



COVID-19 and Seasonal Flu Data Reliability Analysis of New Cases Reported in Croatia

Dragutin Novosel^{1*}, Robert Žunac² and Matija Alanović³¹Glaserbergstrasse 23, CH - 4056 Basel, Switzerland²Petra Kružića 8, HR - 47000 Karlovac, Croatia³SIPA elektronika d.o.o., Ruđera Boškovića 17, HR - 47000 Karlovac, Croatia***Corresponding Author:** Dragutin Novosel, Glaserbergstrasse 23, CH - 4056 Basel.**Received:** April 19, 2021**Published:** May 22, 2021© All rights are reserved by **Dragutin Novosel et al.****Abstract**

Background: Recent reports have raised questions about the reliability of reported COVID-19 cases in Croatia. Analysis of the first-digit distribution in a dataset (Benford's law) is widely employed to review data consistency and is used to indicate data manipulation and/or data inconsistency.

Methods: The frequency distribution of the first digit of the number of newly reported COVID-19 cases in Croatia (2020/02/26 - 2020/12/14) was analyzed using Benford's law. Additionally, the last seven seasons (2013/09/30 - 2020/01/12) of reported cases of seasonal flu were analyzed.

Results: The number of newly reported cases of seasonal flu and newly reported cases of COVID-19 in Croatia do follow a Benford distribution ($P = 0.233$). From July 18, 2020 until August 28, 2020 the number of newly reported COVID-19 cases did not follow a Benford distribution ($P < 0.05$).

Interpretation: Our findings strongly suggest that the COVID-19 and seasonal flu data reported in Croatia are reliable. Further analyses are required to explain the disturbance observed in the summer of 2020.

Keywords: COVID-19; Seasonal Flu; Benford's Law; Data Reliability

Introduction

To manage the spread of the novel COVID-19 virus pandemic, both local and global authorities require reliable data. Countries vary widely in their health-care systems and strategies to collect COVID-19 test results. Zhang [1] analyzed the total number of cases reported from China between January 15, 2020 and February 10, 2020 and observed no indications of irregularities. However, whether the distribution of new cases currently follows Benford's law in other countries is unknown, and growing suspicions about the reliability of the provided data have resulted [2]. Sambridge and Jackson [3] analyzed the cumulative number of cases from the

first case reported until April 2020 in each of 51 countries. They concluded that the data from European countries are reliable but not data from China and South Korea [3]. Their analysis, however, does not include Croatia and this report intends to address and fill that information gap.

Benford's law, also known as the Newcomb-Benford law, the law of anomalous numbers, or the first-digit law, is used in data analysis, particularly in financial analysis and fraud detection. It can be used to analyze flu data [4] and COVID-19 data as well [1,3,5]. Because there is an increasing availability of open-access

data on the numerous parameters measured from healthy humans, from individuals with diseases, and pharmaceutical studies, there is an opportunity to analyze these datasets for their consistency and validity.

Benford’s law states that under defined circumstances the digit one (1) appears most often in a dataset. The quantity (percentage) between numbers is proportional to the space between the two numbers on a logarithmic scale. Therefore, an expected distribution occurs if the mantissas of the logarithms of numbers (but not the numbers themselves) are uniformly and randomly distributed. Notably, the digit zero (0) is not a subject of Benford’s law.

To the best of our knowledge, no reviewed published analysis currently exists in Croatia that compares the reliability of the reported cases of COVID-19 to seasonal flu data. This report fills this gap and seeks to reduce mistrust among political leaders, health-care professionals, and the general population concerning the validity of COVID-19 cases reported in Croatia.

Methodology

Data

COVID-19 raw data were obtained from the European Centre for Disease Prevention and Control [6]. We analyzed the number of newly reported cases per day.

Raw data for the seasonal flu were obtained from the World Health Organization [7]. For seasonal flu data, we analyzed the total number of new influenza-positive cases per week (all subtypes of Influenza A plus B virus).

For this study, the COVID-19 data for Croatia consisted of the first case reported on February 26, 2020 through December 14, 2020. The data of the seasonal flu included the last seven seasons (from 2013/2014 until 2019/2020).

Statistical analysis

Statistical analysis was performed using the χ^2 test. $P < 0.05$ was considered a statistically significant difference. The differences between groups found to be statistically insignificant were noted as $P =$ not significant (n.s.).

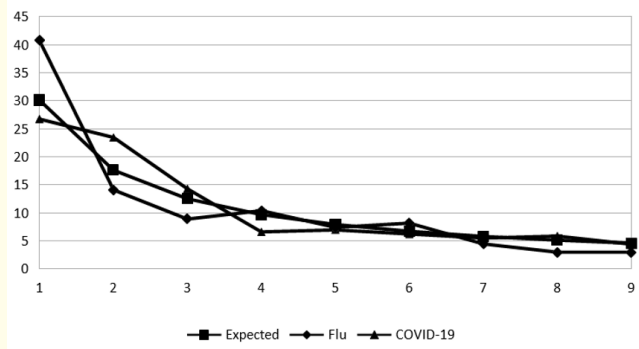
Results

Seasonal flu reports are published weekly; whereas, case data for COVID-19 has been published on a daily basis so far (until

December 14, 2020). Therefore, to increase the validity of the statistical analysis, we accumulated data for the seasonal flu over the last seven seasons (Table 1).

	Seasonal Flu	Covid-19
First Report	30/09/13	26/02/20
Last Report	12/01/20	14/12/20
Number of new Cases > 0	135	273
P Value (X ² -test)	0.239 (n. s.)	0.233 (n. s.)

