Review of Economic Analysis

REA Review of Economic Analysis

Does Digital Awareness Reduce Misreporting of Covid-19 Data? An Empirical Investigation

Kumari Youkta, Rajendra Narayan Paramanik et Shreya Nupur

Volume 15, numéro 3-4, 2023

URI : https://id.erudit.org/iderudit/1113897ar
DOI : https://doi.org/10.15353/rea.v15i3-4.5166

Aller au sommaire du numéro

Éditeur(s)

International Centre for Economic Analysis

ISSN

1973-3909 (numérique)

Découvrir la revue

Citer cet article

Youkta, K., Paramanik, R. & Nupur, S. (2023). Does Digital Awareness Reduce Misreporting of Covid-19 Data? An Empirical Investigation. *Review of Economic Analysis*, *15*(3-4), 271–283. https://doi.org/10.15353/rea.v15i3-4.5166

Résumé de l'article

Manipulation of pandemic induced casualty data poses serious threat to globe. This study is an empirical attempt to assess possible factors causing such data fudging. In order to investigate strategic data misreporting by countries, we have employed acclaimed Benford's law on numbers of deaths reported by 129 countries during both waves of pandemic. Finding reveals that alongside factors like degree of democracy, transparency etc., countries with poor digital awareness are subject to greater manipulation of data.

Review of Economic Analysis 15 (2023) 271-283 1973-3909/2023271

Does Digital Awareness Reduce Misreporting of Covid-19 Data? An Empirical Investigation.

KUMARI YOUKTA
Indian Institute of Technology Patna. Bihar. India

RAJENDRA NARAYAN PARAMANIK
*. Indian Institute of Technology Patna. Bihar. India**

SHREYA NUPUR
Indian Institute of Technology Patna. Bihar. India

Manipulation of pandemic induced casualty data poses serious threat to globe. This study is an empirical attempt to assess possible factors causing such data fudging. In order to investigate strategic data misreporting by countries, we have employed acclaimed Benford's law on numbers of deaths reported by 129 countries during both waves of pandemic. Finding reveals that alongside factors like degree of democracy, transparency, political stability, stringency etc., countries with poor digital awareness are subject to greater manipulation of data.

Keywords: Benford's law, Democracy; Data transparency; Digital awareness

JEL Classifications: D72, P16, I18

1 Introduction

The recent release of World Health Organization (WHO) statistics on Covid-19 induced excess death claims 14.91 million lives across the world and it is much higher than what countries have officially reported at about 5.4 million. It covers the time period of January 2020 to December 2021. This huge disparity between WHO and countries' estimates reignited the debate of data manipulation across the world. Though critics questioned methods of calculation of WHO, such arguments do not possess enough merit to explain this stark disparity. Forging death numbers could pose serious threat. Underreporting of positive or death cases might spread false optimism among masses, thereby encourages non-compliance with preventive measures.

* Corresponding Author. Department of Humanities and Social Sciences. Email: rajendra@iitp.ac.in;

Reliable and freely accessible data of such crisis gives early and better insights to research community, thereby helps to fight the pandemic and save valuable lives. An effective policy response to the pandemic requires transparent and credible flow of data from countries across the world.

This raises the concern that what make(s) a country more likely to manipulate data. According to existing literature foremost reason for strategic data misreporting could be poor quality of political institution. Since statistics on covid-19 are typically managed by governments and local authorities around the world, there is higher chance of data fabrication can be attributed to political motives. Adam and Tsarsitalidou (2022) and Kapoor et al. (2020) highlighted that due to the freedom of institutions, decentralization of power and a strong opposition, democratic regime finds it difficult to manipulate data. On the contrary, autocrats are more likely to misreport data. Some studies show more death cases in democracies than in authoritarian states (Frey et al., 2020 and Cepaluni et al., 2020). Other complementing factors to data misreporting such as data transparency and government effectiveness are highlighted by Annaka (2021) and Farhadi (2021). Their findings suggest that there is less chance of misreporting in countries falling in high transparency scores.

