

Data Quality of Chinese Surveillance of COVID-19: Objective Analysis Based on WHO's Situation Reports

Asia Pacific Journal of Public Health
2020, Vol. 32(4) 165–167
© 2020 APJPH



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1010539520927265
journals.sagepub.com/home/aph



Alvaro Javier Idrovo, MD, PhD,¹  and Edgar Fabián Manrique-Hernández, MD¹

Abstract

Was there quality in the Chinese epidemiological surveillance system during the COVID-19 pandemic? Using data of World Health Organization's situation reports (until situation report 55), an objective analysis was realized to answer this important question. Fulfillment of Benford's law (first digit law) is a rapid tool to suggest good data quality. Results suggest that China had an acceptable quality in its epidemiological surveillance system. Furthermore, more detailed and complete analyses could complement the evaluation of the Chinese surveillance system.

Keywords

coronavirus, pandemic, epidemiological surveillance, data quality, China

Good epidemiological surveillance systems are essential for epidemic management. One of its functions is to provide data with quality that serve to make decisions based on evidence.¹ Unfortunately, it is not easy to know the quality of data during public health emergencies of international concern such as the COVID-19 pandemic. During the influenza A(H1N1) pandemic, Benford's law was proposed as an objective and fast way to assess the performance of surveillance systems during epidemics.² Its usefulness was also evidenced in the dengue epidemic in Paraguay (2009-2011)³ and the Zika epidemic in American countries.⁴

Benford's law,⁵ also called "law of the first digits," "Newcomb-Benford law," or "law of anomalous numbers," states that for a determined set of numbers, those whose first digit is 1 will appear more frequently (30.103%) than those beginning with other digits, following in order from 2 to 9 (17.609%, 12.494%, 9.691%, 7.918%, 6.695%, 5.799%, 5.115%, and 4.576%, respectively).⁶ A very good explanation of Benford's law is the following:

"The good fit of the Newcomb-Benford law to empirical data can be explained by the fact that in many cases the frequency with which objects occur in "nature" is an inverse function of their size. Very small objects occur much more frequently than do small ones, which in turn occur more frequently than do large ones and so on."⁷

The COVID-19 pandemic began in Wuhan (Hubei, China) in early December 2019.⁸ From there, it had a rapid

spread through Asia, Europe, America, and Africa. At the end of March 2020, there were more than 850 000 cases and 40 000 deaths around the world.⁹ This brief report presents the results of an objective evaluation of data quality of the Chinese epidemiological surveillance system during the ongoing epidemic.

To obtain evidence on the level of performance of the Chinese epidemiological surveillance system, we used data included in situations reports 1 to 55 of the World Health Organization (WHO) website (January 21 to March 15, 2020).¹⁰ In these situational reports are found the number of confirmed cases, suspected cases, and deaths in the past 24 hours, and cumulated confirmed cases and deaths, in each Chinese province, region, and city. Given that by March 16, the numbers of cases and deaths outside China had overtaken the total number of cases in China, the WHO decided not to report separately the situation in the Chinese territories.

Cumulative cases reported by Chinese provinces, regions, and cities were evaluated according to how closely they followed the distribution of Benford's law using log-likelihood ratio test. Results of these analyses are presented in Table 1.

¹Universidad Industrial de Santander, Bucaramanga, Santander, Colombia

Corresponding Author:

Alvaro Javier Idrovo, Departamento de Salud Pública, Escuela de Medicina, Universidad Industrial de Santander, Carrera 32 #29-31, Bucaramanga, Santander 680001, Colombia.
Email: idrovoaj@yahoo.com.mx

Table 1. Fulfillment of Benford’s Law of Chinese Epidemiological Surveillance System of COVID-19 (Cumulative Confirmed Cases).

Situation Reports	n	Log-Likelihood Ratio, P			χ^2, P		
		<.05	.05 to .10 ^a	>.10 ^a	<.05	.05 to .10 ^a	>.10 ^a
1	4			Excluded; few data			
2	14	✓				✓	
3 and 4	17			✓		✓	
5 to 12	1			Only all country data available			
13 to 33	34			✓		✓	
34 and 35	34		✓		✓		
36 and 37	34			✓		✓	
38 to 42	34		✓		✓		
43 to 55	34			✓		✓	

^aFulfillment of Benford’s law.

