

Novel Validated Index for the Measurement of Disinformation Susceptibility at the County Level

Michael X. Jin¹, Sangita Rajan², Carlos E. Gary Bicas³, Max Hao¹, Letian Dong⁴, Beckett Mufson⁵, Imran Hafiz⁶

1. Radiology, Stony Brook University Hospital, Stony Brook, USA 2. Data Science, University of California Los Angeles, Los Angeles, USA 3. Planetary Scientist, Stony Brook University, Stony Brook, USA 4. Economics, University of North Carolina, Chapel Hill, USA 5. Journalism, Hofstra University, Hempstead, USA 6. Journalism, Duke University, Durham, USA

Corresponding author: Michael X. Jin, drmichaelxjin@gmail.com

Abstract

In the past decade, disinformation has become an increasingly dangerous enemy of public health, scientific advancement, and social stability. To address and counter this trend, it is essential to first identify communities most at risk for disinformation. The Jin-Hafiz Disinformation Index (JHDI) is developed and validated as a tool to counter disinformation and address deficits of good information on a county level in the United States. Once vulnerable communities are identified with the JHDI, targeted interventions with community partnerships can be conducted to address knowledge concerns.

Categories: Internal Medicine, Healthcare Technology, Epidemiology/Public Health

Keywords: disinformation, index, public health, community health, government, corona virus, measurement validation

Introduction

The growth of social media has allowed an increase in viewpoints and voices. However, there is a small but prominent portion of these voices that are not based on facts or, worse, are intentionally misleading. Disinformation, or “false information which is intended to mislead” as defined by Oxford dictionaries, further degrades the trust the public has for institutions that are responsible for producing and verifying information based on facts, including the media, universities, government, and scientific institutions [1]. Mass media and social networks have been fundamental in the management of public health related information in recent years. However, disinformation, misinformation, and fake reports have deluged this avenue of information delivery [2-4]. Disinformation is particularly damaging towards public health and patient care, especially in the midst of the ongoing coronavirus disease 2019 (COVID-19) pandemic.

The COVID-19 pandemic has thrown this epistemic crisis into sharp relief. COVID-19 has killed more than 2 million people and infected nearly 100 million more worldwide. In the United States alone, COVID-19 has killed more than 500,000 people and infected nearly 30 million, which is near a third of the total cases in the world [2]. Yet, despite the staggering mortality and infection rates, nearly 40% of Americans still say they do not intend to take the vaccine. Most frequently cited reasons include uncertainty of vaccine effectiveness and possibility of unknown side effects despite increasing evidence that the vaccine is safe and effective [3]. Tedros Adhanom Ghebreyesus, Director-General of the World Health Organization (WHO), recognized the overabundance of information, misinformation, disinformation, and outdated information causing people to ignore the empirically backed COVID-19 guidelines as an “infodemic” [4].

For those working to disseminate scientifically verified public health information, a primary challenge is discerning and measuring how misinformation impacts a given population’s behavior. For the purposes of this paper, we are specifically interested in public health misinformation surrounding COVID-19. We will outline a process for indexing publicly available data to measure the susceptibility of communities in the United States to disinformation and to study the efficacy of anti-disinformation campaigns. This index shall be referred to as “The Jin-Hafiz Disinformation Index” (JHDI) throughout this document.

Materials And Methods

A literature review was conducted to determine demographic factors most associated with disinformation [5-14]. Out of 15 factors, five were ultimately selected for the index after evaluating for availability of data in census and publicly available databases with resolution at a county level, including race, gender, poverty status, education, and unemployment (Table 1). Commonalities in variables were also cross-referenced with the Centers for Disease Control and Prevention (CDC) Social Vulnerability Index (SVI) [15]. These factors were used to create three preliminary indices with different weights placed on each variable in which the sum of each index value is added together to produce the index score (Table 2). Each of the three indices was then validated against disinformation variables from the Yale Climate Survey data [16-18]. The index version that most closely predicted both disinformation variables was selected as the final index.