Further, in this digitalized world where information is fluid, the role of social media cannot be ignored. Covid-19 crisis unveiled true potential of digitalization. In many developing and advanced nations, despite government's poor performance to tackle pandemic, social media emerged as a last resort to save lives. Different social media platforms crowd-sourced with posts of people seeking necessary helps like oxygen cylinders, beds and medicines. In such scenario, strategic misreporting of data by government could lead to loss of confidence in the eyes of their electorate and external bodies. Therefore, government's action during the pandemic is crucial for future elections. Thus, it is expected that data misreporting to be low if a large proportion of the population is using internet and they are digitally aware. The present work has tried to fill this gap by emphasizing the role of digital awareness in data misreporting by government. In doing so we utilize cross country data covering 129 countries[1] around the world. And in order to detect strategic misreporting of excess death data we exploit Benford's law which account for such fraud.

2 Materials and Methods

Benford's Law

Benford's law is a widely applied technique for detecting the data manipulation and fraud (Farhadi, 2021). According to the law in a naturally generated data, leading digits of the numbers should be distributed across nine orders of magnitude and follow a particular

[1] The list of countries is in the Appendix, Table A2

logarithmic pattern in which the probability of first digit to be 1 is 30.1%, to be 2 is 17.61% and probability that it would be 3 is 12.5% etc. The expression of the associated probability of the digit (d) to occur as a leading digit is given by

$$P(d) = \log10\ (1+1/d) \quad (1)$$

But in an artificially created or fudged data the frequency of appearance of these numbers would vary from what has been suggested in Benford's law and this deviation of the actual distribution from the theoretical one can be termed as data misreporting. In order to estimate the misreporting of data by countries worldwide, we have used total numbers of death as reported by them. Country wise death data was obtained from World Health Organisation. We preferred using death data as it seems to be more attached with the efficiency and reputation of any government in handling such crisis. Based on availability, data of total of 129 countries were included in the analysis. Time period of the study covers both first and second wave as fatality rate of the virus was higher in first two waves. By employing Benford's law[2] we calculated Mean Absolute Deviation (MAD) for each country (see Appendix Table A3) which later utilised as a dependent variable in the study.

Our key explanatory variables are type of political regime, transparency and digital awareness. Specifically, we expect that countries having high democratic scores are less prone to data fabrication. The autonomous institutions keep checks and balances and make it difficult to manipulate data. The democracy data taken from polity2 (Marshall et al., 2020) and it ranges between -10 (most autocratic) to 10 (most democratic). Digital awareness is captured by percentage of population having internet access, data taken from World Development Indicator (WDI). Countries with larger population having internet access-a proxy for digital awareness-is expected to do less data manipulation. The rapid flow of information through social media is expected to limit the data fabrication. Another substantive explanatory variable is the transparency index created in HRV Transparency Project (Hollyer et al., 2014). It basically estimates the intent of governments to disclose country's internal affairs and its value ranges between -10 (least transparent) to 10 (most transparent). We expect a negative relationship between transparency and data misreporting tendency of countries. Detailed description of employed variables is given in Appendix Table A1.

In order to empirically analyse the relationship between excess death data misreporting (MAD) and political regime, transparency and digital awareness, we employ Ordinary Least Square (OLS) regression. It takes the following specification:

[2] The detailed methodology can be accessed from Benford F. (1938).

$$MAD_i = \alpha + \beta_1 democracy + \beta_2 transparency + \beta_3 digital\ awareness + \beta'_3 X_i + \varepsilon_i \quad (2)$$

where X_i is a set of control variables, which include the stringency index to captures government intervention during pandemic. It is an indicator that measures the extent of government intervention during pandemic and is based on nine indicators including school closures, workplace closures, and travel bans. It takes values from zero to 100 and higher values indicate stricter measures. Other control variable - political stability measures the perceptions of the likelihood of political instability and/or politically-motivated violence, including terrorism in a country. Its estimates ranges from approximately -2.5(weak) to 2.5(strong). Another control variable in the regression model is cardiovascular death rate. It denotes the death rates from cardiovascular disease measured per 100,000 individuals in a country. We have used cardio vascular death rate as a control variable in our study because studies suggests that patients with pre-existing Cardio Vascular Disease (CVD) are more likely to be critically ill and have a higher mortality rate, especially those who simultaneously have different types of CVD (Zhang et al., 2020; Khan et al., 2020). Thus, countries with greater number of people having cardiovascular disease might have faced higher death cases.

