

# Data-Forensic Determination of the Accuracy of International COVID-19 Reporting: Using Zipf's Law for Pandemic Investigation

Aamo Iorliam<sup>1</sup>, Anthony T S Ho<sup>2</sup>, Santosh Tirunagari<sup>3</sup>  
David Windridge<sup>4</sup>, Adekunle Adeyelu<sup>5</sup>, Samera Otor<sup>6</sup>, Beatrice O. Akumba<sup>7</sup>  
Department of Mathematics and Computer Science, BSU, Makurdi, Nigeria<sup>1, 5, 6, 7</sup>  
Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey, United Kingdom GU2 7XH<sup>2, 4</sup>  
Tianjin University of Science and Technology, China<sup>2</sup>  
Wuhan University of Technology, China<sup>2</sup>  
Department of Computer Science, Middlesex University, London, UK<sup>3, 4</sup>

**Abstract**—Severe outbreaks of infectious disease occur throughout the world with some reaching the level of international pandemic: Coronavirus (COVID-19) is the most recent to do so. In this paper, a mechanism is set out using Zipf's law to establish the accuracy of international reporting of COVID-19 cases via a determination of whether an individual country's COVID-19 reporting follows a power-law for confirmed, recovered, and death cases of COVID-19. The probability of Zipf's law (P-values) for COVID-19 confirmed cases show that Uzbekistan has the highest P-value of 0.940, followed by Belize (0.929), and Qatar (0.897). For COVID-19 recovered cases, Iraq had the highest P-value of 0.901, followed by New Zealand (0.888), and Austria (0.884). Furthermore, for COVID-19 death cases, Bosnia and Herzegovina had the highest P-value of 0.874, followed by Lithuania (0.843), and Morocco (0.825). China, where the COVID-19 pandemic began, is a significant outlier in recording P-values lower than 0.1 for the confirmed, recovered, and death cases. This raises important questions, not only for China, but also any country whose data exhibits P-values below this threshold. The main application of this work is to serve as an early warning for World Health Organization (WHO) and other health regulatory bodies to perform more investigations in countries where COVID-19 datasets deviate significantly from Zipf's law. To this end, this paper provide a tool for illustrating Zipf's law P-values on a global map in order to convey the geographic distribution of reporting anomalies.

**Keywords**—COVID-19; power-law; pandemic; Zipf's Law; WHO

## I. INTRODUCTION

The Coronavirus Disease 2019 (COVID-19) broke out in 2019 in Wuhan, China and has caused alarming health crises, unemployment, unimaginable hunger, lockdown of the entire world, constant fear and in some cases death [1]. Up till now, several countries are accused of wrongly reporting the COVID-19 confirmed, recovered, and death cases globally [2,3].

Determining the accuracy of reporting COVID-19 cases internationally is important because it will assist health regulatory bodies to perform in depth investigations into country's suspected data [2]. Moreover, this has a potential to

aid these regulatory bodies in their planning, measures to curb this deadly pandemic, and save more lives [2,3].

The Zipf's law has proved over the years to be very effective in differentiating forged/fabricated data from authentic/original data [4]. Motivated by the capability of the Zipf's law, this paper investigates the COVID-19 datasets using the Zipf's law with the objective of determining the accuracy of international COVID-19 reporting. Any consistency of the COVID-19 datasets to the Zipf's law indicates that such data is reliable, whereas a deviation from the Zipf's law may indicate a wrong reporting of COVID-19 cases.

Results in this paper showed that most countries COVID-19 datasets followed the Zipf's law, whereas some countries COVID-19 datasets deviated from the Zipf's law. The rest of the paper is organized as follows. Related works are described in Section II. Section III explains the experimental setup. Results are presented in Section IV. Section V discusses the results. Conclusions and future work are presented in Section VI.

## II. RELATED WORKS

### A. Historical Perspective on Covid-19 Pandemic Investigations

The first recorded pandemic was in 165 AD to 180 AD. This pandemic was referred to as the Antonine Plague (also known as the plague of Galen) and resulted in about 5 million deaths across the globe. Analysis of symptomology and infection pattern suggest that this was likely smallpox or measles [5].

In around the 735 AD – 737AD, the Japanese smallpox epidemic erupted (believed to be a variola virus), killing up to 1 million persons [6]. Later, around 541 AD – 542 AD, the Plague of Justina killed between 30 – 50 million persons, believed to be the world's first bubonic plague [7]. Procopius described the plague as that “by which the whole human race was near to being annihilated [8 - 9].”

The most devastating pandemic, in terms of its impact of the global population, occurred between 1347 AD and 1352

AD; this is the pandemic referred to as 'The Black Death', which claimed between 75 – 200 million lives. It is believed to have been caused by the bubonic plague [10]; Benedictow in [11] described this plague as “the greatest catastrophe ever”; Michael of Piazza, a Franciscan friar wrote contemporaneously that: “the infection spread to everyone who had any intercourse with the disease” [12].