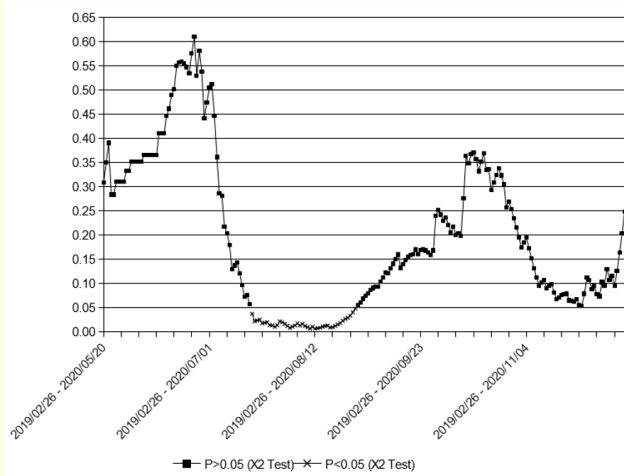
Table 1: Number of observations, periods of observations, and statistical significance (one-digit frequency distribution according to Benford’s law) of the seasonal flu and COVID-19 in Croatia.



Graph 1: Distribution of expected and observed frequency of the first digit of seasonal flu and newly reported COVID-19 cases in Croatia (expressed in %).

Discussion and Conclusion

Epidemiological analysis, which should actually be a foundation for political decisions in the current pandemic, can only be done based on careful analysis of the data. Conclusions can only be made if the collected data are reliable. Although there is currently no strong evidence that European nations are reporting unreliable data, mistrust about reported COVID-19 cases continues to grow between countries and within the countries themselves. For that reason, the data require additional analysis. Therefore, it is



Graph 2: The P-value of the χ^2 test of expected vs. observed frequency distribution of the first digit of newly reported COVID-19 cases in Croatia between May 20, 2020 and December 14, 2020 (cumulative dataset: the first dataset includes reported cases from February 26, 2020 until May 20, 2020, the subsequent datasets include data from February 26, 2020 to May 20, 2020 plus one day). Between July 18, 2020 and August 28, 2020, the reported data distribution did not follow Benford's law ($P < 0.05$).

critical to develop instruments that can enable the analysis of data consistency.

Trust in the government and health-care policy decisionmakers could greatly affect the outcome of the current pandemic. Therefore, it is of the utmost importance to develop instruments that enable data-consistency analysis. Remarkably, some conclusions can still be drawn without an in-depth analysis about why absolute numbers and the number of new cases have different distributions of the first digit.

Seasonal flu data collection has been established for decades and can serve as a control group for the first-digit distribution. Available data from Croatia, based on data published by the WHO, indicate that the first-digit distribution of newly reported COVID-19 cases and seasonal flu cases both follow Benford's law.

We began our observation of COVID-19 data from the beginning of June 2020. On a daily basis, we analyzed whether the first-digit distribution of new cases of COVID-19 followed Benford's law. We found that on every day between July 18, 2020 and August 28, 2020 the first-digit distribution of the number of new COVID-19 cases did not follow Benford's law to the level of $P < 0.05$ (χ^2 test; included data from February 26, 2020 until each day separately and incrementally, starting from May 20, 2020; Graph 2).

This finding deserves particular attention as it overlapped with the summer tourism season in Croatia. At this time, there were several comments in the free press that Croatia was reporting small numbers. Consequently, additional research was necessitated. It should be accentuated that based on our analysis we cannot make any statements regarding the cause of the potential irregularities we observed during the summer of 2020.

This analysis could easily be expanded to include other countries and could be updated as new reports are published by the WHO. Therefore, Benford's law could serve as a viable screening method for validating the consistency of reported COVID-19 data.

This study conclusively supports the hypothesis that the results of newly reported COVID-19 cases from Croatia between February 26, 2020 and December 14, 2020 are reliable with the exception of the Summer 2020 period. However, these contradictory findings require additional in-depth research.

Bibliography

1. Zhang J. "Testing case number of coronavirus disease 2019 in China with Newcomb-Benford Law" (2020). arXiv:2002.05695 [physics.soc-ph]
2. Campbell C and Gunia A. "China says it's beating coronavirus. But can we believe its numbers?" *Time* (2021). Available from: <https://time.com/5813628/china-coronavirus-statistics-wuhan/>
3. Sambridge M and Jackson A. "National COVID numbers—Benford's law looks for errors". *Nature* 581 (2020): 384. doi: 10.1038/d41586-020-01565-5.

4. Idrovo A., *et al.* "Performance of public health surveillance systems during the influenza A(H1N1) pandemic in the Americas: testing a new method based on Benford's Law". *Epidemiology Infection* 139.12 (2011): 1827-1834. doi: 10.1017/S095026881100015X.
5. Novosel D and Alanovic M. "Analysis of Consistency of Prime-boost Covid-19 Baseline and Safety Data". *ACMS* 4.12 (2020): 81-82.
6. European Centre for Disease Prevention and Control (2020). Available from: <https://www.ecdc.europa.eu/en/cases-2019-ncov-eueea>
7. World Health Organization (2020). [Internet] [Accessed November 2, 2020]. Available from: <https://apps.who.int/flumart/Default?ReportNo=12>.

Volume 5 Issue 6 June 2021

© All rights are reserved by Dragutin Novosel *et al.*