Review began 05/09/2021

Review ended 05/24/2021

Published 05/28/2021

© Copyright 2021

Jin et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

How to cite this article

Jin M X, Rajan S, Gary Bicas C E, et al. (May 28, 2021) Novel Validated Index for the Measurement of Disinformation Susceptibility at the County Level. Cureus 13(5): e15305. DOI 10.7759/cureus.15305

Variables selected for index	Variables correlated with disinformation not selected for index due to incomplete of county-level data	Variables correlated with disinformation not selected because political considerations of partisan divide
Population not racial/ethnic minority	Social isolation	Evangelical population
Speaks English less than well	Trust of authorities in science and medicine	
Household below poverty	Physical or mental disability	
Population without high school diploma	Religious belief	
Unemployment	Social influence score	
	Birthplace of parents	
	Population without internet access	
	Property crime rates relative to national mean	
	Higher conservative-leaning population	

TABLE 1: Demographic and Social Variables Identified to Have Significant Correlation With Disinformation

Index Version	Multiplier for RAC	Multiplier for LANG	Multiplier for POV	Multiplier for HSD	Multiplier for EMP
Ver. 1	1x	1x	1x	1x	1x
Ver. 2	2x	1x	1x	2x	2x
Ver. 3	3x	1x	1x	3x	2x

TABLE 2: Preliminary Indices

RAC = z-score of % population not racial/ethnic minority

LANG = z-score of % speaking English less than well

POV = z-score of % household below poverty

HSD = z-score of % population without high school diploma

EMP = z-score of % unemployed

To do this, census data was first filtered based on year (only including the most recent data) and then county codes were converted to Federal Information Processing Standards (FIPS) codes. Columns for population proportions were then created corresponding to:

- Race: people categorized as not racial/ethnic minority
- Speak English less than very well
- Poverty: people below poverty level in the last year
- Educational attainment: people aged 25 years and over who do not have any college degree

- Unemployment: unemployment rate for people aged 20-64 years

County values were calculated for each of the three preliminary indices.

Global warming opinions compiled by the Yale Climate Survey were used to validate the preliminary indices. Four distinct areas from the 63 questions of the survey were selected based on their association with disinformation given the amount of evidence and testable hypotheses presented against such beliefs:

- Estimated percentage who do not think that global warming is happening
- Estimated percentage who think that global warming is caused mostly by natural changes in the environment
- Estimated percentage who believe there is a lot of disagreement among scientists about whether or not global warming is happening
- Estimated percentage who somewhat/strongly disagree that global warming is affecting the weather in the United States

Unweighted means were then calculated for the four variables for each county. Values were normalized and then processed into a heat map of the continental United States (Figure 1). Pearson's tests were used to compare the correlation of each of the indices with the county values obtained from the Yale Climate Survey.

Climate Change Data

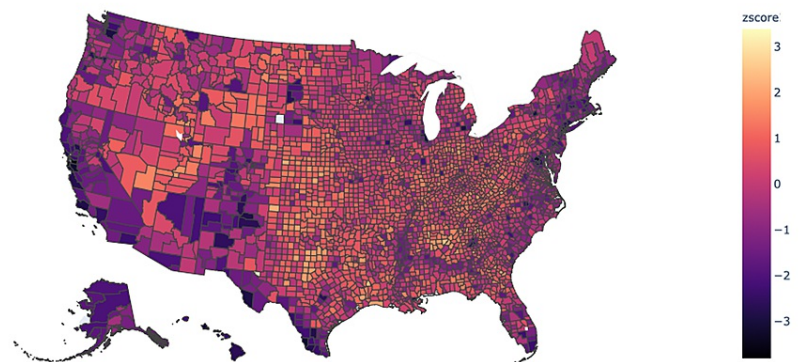


FIGURE 1: Heat Map for Climate Data

Results

Pearson's tests were used to compare the correlation of each of the indices with the county values obtained from the Yale Climate Survey. Values for the indices were normalized and then processed into heat maps of the continental United States. Analysis showed that index version 1 demonstrated weak correlation ($R=0.1$), version 2 demonstrated mild correlation ($R=0.2$), and version 3 demonstrated moderate correlation ($R=0.4$) (Figures 2-4). Version 3 of the JHDI demonstrated the best correlation with the Yale Climate Survey in predicting community susceptibility to disinformation (Table 3).

