



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Short Communication

A forensic analysis of SARS-CoV-2 cases and COVID-19 mortality misreporting in the Brazilian population

D. Galvêas^a, F. Barros Jr.^{b,*}, C.A. Fuzo^{c,*}^a FUCAPE Business School, Vitória-ES, Brazil^b School of Economics, Business Administration and Accounting of Ribeirão Preto, University of São Paulo, Ribeirão Preto-SP, Brazil^c Department of Clinical Analyses, Toxicology and Food Sciences, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto-SP, Brazil

ARTICLE INFO

Article history:

Received 21 February 2021

Received in revised form

5 May 2021

Accepted 12 May 2021

Available online 19 May 2021

Keywords:

COVID-19

Brazil SARS-CoV-2 pandemic

Benford's law

SARS-CoV-2 misreporting

ABSTRACT

Objective: The study aimed to investigate the misreporting number of positively tested individuals for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) succumbed or not to coronavirus disease 2019 (COVID-19) pandemic in Brazil at the city, state, and national scales using statistical forensic analysis.

Study design: This is a register-based study over public health data collected, organized, and maintained by the Ministry of Health covering the Brazilian population.

Methods: We evaluated the Brazilian notifications of positively tested cases for SARS-CoV-2 who have succumbed or not to COVID-19 between February 26th to September 7th of 2020 at the city, state, and national scales for conformity to expected distribution provided by Benford's law (BL).

Results: Statistical analyzes demonstrated a significant rejection of SARS-CoV-2 notification cases at the city and the number of deaths by COVID-19 in all regional levels according to the hypothesis of conformity to BL.

Conclusion: We demonstrated by BL, which has been widely applied to query the quality and reliability of different numerical data sources, the misreporting number of cases and deaths throughout the SARS-CoV-2 pandemic in Brazil. Therefore, we brought to light pieces of evidence that raise questions about the reliability of SARS-CoV-2 data in Brazil. This situation may have led to inconsistencies in public health policy actions, recommendations, and drastic humanitarian, social, and economic consequences such as the intensive unit care overload in some Brazilian regions.

© 2021 The Royal Society for Public Health. Published by Elsevier Ltd. All rights reserved.

Introduction

The ongoing pandemic caused by SARS-CoV-2 infection has been showing increased numbers of infected people worldwide, reaching the limit or collapsing their health systems capabilities. Recently, an unfortunate example of this disaster occurred in Manaus (Brazil), even within the perspective of herd immunity due to the high rate of infection in the local population,¹ reinforcing the idea that great care must be taken to manage this disease. The Brazilian population figures up to more than 200 million people, and more than 70% depends on public health services.² Currently, Brazil presents the second-highest number of COVID-19 infections

and deaths globally according to the Johns Hopkins Coronavirus Resource Center (<https://coronavirus.jhu.edu/>).

Specialized public health researchers have pointed some problems in the response of Brazilian authorities to face COVID-19,³ such as evidence of significant underreporting ratios of cases and deaths in Brazil.⁴ This is a severe problem because, in critical situations just like this pandemic, the data generated by the health system is essential for crafting good responses, allocating resources, and measuring the effectiveness of interventions, such as social distancing. However, incomplete or incorrect data can lead to suboptimal decisions, especially problematic in a developing economy with a fragile health system like Brazil. The SARS-CoV-2 pandemic has generated a considerable amount of data that needs to be precisely curated by government agencies to be used in crisis management. In parallel, health authorities try to process the whole new information and fight against the pandemic impacts.

* Corresponding authors.

E-mail addresses: fabarrojr@usp.br (F. Barros), cafuzo@usp.br (C.A. Fuzo).

The confidence in reported numbers of cases and deaths is an important factor for ensuring robust epidemiological studies and, consequently, adequate public health strategies. In this aspect, the adherence of data to the distribution predicted by Benford's law⁵ (BL) has been evaluated as a criterion for the reliability of extensive numeric data, which states that the sequence of multiple numbers from real-life sources is likely to be distributed in a specific non-uniform way.⁶ This counter-intuitive result can be applied to disparate lists such as stock prices, death rates, population numbers, epidemiological data, and forensic fraud analysis that somehow obey an exponential pattern.⁶ Then, when a dataset does not exhibit conformity to BL, it is indeed a piece of evidence that some misreporting has occurred. In this context, the previous study conducted by Silva and Figueiredo Filho⁷ argued the quality of nationwide Brazilian COVID-19 data and found no conformity to BL. Because of previous inconsistencies and evidence of misreporting, we extend the analysis of Silva and Figueiredo Filho⁷ by testing the agreement of Brazilian COVID-19 data to BL at different regional nation levels.