It is recorded that around 1520 AD there was an outbreak of the new world smallpox, believed to be a Variola virus, resulting in 25 to 55 million deaths. The New world smallpox caused so much damage that Noble David Cook [13-14] estimated that “in the end, the regions least affected lost 80 percent of their populations; those most affected lost their full populations, and a typical society lost 90 percent of its population.”

Around 1629-1631 AD, the Italian Plague erupted, believed to originate from Yersinia Pestis bacteria in rats/fleas. It claimed up to 1 million lives [15-16].

Around 1665 AD, the great plague of London claimed 75,000 to 100,000 lives, also believed to have its source from rats and fleas [17].

From 1817 to 1923 the Cholera Pandemic (caused by V. Cholera bacteria) killed more than 1 million people [18] in Europe. Around 1885, a third plague caused by Yersinia Pestis bacteria carried by rats and fleas resulted in around 12 million deaths in China and India [19]. Also in the late 1800s Yellow fever, its source is believed to be viruses/mosquitoes, resulted in more than 150,000 deaths. It targeted mostly South America and sub-Saharan Africa [20-21].

The Russian Influenza outbreak of around 1889 to 1890, transmitted via the H2N2 virus, claimed around 1 million lives [22-23]. The Spanish flu of 1918 to 1919, its source believed to be H1N1 virus, claimed about 40 to 50 million lives [22-23], by far the most deadly influenza pandemic. Asian flu, believed to be transmitted by the H2N2 virus, claimed around 2 million lives in 1968 – 1970 [23], while the 1968-1970 Hong Kong Flu, transmitted via the H3N2 virus claimed up to 1 million lives [24].

The HIV/AIDS epidemic, which commenced around 1981 has so far claimed 25 to 30 million lives [25-26]. From 2002 to 2003, Severe Acute Respiratory Syndrome (SARS) resulting from Coronavirus in Bats and Civets killed up to 770 people [27]. Later, between 2009 and 2010, Swine flu, its source believed to be the H1N1 virus in pigs, killed about 200,000 people [23].

Around 2012 – 2016, an outbreak of the Ebola virus disease (EVD) killed up to 11,000 persons; it is believed to have arisen from Ebola virus in wild animal. EVD was itself recorded as far back as 1976 [28-29].

From 2015 till the present time, the Middle East Respiratory Syndrome (MERS) believed to be caused by MERS corona-virus (MERS-CoV) in wild animals has so far killed 850 persons [30].

COVID-19 [1], the subject of the current study, erupted in 2019; according to [31], COVID-19 has claimed 146,201 lives as of 17/ 04/2020.

As is clear from the above historical account of pandemic spread, the potential for negative global impact is very substantial indeed if unchecked. In the majority of the above cases, the reporting and compilation of pandemic statistics was substantially after the fact (sometimes by many centuries) given the limited contemporaneous statistical capabilities. In the absence of such statistics, compiled while the outbreak was still live, it would have been very difficult or impossible for authorities to make well-informed policy decisions in order to combat the pandemic spread.

It is therefore critical in the current COVID-19 pandemic that accurate compilation of international reporting is undertaken. However, given the potential for countries/individuals to falsify records, for political, offensive or financial purposes, it is necessary to have methods in place to distinguish authentic from forged records. In this paper, the Zipf's law is proposed as means to achieve this.

### B. Motivation for the use of Zipf's Law

Zipf's law was proposed in 1935 by the US linguist George K Zipf [32] and may be stated succinctly as follows: given some corpus of natural language utterances, the frequency of any given word is inversely proportional to its rank in a frequency table.

Newman [33] made this explicitly stochastic; when considering the probability of measuring a particular quantity (in this case, COVID-19 cases), and it is found that the quantity varies inversely as a power of that value, then the quantity may be said to follow Zipf's law [33]. Mathematically:

$$p(x) = cx^{-\alpha} \quad (1)$$

where  $p(x)$  is the distribution of the quantity  $x$ ,  $\alpha$  is the Zipf's law exponent and  $C$  is a constant [33].

Zipf's law is hence a Power law with small occurrences relatively common and large occurrences very uncommon. The null hypothesis in applying Zipf's law is hence that natural phenomena should follow a Power law and un-natural (or tampered-with) phenomena should deviate from this law [4,34]. Power laws have been applied to human language [4, 33], the city populations [4,33], intensity of earthquakes [4, 33], sizes of power outages [4], ranks of people watching a particular TV station at a given time [4, 35], stock market indices [35], gene expression [35], chess openings [35], the arts [35], paper citations [35], family names [35], personal donations [35], keystroke dynamics [4], the scales of Influenza A (H1N1) and Avian Influenza (H7N9) outbreaks [35, 36] amongst others.

The gap in the related works is that to the best of the researchers' knowledge, the application of Zipf's law to COVID-19 cases has not been done as at April, 2020 when this research was conducted. Hence, this paper proposes the investigation of *reported* COVID-19 datasets using Zipf's law to establish veracity and accuracy, in particular because of presence of widespread allegations that countries may have hidden or systematically underreported the cases of COVID-19 [2,3].

