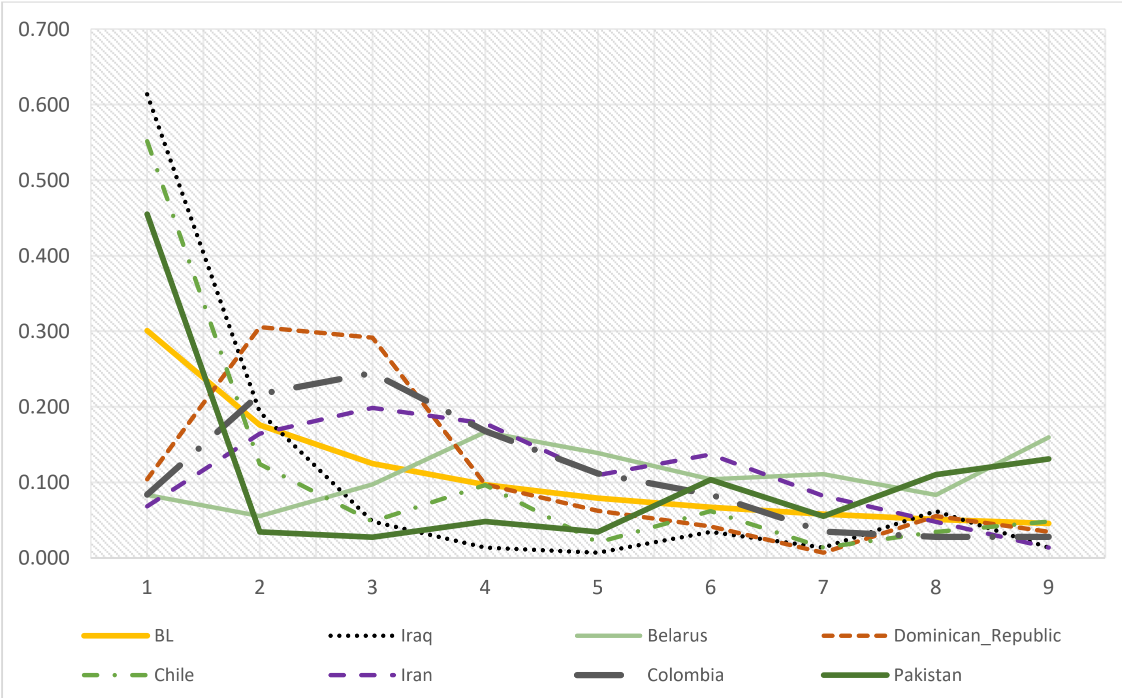

Figure 2. Countries with the lowest level of compliance with BL.

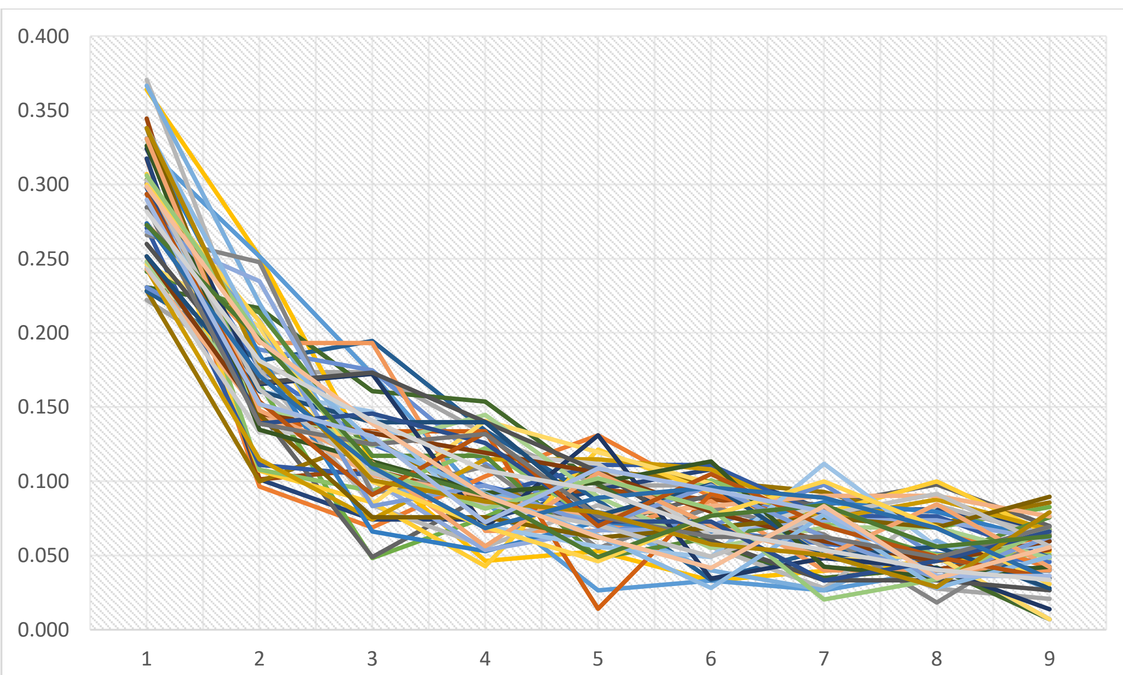

Figure 3. All countries with acceptable compliance with BL.