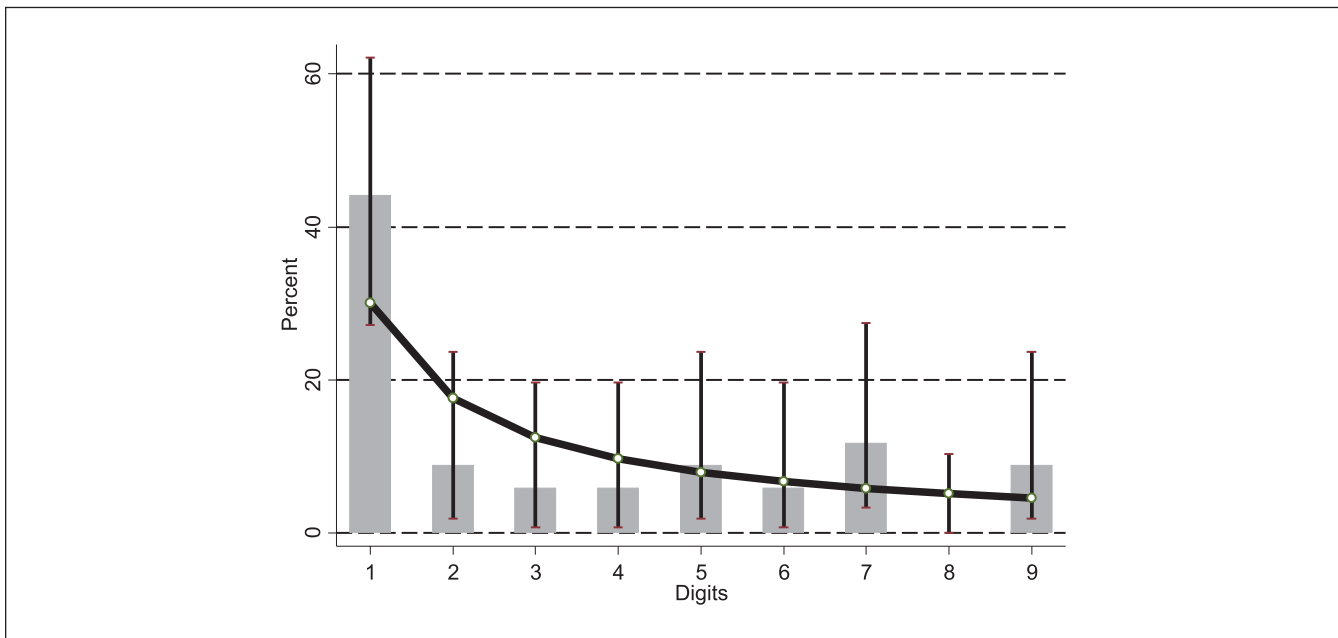


Figure 1. First digit frequencies for the Benford distribution of COVID-19 epidemic situation report 55 in China (March 15, 2020).

As can be seen, in the first situation reports, the information did not maintain a standard nor did the data show stable quality. Since the situation report 13, the data improved its quality in a more stable way. In Figure 1 is the Benford distribution of the last situation report analyzed, and a detailed analysis of each digit with the χ^2 test (Table 2).

The results suggest that the Chinese epidemiological surveillance system has had good data quality during the current health emergency. This must be understood in a pandemic context, and being the country where it originated. In these circumstances, it is highly probable that the reported data are underestimated in great magnitudes, as presented in recent publications.¹¹ It is frequent in these contexts that only most severe cases are reported. With well-designed population studies, the numerator and

denominator of the proportions of the actual occurrence of the infection will be better known. Previous experiences such as Lassa fever in Africa suggests that infections with initial high fatality diminish the severity with widespread epidemiological studies.¹²

However, this evaluation did not incorporate all the elements of the health system involved in the management of the current COVID-19 pandemic. A comprehensive evaluation of the Chinese surveillance system should add to the data quality, characteristics of simplicity, flexibility, acceptability, sensitivity, positive predictive value, representativeness, timeliness, and stability.¹ However, this rapid evaluation serves to provide feedback to officials of the Chinese surveillance system. This type of ongoing evaluation is possible in other countries, and it will allow officials in charge to make

Table 2. Details of Fulfillment of Benford's Law of Chinese Epidemiological Surveillance System of COVID-19 (Cumulative Confirmed Cases) in Situation Report 55.

First Digit	Count	Percentage		P
		Observed	Expected	
1	15	44.118	30.103	.0912
2	3	8.824	17.609	.2580
3	2	5.882	12.494	.4309
4	2	5.882	9.691	.7689
5	3	8.824	7.918	.7489
6	2	5.882	6.695	1
7	4	11.765	5.799	.1322
8	0	0	5.115	.4190
9	3	8.824	4.576	.2025

relevant decisions to improve epidemiological surveillance systems and the response of health care services.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Alvaro Javier Idrovo  <https://orcid.org/0000-0001-9579-1463>

References

- German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN; Guidelines Working Group Centers for Disease Control and Prevention (CDC). Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm Rep.* 2001;50(RR-13):1-35.
- Idrovo AJ, Fernández-Niño JA, Bojórquez-Chapela I, Moreno-Montoya J. Performance of public health surveillance systems during the influenza A(H1N1) pandemic in the Americas: testing a new method based on Benford's law. *Epidemiol Infect.* 2011;139:1827-1834. doi:10.1017/S095026881100015X
- Gómez-Camponovo M, Moreno J, Idrovo AJ, Páez M, Achkar M. Monitoring the Paraguayan epidemiological dengue surveillance system (2009-2011) using Benford's law [in Spanish]. *Biomedica.* 2016;36:583-592. doi:10.7705/biomedica.v36i4.2731
- Manrique-Hernández EF, Fernández-Niño JA, Idrovo AJ. Global performance of epidemiologic surveillance of Zika virus: rapid assessment of an ongoing epidemic. *Public Health.* 2017;143:14-16. doi:10.1016/j.puhe.2016.10.023
- Benford F. The Law of anomalous numbers. *Proc Am Philos Soc.* 1938;78:551-572.
- Hill TP. The first digit phenomenon. *Am Sci.* 1998;86:358-363.
- Formann AK. The Newcomb-Benford law in its relation to some common distributions. *PLoS One.* 2010;5:e10541. doi:10.1371/journal.pone.0010541
- Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395:497-506. doi:10.1016/S0140-6736(20)30183-5
- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time [published online February 19, 2020]. *Lancet Infect Dis.* doi:10.1016/S1473-3099(20)30120-1
- World Health Organization. Coronavirus disease (COVID-2019) situation reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>. Accessed April 30, 2020.
- Bommer C, Vollmer S. Average detection rate of SARS-CoV-2 infections is estimated around six percent. *Lancet Infect Dis.* In press.
- Binns C, Low WY, Lee MK. The COVID19 pandemic: public health and epidemiology. *Asia Pac J Public Health.* In press.