Thus, the aim is to establish probability values (P-values) in relation to Zipf's law calculation for each country affected by COVID-19. Furthermore, the P-values of each country based on the Zipf's law calculation on a global map are presented. This is hence an ongoing work as more data is compiled throughout the current COVID-19 outbreak.

### III. EXPERIMENTS

The primary goal is to investigate internationally reported cases of COVID-19 in order to determine consistency with Zipf's law. A secondary goal is to calculate the P-value for Zipf's law on each country's COVID-19 datasets. Lastly, Zipf's law P-values on a global map to convey the geographic distribution of reporting anomalies is illustrated.

The Power-law package developed by Clauset, Shalizi, and Newman [37] is used to obtain P-values for reported cases of COVID-19 per country. As methodologically reported in [37] and evaluated in [4], experiments are carried out 1000 times on the COVID-19 datasets in order to obtain P-values in each case. The steps followed to test whether COVID-19 datasets follow a Power-law are set out in [4] and [37]. It should be noted that the P-values are generated using the Kolmogorov-Smirnov (KS) statistic goodness-of-fit test as specified in [4] and [37]. The standard COVID-19 datasets available at: <https://github.com/CSSEGISandData/COVID-19> are used. The data consists in rows representing countries and columns representing the number of COVID-19 cases per each day from 22/1/2020 to 6/4/2020.

The evaluated hypotheses are thus:

- 1) Is Zipf's law applicable to COVID-19 case data?
- 2) How is Zipf's law behavior characterized if applicable?

### IV. RESULTS

The P-values for all the countries/regions using COVID-19 confirmed cases based on the Power law calculation is performed. The complete P-values for the COVID-19 confirmed cases can be accessed at: <https://www.preprints.org/manuscript/202004.0531/v1>.

Furthermore, the distribution of P-values across countries/regions for COVID-19 confirmed cases is shown in Fig. 1. These P-values range from 0 to 1, with any country having a value of less than 0.1 deemed to deviate from Zipf's law.

Table I shows four countries'/regions' P-values to illustrate score dichotomy: Uzbekistan, Belize, Qatar have high P-values close to 1 while China has a P-value of 0 (to within the measured accuracy of the test). Countries/regions with the higher P-values indicate COVID-19 datasets in very close compliance with Zipf's law; China's reported statistics, along with those scoring similar P-values such as Australia, US, etc. are not possible to reconcile with Zipf's law.

Furthermore, P-values on all recovered cases of COVID-19 are calculated and are available at: <https://www.preprints.org/manuscript/202004.0531/v1>. Due to the number of recovered cases being small in some countries, Zipf's law is not statistically robust in such cases.

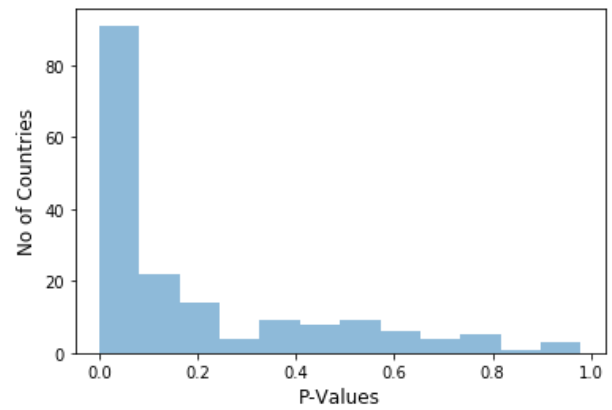


Fig. 1. Distribution of P-values across Countries / Regions for Confirmed Cases.

TABLE I. P-VALUES OF FOUR COUNTRIES / REGIONS FOR CONFIRMED CASES

Country	P-value
Uzbekistan	0.940
Belize	0.929
Qatar	0.897
China	0.000

The distribution of P-values across countries/regions for COVID-19 recovered cases are also illustrated in Fig. 2.

Again, Table II shows four external countries'/regions' P-values: Iraq, New Zealand, Austria and China (with China having an extremely low P-value of 0.002). It should however be noted that other countries such as Ghana, US, etc. have P-values of 0 which are lesser than China when considering P-values for recovered cases.

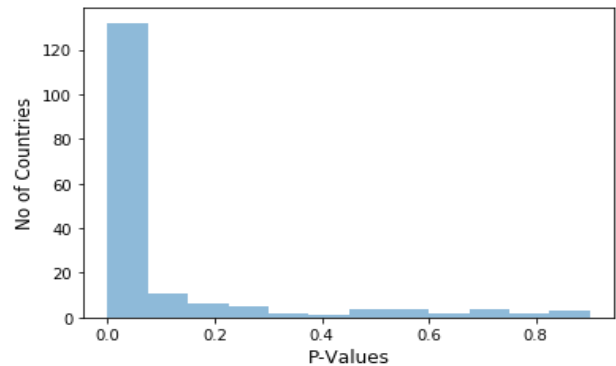


Fig. 2. Distribution of P-values across Countries / Regions for Recovered Cases.