4. Conclusions

Two years after the COVID-19 outbreak, can we rely on the pandemic data published by countries worldwide? It is commonly accepted that the number of cases by country of almost any type of infectious diseases reported to the WHO follow Benford's law. Researchers have conducted multiple studies worldwide, addressing the conformality of pandemic data to Benford's law. Today, we better understand pandemic growth and the emergent data from territories worldwide. Societies and politicians initiate local countermeasures against the epidemics and mobilize policies based on evidence-based data. However, can the world rely on the COVID-19 data?

To answer this crucial question, researchers recurrently applied multiple statistical tests to examine respective data quality. In our eyes, these papers are inconsistent and, to some extent, misleading. Rudimentary application of cumulative incidents of new cases and new deaths, problematic use of goodness-of-fit tests, and lack of attention to the detailed data quality may have created contradictory results.

We collected pandemic data disclosed by 201 countries. To improve the accuracy of the measurements, our sample was focused on cases that showed a significant order of magnitude of at least four. We operationalized four statistical tests. In prior research, Chi-square, M-statistics, and d-factor were commonly used [2,4,11]. This study confirmed that up to 48 countries did not adhere to BL after two years. For the most part, our outcomes are in agreement with prior research [3–10]. The United States, United Kingdom, Australia, and most European countries demonstrated BL conformity by satisfying at least two goodness-of-fit tests. Spain and Israel overwhelmingly revealed compliance with the law.

Once countries experience rapid growth of new COVID-19 cases, they tend to demonstrate shrinking conformity to the common law of leading digits. This behavior was extensively studied, empirically tested, and reported in earlier research [9,11]. BL conformity works well for data originating from phases of progressive growth along the logistic curve [6,9,13].

The most significant irregularities occurred in Iran, Belarus, Chile, and Iraq. These are socially and economically distressed countries under challenging public health conditions and therefore demonstrate the poor conformity of pandemic data to BL. These findings affirm that the leading violators of the BL in previous studies exhibited similar behaviors. In previous studies, Belarus and Iran—the leaders in BL violations—showed the greatest distance to BL frequencies. COVID-19 has exacerbated existing and, in some cases, deep-rooted political, economic, social, and security problems in these countries.

In early 2020, Belarusian President Alexander Lukashenko denied the threat of Coronavirus in his country [32]. He called on people to go to work in the fields and ride tractors in order to get rid of the infectious disease. In the case of Iran, the British news channel BBC reported in August 2021 that the number of deaths and new cases in Iran was almost triple and double the official figures [33]. In 2020, the Washington Post released satellite images of mass graves for Coronavirus casualties in Qom, the main hub of the outbreak in Iran [34]. At the same time, WHO confirmed that the actual figures on the COVID-19 spread out of Iran must be five times higher than the official numbers made public by the government [35].

Johns Hopkins University (JHU) studied national health care systems and policies in the context of epidemics [36]. In an extensive study conducted for the first time in 2019—the *Global Health Security Index* (GHSI)—JHU identified the highest scores for "early detection and reporting for epidemics of potential international concern" for the countries with significant BL conformity, such as the US (ranked 1), Australia (ranked 2), the UK (ranked 6), Germany (ranked 10), or Spain (ranked 11). Later in 2021, JHU shared its 2021 GHSI report. In this context, we replicated Farhadi's and Farhadi and Lahooti's approach [3,11] based on 2019 and 2021 GHSIs. In a further effort, we operationalized Pearson's product–moment correlation analysis to explore potential relationships between the GHSI scores and the goodness-of-fit tests applied here. A preliminary examination of statistics was conducted to ensure no violation of normality, linearity, and homoscedasticity assumptions occurred. We identified negative correlations between the GHSI scores and Chi-square statistics (r: −0.27, n: 98, $p < 0.001$) in 2019 and (r: −0.23, n: 98, $p < 0.001$) in 2021, which suggest moderate and weak relationships among the variables respectively.