3 Results and Discussion

Table 1 presents the results of the whole empirical analysis. Results confirms the correlation between data misreporting, democracy, transparency in the governance structure and digital awareness among population. Model 1 is our baseline model in which the relationship between dependent and independent variables has been studied for the entire time period (March 2020 to August 2021) covering both the first and second wave. A significant and negative relationship ($p<0.05$) has been found between the extent of data misreporting and the political regime or degree of democracy in countries. Democratic countries have lesser tendency to hide actual death data as compared to countries where there is autocracy. The result aligns with the findings of Kapoor et al. (2020) and Adiguzel et al. (2020).

Existing literature also suggests that there is positive correlation between transparency and reported deaths during pandemic (Annaka, 2021). Countries having transparent governance structure and decision making are less involved in hiding data. In present study, transparency is found to have negative and significant impact over extent of data misreporting. Another political variable, political stability is also found to having positive and highly significant ($p<0.01$) relationship with the dependent variable. Intuitively, it is possible that if a ruling party remains in power for longer period of time then it starts behaving like an authoritarian government and in autocracy data manipulation is a common phenomenon.

Table 1: Regression result for each model

	(1)	(2)	(3)
	Baseline Model	March 2020 to August 2020	September 2020 to August 2021
Democracy	-0.095**	-0.101	-0.095
	(0.0459)	(0.065)	(0.065)
Digital awareness	-0.027**	-0.046***	-0.046***
	(0.0124)	(0.017)	(0.017)
Transparency	-0.294**	-0.486**	-0.503**
	(0.149)	(0.203)	(0.205)
Political stability	0.931***	1.312***	1.403***
	(0.333)	(0.477)	(0.489)
Stringency	-0.089***	-0.036	-0.034
	(0.017)	(0.024)	(0.025)
Cardiovascular death rate	-0.002	-0.005*	-0.004
	(0.002)	(0.003)	(0.003)
Constant	12.443***	13.676***	13.421***
	(1.376)	(2.004)	(2.050)
Observations	99	97	96

*** p<0.01, ** p<0.05, * p<0.1; Robust standard errors in parentheses.

The internet has profoundly revolutionized the way people share and access information (Tonsaker et al, 2014). Information retrieval is now easier than ever. Nowadays with digitalization and with the help of internet, information is in the reach of everyone at their fingertips. In this pandemic we realized that social media and internet can play an instrumental role not only in spreading awareness among people but also in reaching out to the administration, government and policymakers directly for their demand, appeals and oppositions. The well-informed citizen is better positioned to monitor the governance process, advocate for transparency, and hold government officials accountable (Thorbjornsrud and Figenschou, 2020). Moreover, informed citizens exert a signaling effect on government to maintain public trust for next election outcome. As pointed out by Yu et al. (2022), public trust

in government mediates the relationship between COVID-19 cases and ruling party's vote share. Government personnel may be more likely to act prudently and accurately report COVID-19 death data if they are aware of the fact that they are closely monitored by a well-informed public. This get highlighted with our results also. We have found a significant ($p<0.05$) negative relationship between data misreporting and digital awareness variable in our study which signifies that if digital awareness is high among the population then in those countries government hesitate in fabricating the data.

Further, we also split the entire time period into two parts, March 2020 to August 2020 and September 2020 to August 2021 representing both the first and second wave of the epidemic and repeated the whole analysis for each time period separately as results are reflected in Model 2 and Model 3 of Table 1. This whole exercise is a part of robustness test in our study and it is apparent that the results are not sensitive to subsampling. We have found significant relationship between political variables (political stability and transparency) and our focal independent variable-digital awareness with the dependent variable in both the models as has been obtained in the baseline model. However, the relationship between political regime and data misreporting was found to be significant only in the baseline model and not in the subsamples.