TABLE II. P-VALUES OF FOUR COUNTRIES / REGIONS FOR RECOVERED CASES

Country	P-value
Iraq	0.901
New Zealand	0.888
Austria	0.884
China	0.002

The P-values of all death cases of COVID-19 across each country are also calculated and can be accessed at: <https://www.preprints.org/manuscript/202004.0531/v1>. In countries in which there are only a few recorded cases of COVID-19 deaths, power-law P-values are not presented.

The P-value distribution across countries/regions for COVID-19 death cases are indicated in Fig. 3.

Four extremal P-values are also shown in Table III; those for Bosnia and Herzegovina, Lithuania, Morocco and China; Bosnia and Herzegovina, Lithuania, Morocco have high P-values while China has a P-value of 0.000. Countries such as Costa Rica, Jordan, etc. have P-values of 0 similar to that of China.

Zipf's law power-law graph fits for each of the four countries identified in Tables I, II, and III are shown in Fig. 4, 5, and 6 for COVID-19 confirmed cases, recovered cases, and death cases, respectively.

Fig. 7a, b, and c indicate, on a global map, P-values per country for the COVID-19 confirmed cases, recovered cases, and death cases respectively. These maps are interactive and can also be viewed online for the COVID-19 confirmed cases<sup>1</sup>, recovered cases<sup>2</sup>, and death cases<sup>3</sup> showing the P-values and country names when hovered on it.

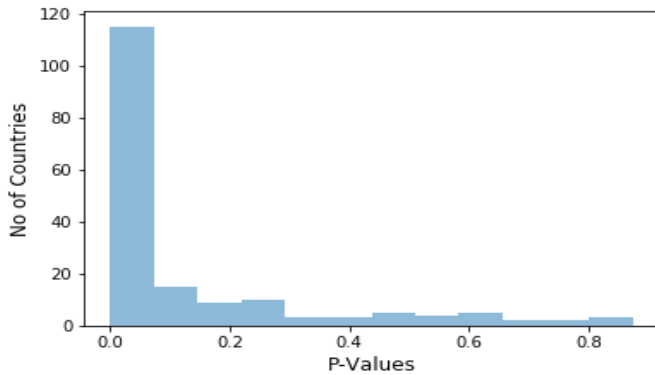


Fig. 3. Distribution of P-values across Countries / Regions for Death Cases.

TABLE III. P-VALUES OF FOUR COUNTRIES / REGIONS FOR DEATH CASES

Country	P-value
Bosnia and Herzegovina	0.874
Lithuania	0.843
Morocco	0.825
China	0.000

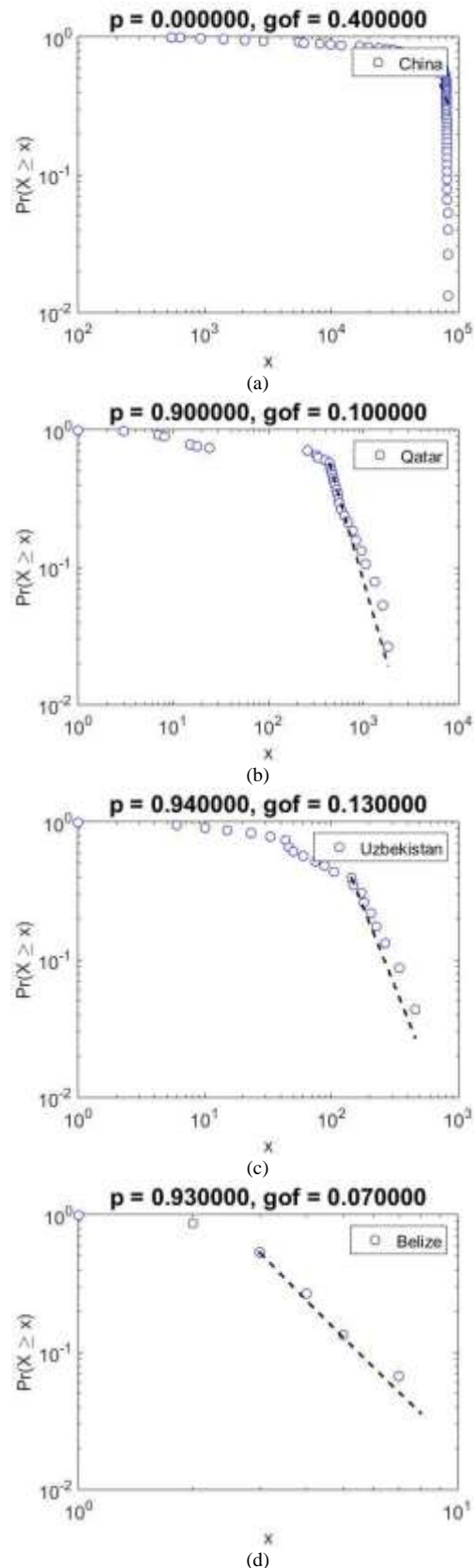


Fig. 4. Zipf's Law Power-Law Fits of COVID-19 Confirmed Cases for: (a) China, (b) Qatar, (c) Uzbekistan, (d) Belize.