Jin-Hafiz Index ver. 1

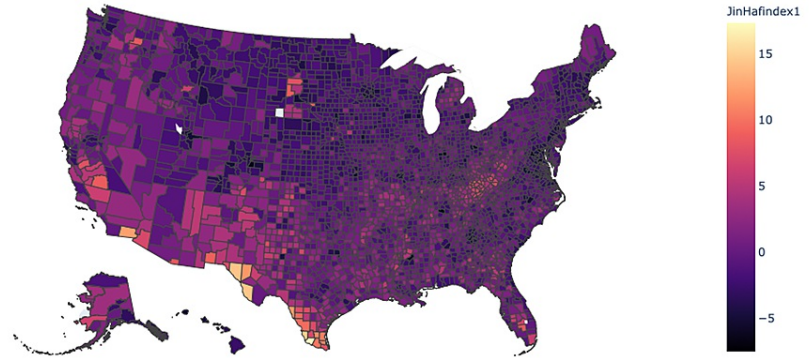


FIGURE 2: Heat Map of JHDI Version 1

correlation: 0.071; determination: 0.005

JHDI, Jin-Hafiz Disinformation Index

Jin-Hafiz Index ver. 2

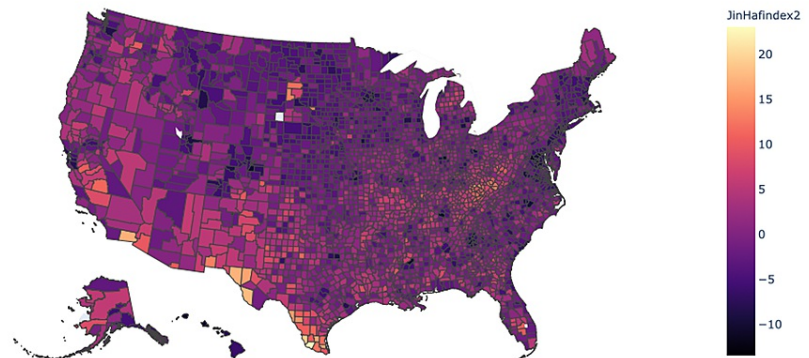


FIGURE 3: Heat Map of JHDI Version 2

correlation: 0.224; determination: 0.050

JHDI, Jin-Hafiz Disinformation Index

Jin-Hafiz Index ver. 3

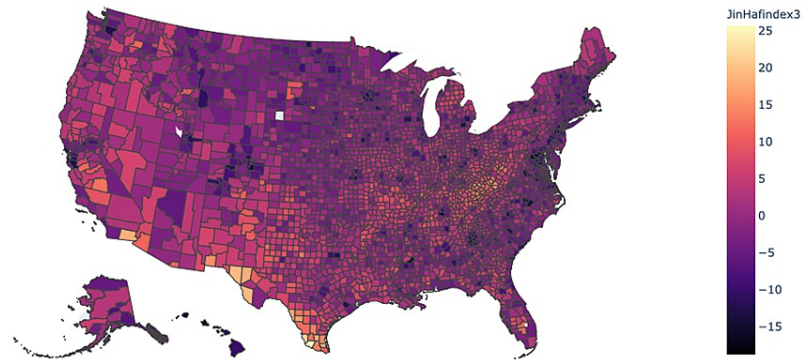


FIGURE 4: Heat Map of JHDI Version 3

correlation: 0.361; determination: 0.130

JHDI, Jin-Hafiz Disinformation Index

Version 3

z-score of % racial/ethnic minority (x3)

z-score of % speaking English less than well

z-score of % household below poverty

z-score of % population without high school diploma (x3)

z-score of % unemployed (x2)

Total score:

TABLE 3: Final Index

Score interpretation:

≤10: Extremely resistant to disinformation

-10 to -5: Above average resistance to disinformation

-5 to 5: Average susceptibility to disinformation

5 to 10: Above average susceptibility to disinformation

>10: Extremely susceptibility to disinformation

Discussion

Much of the current discussion regarding disinformation focuses on why and how it spreads. The JHDI seeks to understand the “who” piece of the puzzle: the populations at greatest risk of falling prey to disinformation. In order to fight disinformation, it is essential to determine which communities are most vulnerable to information, especially since falsehoods can spread six times as fast as the truth on social platforms [19]. A recent NPR/Marist poll found that one in four people living in the United States said they would “refuse a COVID-19 vaccine outright if offered” due to largely unfounded fears regarding the vaccine, threatening the prospects of herd immunity in America [20]. In these situations, the JHDI can be used to triage the most vulnerable areas and create targeted campaigns to counter viral disinformation and encourage vaccination across platforms to fight disinformation.

To our knowledge, the JHDI is the first of its kind. Other indices such as the Global Disinformation Index focuses on disinformation produced by national governments [21]. Another similar index was formulated by CDC's SVI, which endeavors to identify communities most vulnerable to natural disasters by using the data from the 2020 U.S. Census of Population and Housing and ranking each U.S. county (N=65,081) by the 15 variables within the U.S. census, such as income per capita, education, and single parenting [15]. The rank score is totaled for each county, and the county with the lowest total rank when all 15 variables were combined is deemed most vulnerable. However, there are several limitations to the CDC's SVI.

First, CDC's SVI is primed for identifying primarily communities with social and logistical barriers to receiving aid, whereas the JHDI identifies susceptibility of communities to disinformation. Additionally, CDC's SVI gave equal weight to all 15 factors when ranking vulnerability. However, as we have shown with our index validation, applying significance factors can have a marked impact on index performance and provide more accurate data (Figures 2-4). Finally, the CDC's SVI has not been validated against a national database for disinformation, whereas the JHDI was tested against the Yale Climate Change data with moderate correlation (R=0.4).

Limitations and next steps

The sourced data is fairly robust and reliable, and provides sufficient granularity for our purposes. However, there are several nuances that complicate this analysis, for example, misinformation on public health issues could be driven by microcultural circumstances that are not relevant to misinformation on other issues such as climate change. One documented example of this is the anti-vaccination tendencies among Somali communities; these populations may not have as strong opinions towards other issues such as global warming but feel strongly against anti-vaccination, which makes finding an association more difficult [5]. However, the rationale for exploring microcultural associations is strong with associations influenced by confounding factors. These associations can be explored in future studies with pattern analysis of social media platforms to find overlaps of association.

Furthermore, it would be interesting to understand the impact of local business patterns on public health misinformation. Misinformation has been used to protest lockdown measures in places where local economies have been hit the hardest; therefore, it might be interesting to investigate local economies. Starting with industries most impacted by the pandemic, counties that rely on them most can be selected for closer review. For example, it has been reported that meatpacking firms are facing investigation due to cases that have originated at packing locations [6]. It would be interesting to examine the nature of COVID-19 information dissemination in regions that rely on meatpacking for economic stability.

Additionally, variables such as lack of internet access, property crime rates, and population that are politically conservative are highly correlated with disinformation but were not included in this editions of the index due to incomplete county-level data [13]. Future studies can include these variables to the index to potentially augment disinformation susceptibility predictions. Furthermore, other databases such as non-medical vaccination exemption, mask-wearing belief, and Holocaust denial can also be used for validation of the index. However, caution should be used as the disinformation variables these databases contain may be influenced by additional social and political factors and are often less substantiated than the data from the Yale Climate Survey, resulting in a less accurate measure of disinformation.

Finally, it would be desirable to analyze a more dynamic data set. Currently, our data is relatively static; it captures snapshots of communities at specific points in accordance with census data. Moving forward, the model could be made more dynamic by incorporating news and social media sentiment analysis with advances in artificial intelligence. Doing so will make the JHDI an even more powerful tool in addressing disinformation.