4 Robustness Analysis

We suspect the presence of correlation among predictors thus tested for multicollinearity in the data. The results of Variance Inflation Factor (VIF) (see Appendix Table A4) confirmed the absence of multicollinearity among explanatory variables. We also performed Bruesch-Pegan test to check whether data is homoscedastic or not (See Appendix Table A5). Results confirmed the presence of heteroscedasticity in the data. Thus in order to confirm the robustness of the results that we obtained with linear regression analysis we also performed Feasible Generalised Least Square model (See Table 2) for the entire period covering both the waves and for each first and second wave separately. Fortunately, results remained unaltered.

5 Conclusions

After conducting a thorough analysis of the factors that might have determined the data underreporting of the actual casualties during pandemic we found several political factors and the digital awareness among population as major players. Results suggest that as democratic nations misreports death data less in comparison to authoritarian government.

The transparency in decision making and governance also has minimising impact over data fudging tendency. Above all, our analysis also highlights the importance of digital awareness among population on the data misreported during pandemic. The empowering impact of social media and internet has clearly been reflected in the study. Countries where population has

higher access to internet are supposed to be more informed thus in those countries the extent of data underreporting is comparatively lower.

Table 2: Results of Feasible Generalized Least square model

	(1)	(2)	(3)
	Baseline Model	March 2020 to August 2020	September 2020 to August 2021
Democracy	-0.096**	-0.101	-0.096
	(0.044)	(0.063)	(0.063)
Digital awareness	-0.027**	-0.047***	-0.047***
	(0.012)	(0.017)	(0.017)
Transparency	-0.295**	-0.486**	-0.504**
	(0.144)	(0.196)	(0.198)
Political stability	0.931***	1.313***	1.404***
	(0.321)	(0.466)	(0.471)
Stringency	-0.089***	-0.037	-0.034
	(0.017)	(0.024)	(0.024)
Cardiovascular death rate	-0.002	-0.006*	-0.005
	(0.002)	(0.003)	(0.003)
Constant	12.443***	13.676***	13.421***
	(1.327)	(1.931)	(1.974)
Observations	99	97	96
Wald Chi2(6)	69.89***	36.65***	36.44***

*** p<0.01, ** p<0.05, * p<0.1; Robust standard errors in parentheses.

The transparency in decision making and governance also has minimising impact over data fudging tendency. Above all, our analysis also highlights the importance of digital awareness among population on the data misreported during pandemic. The empowering impact of social media and internet has clearly been reflected in the study. Countries where population has

higher access to internet are supposed to be more informed thus in those countries the extent of data underreporting is comparatively lower.

The pandemic has made us realize true power of digital connect in preventing thousands to precious lives. Many countries witnessed a spike in death rates because of hiding actual data from epidemiologists and health researchers. Timely release of true data would have had saved more lives. Data are not mere figures but are factual sources of information on the basis of which policies can be framed which in turn decides the fate of entire population. Thus, data misreporting is a serious issue and should be discussed at global forum. The study is not free from limitations. Due to data related constrains we are unable to include some more countries in our study but by doing so results can be made more generalizable.

Data availability statement: Data will be shared upon reasonable request.
Disclosure Statement: Authors declare no potential competing interests.
Funding Information: This research is not funded.

References

Adam, A., and Tsarsitalidou, S. 2022. "Data misreporting during the COVID19 crisis: The role of political institutions". *Economics Letters,* 213, 110348.

Adiguzel, F. S., Cansunar, A., and Corekcioglu, G. 2020. "Truth or dare? Detecting systematic manipulation of COVID-19 statistics". *Journal of Political Institutions and Political Economy*, 1, 543–557.

Annaka, S. 2021. "Political regime, data transparency, and COVID-19 death cases". *SSM-population health,* 15, 100832.

Benford F. 1938. The law of anomalous numbers. *Proc Am Philos Soc;* 78: 551–572.

Cepaluni, G., Dorsch, M., and Branyiczki, R. 2020. Political Regimes and Deaths in the Early Stages of the COVID-19 Pandemic. Available at SSRN: https://ssrn.com/abstract=3586767. or https://doi.org/10.2139/ssrn.3586767(2020/09/11). (Accessed 27 April 2020).