<sup>1</sup> Confirmed Cases: <https://covid0.volitionlabs.xyz/zipfs/confirmed>

<sup>2</sup> Recovered Cases: <https://covid0.volitionlabs.xyz/zipfs/recovered>

<sup>3</sup> Death Cases: <https://covid0.volitionlabs.xyz/zipfs/death>

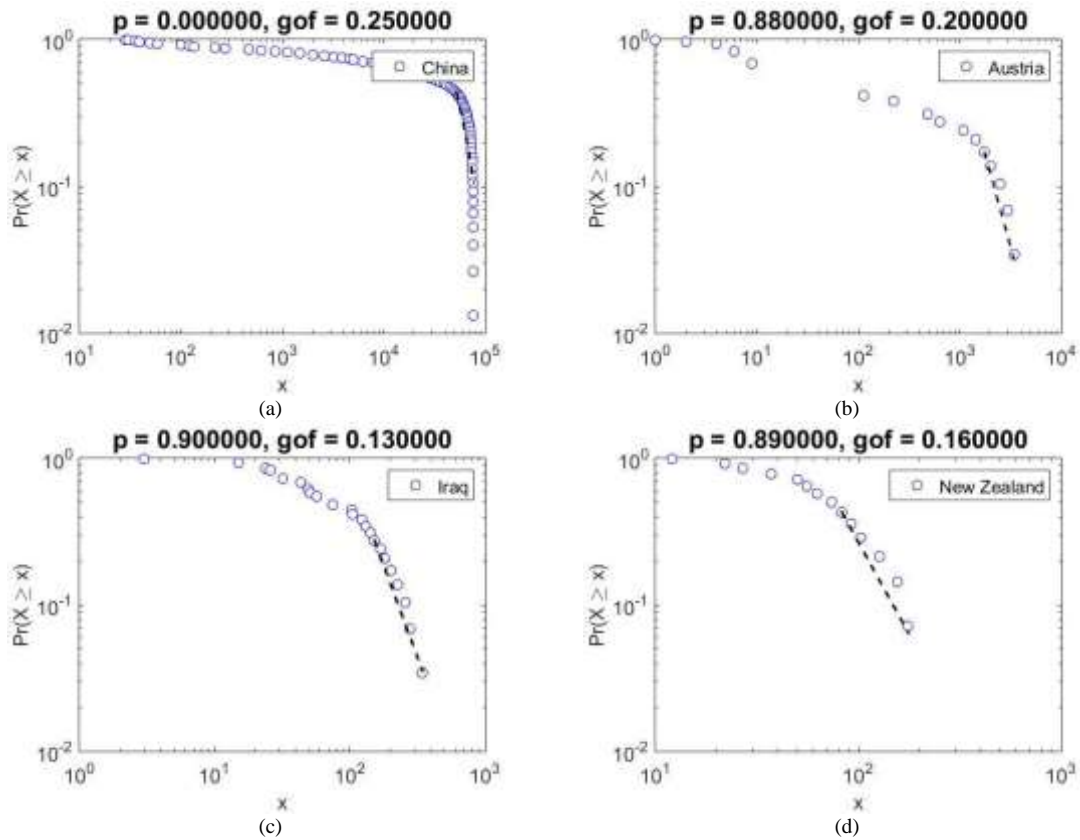


Fig. 5. Zipf's Law Power-Law Fits of COVID-19 Recovered Cases for: (a) China, (b) Austria, (c) Iraq, (d) New Zealand.