Disobeying BL does not indicate fraud by default [3–11]. Irregularities with BL show the distance between observed and expected leading digits' frequencies only. This may pertain to varying national public health policies and boundaries in capacity management or poor reporting. The Coronavirus mainly affects frail or immunocompromised people. The situation can worsen when new variants of the deadly virus accelerate to spread worldwide. For instance, the new variant, Omicron, affects the elderly, adolescents, and even children [37]. While the new variant has a very high transmission rate, it does not cause severe diseases like the previous variants. Young people infected with Omicron may develop mild flu-like symptoms, and after a few days, they recover from the illness. Perhaps, they do not bother to report positive test results from rapid antigen tests to health authorities. This may have a hidden impact on incident reporting, especially in territories severely affected by the virus. We recommend establishing global governance over pandemic data to effectively address these issues. Access to reliable data is vital in the fight against submicroscopic infectious organisms.

5. Future Research

Our study found substantial problems with the quality of prior BL research into COVID-19 worldwide. It is of particular importance to evaluate these studies and define the essential criteria for Benfordness analysis. To address the national issues with BL accurately, one shall further examine countries' economic, political, and social issues. In the case of the red-flagged countries, we still need to understand the domestic healthcare processes and epidemic reporting policies. The authors recommend replicating the BL assessment in a later re-examination.

6. Limitation

The data collected in our study are affected by diverging public health systems and policies. Lack of common and shared practices in reporting epidemic incidents globally, particularly in developing countries, might have affected the quality of COVID-19 reports and thus caused noncompliance with BL. Furthermore, the emergence of new variants of Coronavirus, such as Omicron, in connection with extensive vaccination programs, may have also affected the pandemic growth and the BL conformity, testing processes, and policies. Although we know that Benfordness decreases as the growth factors of daily incidence flatten, the phenomenon is still poorly understood.

Author Contributions: Conceptualization, N.F., H.L.; methodology, N.F., H.L.; software, N.F.; validation, H.L.; formal analysis, N.F., H.L.; investigation, N.F., H.L.; resources, N.F., H.L.; data curation, N.F., H.L.; writing—original draft preparation, N.F.; writing—review and editing, N.F., H.L.; visualization, N.F., H.L.; supervision, N.F., H.L.; project administration, N.F., H.L.; funding acquisition, not applicable. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Our data on confirmed cases come from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (see https://systems.jhu.edu, last accessed on 31 January 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. *Coronavirus Disease (COVID-19) Outbreak*; WHO: Geneva, Switzerland, 2020. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 13 June 2020).
2. Farhadi, N. *Cross-Industry Ecosystems*; Springer: Wiesbaden, Germany, 2019. https://doi.org/10.1007/978-3-658-26129-0.
3. Farhadi, N. Can we rely on COVID-19 data? An assessment of data from over 200 countries. *Sci. Prog.* **2021**, *104*, 00368504211021232.