Farhadi, N. 2021. "Can we rely on COVID-19 data? An assessment of data from over 200 countries worldwide". *Science Progress*, 104(2), 00368504211021232.

Frey, C. B., Presidente, G., and Chen, C. 2020. "Democracy, culture, and contagion: Political regimes and countries responsiveness to covid-19". *Covid Economics*, 18.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., Tatlow, H., 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat. Hum. Behav. 5, 529–538. https://doi.org/10.1038/s41562-021-01079-8

Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. 2014. "Measuring transparency". *Political Analysis*, 22, 413–434.

Kapoor, M., Malani, A., Ravi, S., and Agrawal, A. 2020. "Authoritarian governments appear to manipulate COVID data". arXiv, https://arxiv.org/abs/2007.09566(2020/09/11

Khan, I. H., Zahra, S. A., Zaim, S., & Harky, A. (2020). At the heart of COVID-19. *Journal of cardiac surgery*, *35*(6), 1287-1294.

Marshall, M. G., Jaggers, K., and Gurr, T. R. 2020. "Polity V Project, political regime characteristics and transitions", 1800-2018. *Center for Systemic Peace*.

Thorbjørnsrud, K., & Figenschou, T. U. (2022). The alarmed citizen: Fear, mistrust, and alternative media. *Journalism Practice*, *16*(5), 1018-1035.

The World Bank. *World Bank Open Data*, <https://data.worldbank.org/> (2020).

Tonsaker, T., Bartlett, G., & Trpkov, C. (2014). Health information on the Internet: gold mine or minefield?. *Canadian Family Physician*, *60*(5), 407-408.

Yu, S., Yoo, E. J., & Kim, S. (2022). The effect of trust in government on elections during the COVID-19 pandemic in South Korea. *Asian Politics & Policy*, *14*(2), 175-198.

Zhang, J., Lu, S., Wang, X., Jia, X., Li, J., Lei, H., & Dong, W. (2020). Do underlying cardiovascular diseases have any impact on hospitalised patients with COVID-19? *Heart*, *106*(15), 1148-1153.

Appendix

Table A1: Descriptive Statistics

Variable	Mean	Std Deviation	Definition	Source
Democracy	4.98	5.71	The democracy data taken from polity2 and it codes democracy between -10 (most autocratic) to 10 (most democratic).	Marshall et al.(2020)
Transparency	1.29	1.93	It basically estimates the intent of governments to disclose country's internal affairs and its value ranges between -10 (least transparent) to 10 (most transparent).	Hollyer et al.(2014)
Stringency Index	59.91	12.70	A Composite measure that captures government intervention during pandemic and based on nine indicators including school closures, workplace closures, and travel bans, taking values from zero to 100 (higher values indicate stricter measures).	Hale et al. (2021)
Political stability	-0.21	0.94	Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically motivated violence, including terrorism. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5.	World Bank's Development Indicators (WDI)
Digital awareness	62.46	27.15	Internet users are individuals who have used the Internet (from any location) in the last 3 months. The Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc.	WDI
Cardiovascular death rate	230.29	109.17	Death rates from cardiovascular disease measured per 100,000 individuals	Global Burden of Disease Collaborative Network, Global Burden of Disease Study 2017

Table A2: List of Countries

Afghanistan
Angola
Argentina
Australia
Austria
Belgium
Benin
Burkina Faso
Bangladesh
Bolivia
Botswana
Brazil
Burundi
Bulgaria
Cambodia
Canada
Cameroon
Central African Republic
Chad
Chile
China
Colombia
Congo-Brazzaville
Costa Rica
Cuba
Cyprus
Denmark
Dominican Republic
Ecuador
Egypt
Ethiopia
Finland
Fiji
France
Gabon
Gambia
Ghana
Guinea-Bissau
Greece
Guatemala
Guinea
Guyana
Haiti
Honduras
Hungary
India
Indonesia
Ireland
Iran
Iraq
Israel
Italy
Cote D'lvoire
Jamaica
Jordan
Japan
Kenya
Kuwait
Laos
Lebanon
Lesotho
Libya
Mauritania
Madagascar
Malaysia
Mauritius
Malawi
Mexico
Mali
Mongolia
Morocco
Mozambique
Nepal
New Zealand
Nicaragua
Nigeria
Niger
Norway
Netherlands
Oman
Pakistan
Panama
Paraguay
Peru
Philippines
Papua New Guinea
Poland
Portugal
Romania
Russia
Rawanda
South Africa
El salvador
Saudi Arabia
Sudan-North
Senegal
Serbia
Sierra Leone
Singapore
Slovak Republic
Slovenia
Somalia
Spain
Sri Lanka
South Sudan
Suriname
Sweden
Switzerland
Syria
Tajikistan
Tanzania
Thailand
Togo
Trinidad and Tobago
Tunisia
Turkey
UAE
Uganda
United Kingdom
Ukraine
Uruguay
United States
Uzbekistan
Venezuela
Vietnam
Yemen
Zambia
Zimbabwe