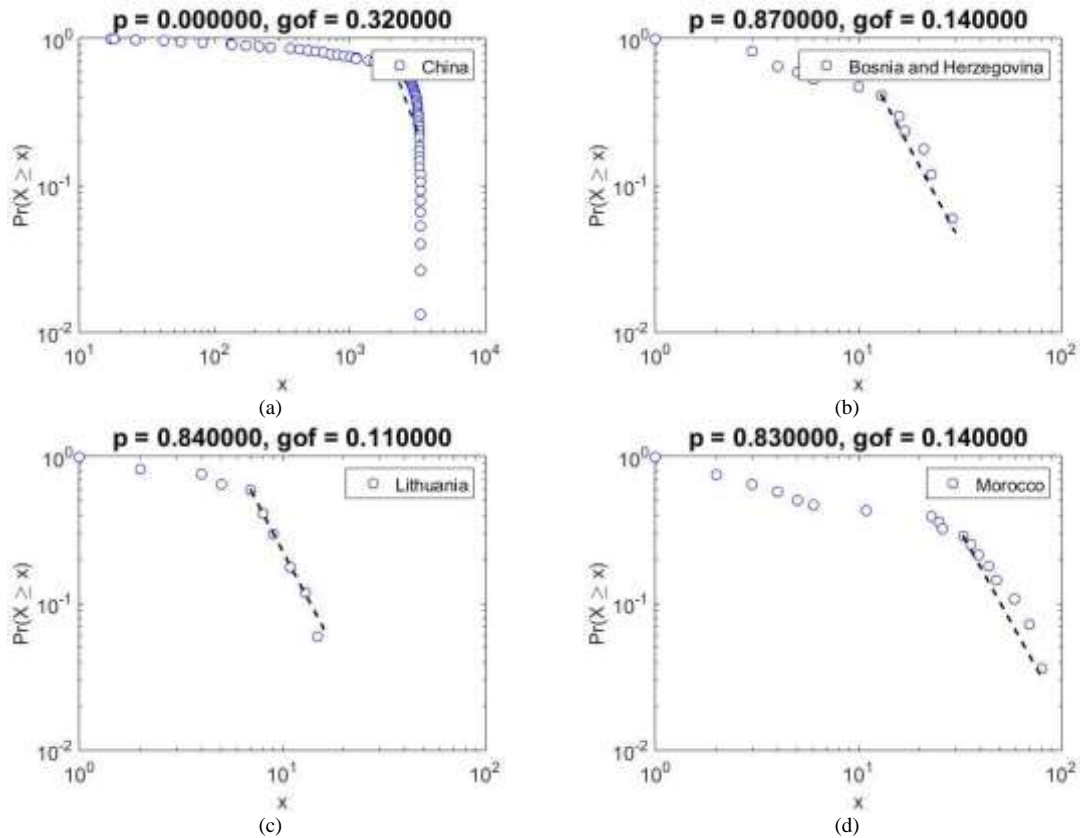


Fig. 6. Zipf's Law Power-Law Fits of COVID-19 Death Cases for (a) China, (b) Bosnia and Herzegovina, (c) Lithuania, (d) Morocco.



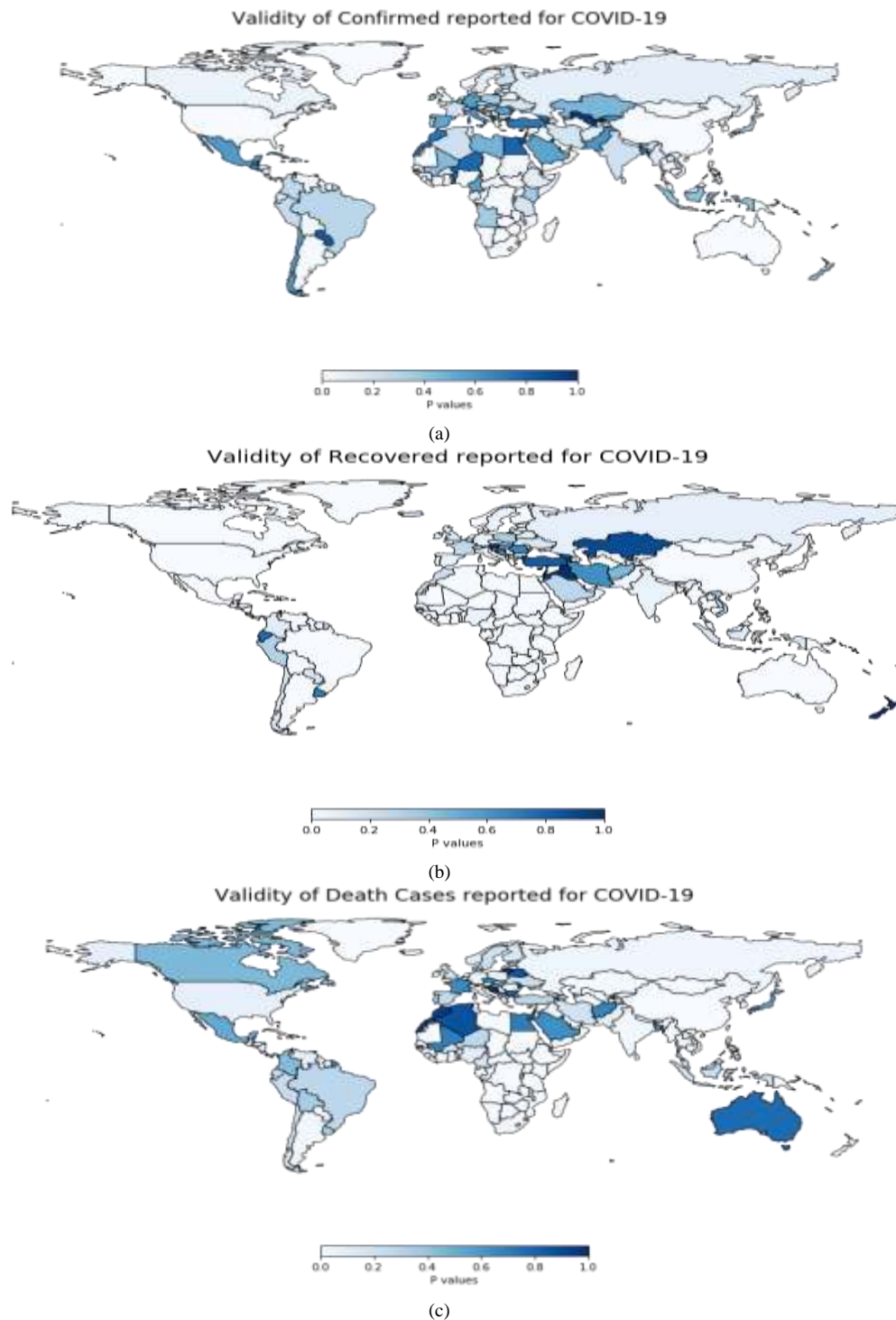


Fig. 7. Maps of: (a) COVID-19 Confirmed Cases, (b) COVID-19 Recovered Cases, (c) COVID-19 Death Cases.