4. Newcomb, S. Note on the Frequency of Use of the Different Digits in Natural 242 Numbers. *Am. J. Math.* **1881**, *4*, 39–40.
5. Benford, F. The Law of Anomalous Numbers. *Proc. Am. Philos. Soc.* **1938**, *78*, 551–572.
6. Sambridge, M.; Jackson, A. National COVID Numbers—Benford's Law Looks for Errors. Nature [Internet]. 2020. Available online: https://www.nature.com/articles/d41586-020-01565-5 (accessed on 31 January 2022).
7. Koch, C.; Okamura, K. Benford's Law and COVID-19 Reporting. 2020. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3586413 (accessed on 31 January 2022).
8. Idrovo, A.J.; Manrique-Hernndez, E.F. Data Quality of Chinese Surveillance of 270 COVID-19: Objective Analysis Based on WHO's Situation Reports. *Asia Pac. J. Public Health* **2020**, *32*, 165–167.
9. Lee, K.-B.; Han, S.; Jeong, Y. COVID-19 flattening the curve, and Benford's law. *Phys. A Stat. Mech. Appl.* **2020**, *559*, 125090. https://doi.org/10.1016/j.physa.2020.125090.
10. Isea, R. How Valid are the Reported Cases of People Infected with COVID-19 in the World? *Int. J. Coronaviruses* **2020**, *1*, 53–56.
11. Farhadi, N.; Lahooti, H. Are COVID-19 Data Reliable? A Quantitative Analysis of Pandemic Data from 182 Countries. *COVID* **2021**, *1*, 137–152. https://doi.org/10.3390/covid1010013.
12. Morillas-Jurado, F.G.; Caballer-Tarazona, M.; Caballer-Tarazona, V. Applying Benford's Law to Monitor Death Registration Data: A Management Tool for the COVID-19 Pandemic. *Mathematics* **2022**, *10*, 46. https://doi.org/10.3390/math10010046.
13. Farhadi, N.; Lahooti, H. Pandemic Growth and Benfordness: Empirical Evidence from 176 Countries Worldwide. *COVID* **2021**, *1*, 366–383. https://doi.org/10.3390/covid1010031.
14. Balashov, V.S.; Yan, Y.; Zhu, X. Using the Newcomb–Benford law to study the association between a country's COVID-19 reporting accuracy and its development. *Sci. Rep.* **2021**, *11*, 22914. https://doi.org/10.1038/s41598-021-02367-z.
15. Kilani, A. Authoritarian regimes' propensity to manipulate Covid-19 data: A statistical analysis using Benford's Law. *Commonw. Comp. Polit.* **2021**, *59*, 319–333. https://doi.org/10.1080/14662043.2021.1916207.
16. Ley, L. On the Peculiar Distribution of the U.S. Stock Indexes' Digits. *Am. Stat.* **1996**, *50*, 311–313. https://doi.org/10.2307/2684926.
17. Kossovsky, A. *Benford's Law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications*; World Scientific: Singapore, 2015.
18. Silva, L.; Filho, D.F. Using Benford's law to assess the quality of COVID-19 register data in Brazil. *J. Public Health* **2022**, *10*, 107–110.
19. Nigrini, M.J. *Benford's Law Applications for Forensic Accounting, Auditing and Fraud Detection*; Wiley Corporate F&A; Wiley: Hoboken, NJ, USA, 2012.
20. Fairweather, W.R. Sensitivity and Specificity in the Application of Benford's Law to Explore for Potential Fraud. *J. Forensic Investig. Account.* **2017**, *9*, 953–961.
21. Durtschi, C.; Hillison, W.; Pacini, C. The Effective Use of Benford's law to Assist in Detecting Fraud in Accounting Data. *J. Forensic Account*. **2004**, *5*, 17–34.
22. Grammatikos, T.; Papanikolaou, N.I. Applying Benford's law to Detect Accounting 250 Data Manipulation in the Banking Industry. *J. Financ. Serv. Res*. **2020**, *59*, 115–142.
23. Cho, T.W.; Gaines, B.J. Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance. *Am. Stat.* **2007**, *61*, 218–223.
24. Roukema, B.F. A first-digit anomaly in the 2009 Iranian presidential election. *J. Appl. Stat*. **2014**, *41*, 164–199.
25. D'Alessandro, A. Benford's law and metabolomics: A tale of numbers and blood. *Transfus. Apher. Sci.* **2020**, *59*, 103019. https://doi.org/10.1016/j.transci.2020.103019.
26. Leemis, L.M. Reliability: Probabilistic Models and Statistical Methods; Prentice-Hall: New York, NY, USA, 1995.
27. Leemis, L.M.; Schmeiser, B.W.; Evans, D.L. Survival Distributions Satisfying Benford's Law. *Am. Stat*. **2000**, *54*, 236–241. https://doi.org/10.2307/2685773.
28. Rayner, J.C.W.; Best, D.J. *Smooth Tests of Goodness of Fit*; Oxford University Press: New York, NY, USA, 1989.
29. Dadakis Horn, S. Goodness-of-Fit Tests for Discrete Data: A Review and an Application to a Health Impairment Scale. *Biometrics* **1977**, *33*, 237–247. https://doi.org/10.2307/2529319.
30. Goodman, Q. The promises and pitfalls of Benford's law. *Significance* **2016**, *13*, 38–41. https://doi.org/10.1111/j.1740-9713.2016.00919.x.
31. Morrow, J. Benford's Law, Families of Distributions and a Test Basis. 2014. Centre for Economic Performance Discussion Paper No 1291, London. Available online: http://johnmorrow.info/pdf/benford.pdf (accessed on 2 February 2022).
32. Vector, D. What's Happening in Belarus? Here Are the Basics. *New York Times*, 26. May 2021. Available online: https://www.nytimes.com/2021/05/26/world/europe/whats-happening-in-belarus.html (accessed on 23 January 2022).
33. BBC. Coronavirus: Iran Cover-Up of Deaths Revealed by Data Leak. Available online: https://www.bbc.com/news/world-middle-east-53598965 (accessed on 15 August 2021).
34. Cunningham, E.; Bennett, D. Coronavirus Burial Pits So Vast They're Visible from Space. 2020. Available online: https://www.washingtonpost.com/graphics/2020/world/iran-coronavirus-outbreak-graves (accessed on 25 February 2022).

35. Farge, E. WHO to Start Coronavirus Testing in Rebel Syria; Iran Raises Efforts, Official Says. 16 March 2020. Available online: https://web.archive.org/web/20200316212446if_/https://www.reuters.com/article/us-health-coronavirus-mideast/who-to-start-coronavirus-testing-in-rebel-syria-iran-raises-efforts-official-says-idUSKBN2133PK (accessed on 16 March 2020).
36. Johns Hopkins University. Global Health Security Index. Available online: https://www.ghsindex.org/ (accessed on 25 January 2022).
37. CDC. SARS-CoV-2 Variant Classifications and Definitions. 1 December 2021. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html (accessed on 28 January 2020).