Table A3: Estimated MADs for the complete sample of countries

Country	MAD	Country	MAD
Afghanistan	0.979	Israel	2.191
Angola	2.797	Italy	3.055
Argentina	1.974	Cote D'lvoire	7.218
Australia	2.854	Jamaica	1.759
Austria	2.2	Jordan	3.452
Belgium	2.639	Japan	1.56
Benin	5.139	Kenya	2.379
Burkina Faso	7.318	Kuwait	2.028
Bangladesh	3.678	Laos	11.829
Bolivia	2.319	Lebanon	2.195
Botswana	2.32	Lesotho	4.064
Brazil	2.453	Libya	1.969
Burundi	11.62	Mauritania	2.543
Bulgaria	1.066	Madagascar	2.697
Cambodia	4.685	Malaysia	1.215
Canada	1.694	Mauritius	10.65
Cameroon	1.752	Malawi	4.053
CentralAfrican	6.852	Mexico	4.842
Chad	9.872	Mali	4.675
Chile	2.115	Mongolia	3.523
China	5.234	Morocco	2.677
Colombia	3.004	Mozambique	4.383
Congo-	6.136	Nepal	3.148
Costa Rica	5.398	New Zealand	10.45
Cuba	1.579	Nicaragua	10.254
Cyprus	4.661	Poland	1.027
Denmark	4.399	Portugal	1.348
Dominican	0.863	Romania	2.505
Ecuador	1.958	Russia	5.109
Egypt	3.421	Rawanda	4.062
Ethiopia	2.384	South Africa	1.569
Finland	3.156	El salvador	6.394
France	1.648	Saudi Arabia	3.06
Gabon	7.669	Sudan-North	3.015
Gambia	5.291	Senegal	2.163
Ghana	1.519	Serbia	1.951
Guinea-Bissau	7.242	Sierra Leone	9.12

Greece	1.689	Singapore	12.755
Guatemala	3.004	Slovak	2.671
Guinea	7.509	Slovenia	1.936
Guyana	5.456	Somalia	2.957
Haiti	3.178	Spain	1.486
Honduras	0.968	Sri Lanka	3.047
Hungary	4.114	South Sudan	8.316
India	2.833	Suriname	3.59
Indonesia	4.548	Sweden	2.111
Ireland	1.8	Switzerland	2.294
Iran	2.543	Syria	4.129
Iraq	3.922	Tajikistan	9
Yemen	2.823	Tanzania	9.014
Zambia	1.623	Thailand	4.871
Zimbabwe	3.145	Togo	9.725
Fiji	4.696	Vietnam	6.447

Source: Author's calculation

Note: Table A3 presents the calculated MADs for 129 countries. However, regression analysis has small number of observation due to missing data for some variables.

Table A4: Multicollinearity Test (VIF)

Independent Variables	VIF	1/VIF
Internet access	2.41	0.41
Political stability	2.07	0.48
Transparency	1.73	0.57
Degree of democracy	1.48	0.67
Cardio vascular death rate	1.48	0.67
Stringency	1.15	0.86
Mean VIF	1.72	

Source: Author's calculation

Table A5: Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

BP test Value	p-value	Remarks
22.79	0.0000	Null hypothesis of constant variance rejected.

Source: Author's calculation