## V. DISCUSSION

Experiments indicate that many countries' COVID-19 datasets follow a consistent power law for confirmed cases of COVID-19, recovered cases of COVID-19, and death cases of COVID-19. (Table I indicates that Uzbekistan had the highest P-value of 0.94, followed by Belize with a P-value of 0.929, and Qatar with a P-Value of 0.897; Table II indicates that

Iraq's recovered cases data most closely follows Zipf's law with a P-value of 0.901, followed by New Zealand with a P-value of 0.888, and Austria with a P-value of 0.884; Table III indicates that Bosnia and Herzegovina had the highest P-value of 0.874, followed by Lithuania with a P-value of 0.843, and Morocco with a P-value of 0.825).

As can be seen in Fig. 4, 5 and 6, the CDF represented as  $Pr(x)$  is plotted as a function of frequency ( $x$ ) for the COVID-

19 dataset for confirmed, recovered, and death cases. The straight line (black line) shows the section of the plots where the Power law provided a good model fit [35] of the data considered.

From Fig. 1 to 6 and Tables I to III, however, it is observed that several countries have P-values that are less than 0.1. Notably, China's confirmed cases, recovered cases, and death cases of COVID-19 all failed to follow Zipf's law despite an earlier commencement of data recording than those of other countries (being the pandemic origin).

This has raised some questions, not only for China, but also every other country whose power-law P-values are less than 0.1 (this threshold being the one selected to establish compliance with Zipf's law according to the reasoning in [4, 37]).

Based on the above discussion, the paper can conclude that:

- 1) Zipf's law can be applied to COVID-19 case data with reliability monotonically improving in relation to dataset size.
- 2) This analysis can potentially be used as an 'early warning system' for further investigation into COVID-19 datasets not consistent with Zipf's law.

## VI. CONCLUSIONS AND FUTURE WORK

In this paper, it is established that COVID-19 datasets for many countries can be shown to be consistent with Zipf's law. However, experiments also indicate that deviation of COVID-19 datasets from Zipf's law may be indicative of incorrect data reporting. The main application of this work is thus to serve as a potential early warning system for international health regulatory bodies such as the World Health Organization (WHO) in performing further investigations in countries where COVID-19 datasets have deviated from Zipf's law.

In future work, the plan is to:

- 1) Re-perform this experiment with complete statistics once the COVID-19 pandemic has ended.
- 2) Carry out experiments with other power laws variants (in particular Benford's law and Heap's law) both for COVID-19 and other pandemic datasets.