Brazilian public health data

Our analysis covers both the number of deaths and the number of confirmed cases of SARS-Cov-2 infection. Besides, we conduct experiments using data at the country, state, and city levels. We have mixed results of confirmed infections in which some tests reject the conformity to BL, and others do not. For this purpose, we analyzed the report of COVID-19 cases and deaths from the Brazilian data maintained by the Minister of Health from February 26th to September 7th of 2020 (<https://covid.saude.gov.br>), marking respectively the first case registered in Brazil and the ending of the exponential path in both Ceará and Goiás (Brazilian states). Such data were obtained at the national, state, and city levels on October 14th of the same year. Although we have collected a wider data set, we trim it at the highest registered number of cases and deaths of COVID-19 both at the national and state levels to represent the initial exponential growth. After trimming the data set, we analyze all three data sets separately. Thus, we tested the national number of cases and deaths and then we tested the merged data set from all states. The same method was used for the cities dataset. Although it might seem counter-intuitive, this is a valid way of testing,⁶ especially considering that a wider data set provides better results to such tests. **Supplementary Table 1** shows the statistical description of the analyzed data set showing an exponential pattern and cuts on the same day, i.e., 156 days after the first case registered by Brazilian authorities. Details of statistical methods are in Extended Methods in **Supplementary Information**.

Table 1
Significance test and empirical distribution of first digit according to BL and distribution of number of SARS-CoV-2 cases and COVID-19 deaths at different regional levels in Brazil.

1st Digit	BL	Cases			Deaths		
		Brazil	States	Cities	Brazil	States	Cities
1	0.301	0.349	0.295	0.296***	0.182***	0.258***	0.156***
2	0.176	0.233*	0.169	0.187***	0.167	0.212***	0.303***
3	0.125	0.089	0.132	0.122***	0.098	0.138***	0.170***
4	0.097	0.075	0.104	0.107***	0.121	0.104	0.117***
5	0.079	0.055	0.072*	0.085***	0.114	0.082	0.077***
6	0.067	0.055	0.073	0.063***	0.091	0.064	0.062***
7	0.058	0.048	0.059	0.052***	0.106**	0.055	0.049***
8	0.051	0.041	0.053	0.044***	0.083	0.047	0.036***
9	0.046	0.055	0.043	0.044***	0.038	0.042	0.031***
MAD	—	0.025 ^N	0.006 ^A	0.006 ^A	0.036 ^N	0.013 ^M	0.043 ^N
Chi-squared	—	8.233***	11.71***	17.97***	19.07***	68.61***	815.46***
Kupier V	—	1.398	1.044	20.05***	1.901**	3.762***	158.66***

Notes. *** $P < 0.01$, ** $P < 0.05$, * $P < 0.1$.
A = acceptable; M = marginal acceptable; N = nonconformity.

Results

We computed the first digit frequency in Brazil, state, and city levels (**Table 1** and **Supplementary Fig. 1**). Although the empirical distribution of the first digit of confirmed number cases fits BL, we can easily reject the hypothesis in the case of confirmed deaths. Before we move to more detailed tests, we point our first issue with the Brazilian data's reliability. Even though it is difficult to precisely estimate the fatality rate (deaths/cases) caused by SARS-CoV-2 infection as Atkeson points,⁸ it is commonly assumed to be a constant proportion of the cases. Thus, in accordance with Theorem 1 of Nigrini,⁶ if the confirmed cases data have conformity to BL we also should expect the same for confirmed deaths. However, deaths do not follow BL expected distribution in Brazil (**Supplementary Fig. 1**).

The significance tests for adherence of cases and deaths to BL are shown in **Table 1**. First, regarding the individual digits Z test, we notice that at a more aggregate level (country and states), only some Z statistics reject the null hypothesis of conformity to BL, but at city-level data, the opposite occurs. This happens for both confirmed cases and deaths. Considering the general test of conformity to BL, all three tests—MAD, Chi-squared, and Kuiper—do not support a BL distribution to confirmed Brazil deaths. On the other hand, we have mixed results in the tests using confirmed case data. Although the Chi-squared test rejects the null hypothesis in all cases, the MAD test indicates that only country-level data presents nonconformity to BL and Kupier V, indicating that only city-level data do not follow a BL distribution.

Discussion

Epidemiological studies are fundamental pieces of public health that guide, through mathematical models and population data, strategies for the management and coping of epidemics or pandemics, such as COVID-19, by governmental agencies. The detection of anomalies in the reported data used in these epidemiological models is essential for the reliability and predictive power that can be detected by some approaches like the BL law (see extended discussion in **Supplementary Information** for more details on the importance of BL for anomaly detection in COVID-19 data). In general, our results show that Brazilian data on COVID-19 do not follow the BL. One possible explanation is the underreporting of confirmed cases and deaths.⁴ It is important to mention that if the misreporting in the number of the pandemic is represented by a constant fraction of the actual numbers of the pandemics in Brazil, Theorem 1 of Nigrini⁶ assures that we should find conformity to BL even in this case, hence showing that even a specific misreporting

might fit the BL proportion. However, at the municipality level, our results unanimously indicate no conformity to BL. It contrasts with Koch and Okamura⁹ results, where it was found that Chinese confirmed infections match the distribution expected in BL and are similar to what was observed in the United States and Italy. Our analysis of the number of deaths rejects the conformity to BL at all regional levels.

The reliability in the reported data of COVID-19 pandemic is extremely relevant, being the core of any reasonable response to this health crisis that started in early 2020. In a big developing country such as Brazil, data accountability is especially important because it should drive an adequate policy response to the chaotic situation of a continental country with a vulnerable population. Although the city data seems to fit BL, if one aggregate and compares it with the national vector, one will find different figures and nonconformity of this aggregated, suggesting that the number of observations plays an important role in the statistical significance. The data of confirmed cases and deaths are a raw input of many works that investigate policy responses to the COVID-19 pandemic. For example, many investigations use some of those numbers to calibrate SIR (Susceptible, Infected, Recovered) models and trace different scenarios given alternative policies.¹⁰ Hence, our study is an alert to those who strongly rely on COVID-19 data to study or confront the pandemic effects in Brazil.

Finally, even though we could not distinguish the sources of data nonconformity to BL, our results reinforce the alert for Brazilian COVID-19 data that have been applied in the study or construction of public health policy strategies for the management of pandemic in Brazil. Since other studies find COVID-19 data in conformity to BL in other countries, one should be careful when making conclusions using Brazilian data.

Author statements

Acknowledgments

The authors thank Prof. Dr. Andrew W. Horowitz for the helpful comments.

Ethical approval

Not required.

Funding

During the course of this work, C.A. Fuzo was supported by a postdoctoral fellowship from Coordination for the Improvement of Higher Educational Personnel (CAPES-Finance Code 001).

Competing interests

The authors declare no conflict of interest.

Author contributions

FBJ designed the study; DG and FBJ performed the acquisition, analysis, interpretation, and drafting the article; CAF revisited it critically for intellectual content and wrote the final version of the article. All authors have approved the final article.

Data availability

The raw data set analyzed during the present study is freely available by Brazilian Data for the COVID-19 at <https://covid.saude.gov.br>. Codes and related data will be available upon request for the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.puhe.2021.05.010>.

References

- Sridhar D, Gurdasani D. Herd immunity by infection is not an option. *Science* 2021;**371**(6526):230–1. <https://doi.org/10.1126/science.abf7921>.
- Paim JS. Sistema Único de Saúde (SUS) aos 30 anos. *Ciencia Saude Coletiva* 2018;**23**(6):1723–8. <https://doi.org/10.1590/1413-81232018236.09172018>.
- Freitas CM, Silva IVM, Cidade NC. COVID-19 as a global disaster: challenges to risk governance and social vulnerability in Brazil. *Ambiente Sociedade* 2020;**23**:e0115. <https://doi.org/10.1590/1809-4422asoc20200115vu2020l3id>.
- Silva LV, Harb MPAA, Santos AMTB, Teixeira CAM, Gomes VHM, Cardoso EHS, et al. COVID-19 mortality underreporting in Brazil: analysis of data from government internet portals. *J Med Internet Res* 2020;**22**(8):e21413. <https://doi.org/10.2196/21413>.
- Benford F. The law of anomalous numbers. *Proc Am Phil Soc* 1938;**78**(4):551–72.
- Nigrini MJ. *Benford's Law: applications for forensic accounting, auditing, and fraud detection*. New Jersey: John Wiley & Sons; 2012.
- Silva L, Filho DF. Using Benford's law to assess the quality of COVID-19 register data in Brazil. *J Public Health* 2020;**43**(1):107–10. <https://doi.org/10.1093/pubmed/fdaa193>.
- Atkeson A. *How deadly is COVID-19? understanding the difficulties with estimation of its fatality rate*. National Bureau of Economic Research Working Paper No. 26965; 2020.
- Koch C, Okamura K. Benford's Law and COVID-19 reporting. *Econ Lett* 2020;**196**:109573. <https://doi.org/10.1016/j.econlet.2020.109573>.
- Brotherhood L, Kircher P, Santos C, Tertilt M. *An economic model of the COVID-19 pandemic with young and old agents: behavior, testing and policies*. Banco de Portugal, Economics and Research Department Working Paper No. 202014; 2020.