## REFERENCES

- [1] Z. Y. Zu, M. D. Jiang, P. P. Xu, W. Chen, Q. Q. Ni, G. M. Lu, and L. J. Zhang, "Coronavirus disease 2019 (COVID-19): a perspective from China." *Radiology*, 200490, 2020.
- [2] L. Winter, Data fog: Why some countries' coronavirus numbers do not add up, Accessed on April, 30, 2020. [Online]. Available: <https://www.aljazeera.com/features/2020/6/17/data-fog-why-some-countries-coronavirus-numbers-do-not-add-up>.
- [3] A. Lachmann, "Correcting under-reported COVID-19 case numbers." *medRxiv*, 2020.
- [4] A. Iorliam, A. T. Ho, N. Poh, S. Tirunagari, and P. Bours, "Data forensic techniques using Benford's law and Zipf's law for keystroke dynamics." In 3rd Int. Workshop on Biometrics and Forensics (IWBF 2015) pp. 1-6. IEEE, Mar., 2015.
- [5] R. P. Duncan-Jones, "The impact of the Antonine plague," *J. of Rom. Archeo.*, vol. 9, pp. 108-136, 1996.
- [6] A. M. Geddes, "The history of smallpox." *Clinics in dermatology*, vol. 24 no. 3, pp. 152-157, 2006.
- [7] F. M. Snowden, "Epidemics and Society: From the Black Death to the Present. Yale Uni. Press," 2019.
- [8] R. S. Bray, "Armies of pestilence: the impact of disease on history." James Clarke & Co., 2004.
- [9] R. Findlay, and M. Lundahl, "Demographic shocks and the factor proportions model: from the plague of Justinian to the Black Death." In *The Economics of the Frontier* (pp. 125-172). Palgrave Macmillan, London, 2017.
- [10] R. Horrox, "The black death." 2013.
- [11] O. J. Benedictow, "The Black Death: the greatest catastrophe ever." *History Today*, vol. 55, no. 3, pp. 42-49, 2005.
- [12] C. J. Duncan, and S. Scott, "What caused the black death?." *Postgraduate. Medical J.*, vol. 81, no. 955, pp. 315-320, 2005.
- [13] N. D. Cook, "Born to die: disease and New World conquest," pp. 1492-1650 vol. 1. Cambridge University Press, 1998.
- [14] N. Nunn, and N. Qian, "The Columbian exchange: A history of disease, food, and ideas." *J. of Econ. Perspective*, vol. 24, no. 2, pp. 163-88, 2010.
- [15] C. B. Vicentini, S. Manfredini, D. Mares, T. Bonacci, C. Scapoli, M. Chicca, and M. Pezzi, "Empirical 'integrated disease management' in Ferrara during the Italian plague" (1629-1631). *Parasitology Int.*, vol. 75, pp. 102046, 2020.
- [16] G. Alfani, and T. E. Murphy, "Plague and lethal epidemics in the pre-industrial world." *the J. of economic History*, vol. 77 no.1, pp. 314-343, 2017.
- [17] R. S. Roberts, "The place of plague in English history." *Proceedings of the Royal Society of Medicine*, 59(2), 101, 1966.
- [18] D. Barua, "History of cholera." In *Cholera* pp. 1-36. Springer, Boston, MA. 1992.
- [19] K. L. Kausrud, M. Begon, T. B. Ari, H. Viljugrein, J. Esper, U. Buntgen, and L. Xu, "Modeling the epidemiological history of plague in Central Asia: palaeoclimatic forcing on a disease system over the past millennium." *Bmc Biology*, vol. 8 no.1, pp. 112, 2010.
- [20] F. Delaporte, "History of yellow fever: an essay on the birth of tropical medicine." MIT Press, 1991.
- [21] A. J. Haddow, "X.—The Natural History of Yellow Fever in Africa." *Proceedings of the Royal Society of Edinburgh, Section B: Biological Sci.*, vol. 70 no. 3, pp. 191-227, 1969.
- [22] A. J. Valleron, S. Meurisse, and P. Y. Boelle, "Historical Analysis of the 1889-1890 Pandemic in Europe." *Int. J. of Infectious. Diseases.*, vol. 12, e95, 2008.
- [23] A. Choi, and A. García-Sastre, "Influenza forensics." In *Microbial. Forensic*, pp. 89-104. Academic Press, 2020.
- [24] W. G. Laver, and R. G. Webster, "Studies on the origin of pandemic influenza: II. Peptide maps of the light and heavy polypeptide chains from the hemagglutinin subunits of A2 influenza viruses isolated before and after the appearance of Hong Kong influenza." *Virology.*, vol. 48 no.2, pp. 445-455, 1972.
- [25] M. R. Ramogale, and J. Moodley, "HIV-associated maternal mortality—primary causes of death at King Edward VIII Hospital," *Durban. South Afric. Med. J.*, vol. 97 no. 5, pp. 363-366, 2007.
- [26] A. D. Grant, G. Djomand, and K. C. De, "Natural history and spectrum of disease in adults with HIV/AIDS in Africa." *AIDS (London, England)*, 11, S43-54, 1997.
- [27] C. C. Hon, T. Y. Lam, Z. L. Shi, A. J. Drummond, C. W. Yip, F. Zeng, and F. C. C. Leung, "Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus." *J. of Virology.*, vol. 82 no. 4, pp. 1819-1826, 2008.
- [28] J. Weyer, A. Grobbelaar, and L. Blumberg, "Ebola virus disease: history, epidemiology and outbreaks." *Current. Infectious. Disease. Report.*, vol. 17 no.5, pp. 21, 2015.
- [29] X. Pourrut, B. Kumulungui, T. Wittmann, G. Moussavou, A. Délicat, P. Yaba, and E. M. Leroy, "The natural history of Ebola virus in Africa." *Microbes and infection*, vol. 7 no. 7-8, pp. 1005-1014, 2005.
- [30] A. M. Ajlan, R. A. Ahyad, L. G. Jamjoom, A. Alharthy, and T. A. Madani, "Middle East respiratory syndrome coronavirus (MERS-CoV) infection:

- chest CT findings." American Journal of Roentgenology, vol. 203 no. 4, pp. 782-787, 2014.
- [31] Novel Coronavirus (COVID-19) Cases Data, Accessed on April, 17, 2020. [Online]. Available: <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>.
- [32] C. D. Manning, and H. Schütze, "Foundations of statistical natural language processing." MIT press, 1999.
- [33] M. E. Newman, "Power laws, Pareto distributions and Zipf's law." Contemporary physics, vol. 46 no. 5, pp. 323-351, 2005.
- [34] L. R. Kalankesh, R. Stevens, and A. Brass, "The language of gene ontology: a zipf's law analysis." BMC bioinformatics, vol. 13 no. 1, pp. 127, 2012.
- [35] L. Wang, X. Li, Y. Q. Zhang, Y. Zhang, and K. Zhang, "Evolution of scaling emergence in large-scale spatial epidemic spreading." PloS one, vol. 6 no. 7, 2011.
- [36] Y. N. Pan, J. J. Lou, and X. P. Han, "Outbreak patterns of the novel avian influenza (H7N9)." Physica A: Statistical Mechanics and its Applications, vol. 401, pp. 265-270, 2014.
- [37] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data." SIAM review, vol. 51 no. 4, pp. 661-703, 2009.