Conclusions

The JHDI has applications beyond improving public health outcomes around COVID-19. In an unprecedented era of mass communication, communities are at risk of losing empiricism and factfulness due to the mechanisms of sharing on social media. It is clear that as society evolves to incorporate new forms of media, institutions should evolve to adapt to new information ecosystems to promote fact-based information over mistruth. Using this index, local governments can now get a sense of how pervasive the problem of disinformation is and work to address disinformation proactively by using policy regulating social media, "info-interventions" to serve low-information populations with facts, and new civic media platforms designed with this problem in mind.

However, as with many types of repair work, the first step is figuring out how to measure where reality is broken.

Additional Information

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Americans' Views of Government: Low Trust, but Some Positive Performance Ratings . (2020). Accessed: March 2, 2021: <https://www.pewresearch.org/politics/2020/09/14/americans-views-of-government-low-trust-but-some-positive-performance...>
2. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). (2021). Accessed: March 17, 2021: <https://coronavirus.jhu.edu/map.html>.
3. Intent to Get a COVID-19 Vaccine Rises to 60% as Confidence in Research and Development Process Increases. (2020). Accessed: January 7, 2021: <https://www.pewresearch.org/science/2020/12/03/intent-to-get-a-covid-19-vaccine-rises-to-60-as-confidence-in-research...>
4. UN tackles 'infodemic' of misinformation and cybercrime in COVID-19 crisis . (2020). Accessed: March 2, 2021: <https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-c...>
5. Cutilli CC, Simko LC, Colbert AM, Bennett IM: Health literacy, health disparities, and sources of health information in U.S. older adults. *Orthop Nurs*. 2018, 37:54-65. [10.1097/NOR.0000000000000418](https://doi.org/10.1097/NOR.0000000000000418)
6. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K: Health literacy and health information technology adoption: the potential for a new digital divide. *J Med Internet Res*. 2016, 18:e264. [10.2196/jmir.6549](https://doi.org/10.2196/jmir.6549)
7. Chen X, Hay JL, Waters EA, et al.: Health literacy and use and trust in health information . *J Health Commun*. 2018, 23:724-3. [10.1080/10810730.2018.1511658](https://doi.org/10.1080/10810730.2018.1511658)
8. Protheroe J, Whittle R, Bartlam B, Estacio EV, Clark L, Kurth J: Health literacy, associated lifestyle and demographic factors in adult population of an English city: a cross-sectional survey. *Health Expect*. 2017, 20:112-9. [10.1111/hex.12440](https://doi.org/10.1111/hex.12440)
9. Estacio EV, Whittle R, Protheroe J: The digital divide: Examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. *J Health Psychol*. 2019, 24:1668-75. [10.1177/1359105317695429](https://doi.org/10.1177/1359105317695429)
10. Clouston SAP, Manganello JA, Richards M: A life course approach to health literacy: the role of gender, educational attainment and lifetime cognitive capability. *Age Ageing*. 2017, 46:495-9. [10.1093/ageing/afw229](https://doi.org/10.1093/ageing/afw229)
11. Christy SM, Gwede CK, Sutton SK, et al.: Health literacy among medically underserved: the role of demographic factors, social influence, and religious beliefs. *J Health Commun*. 2017, 22:923-31. [10.1080/10810730.2017.1377322](https://doi.org/10.1080/10810730.2017.1377322)
12. Stormacq C, Van den Broucke S, Wosinski J: Does health literacy mediate the relationship between socioeconomic status and health disparities? Integrative review. *Health Promot Int*. 2019, 34:e1-e17. [10.1093/heapro/day062](https://doi.org/10.1093/heapro/day062)
13. Yang TC, Matthews SA, Shoff C: Individual health care system distrust and neighborhood social environment: how are they jointly associated with self-rated health?. *J Urban Health*. 2011, 88:945-58. [10.1007/s11524-011-9561-x](https://doi.org/10.1007/s11524-011-9561-x)
14. Goreis A, Voracek M: A systematic review and meta-analysis of psychological research on conspiracy beliefs: field characteristics, measurement instruments, and associations with personality traits. *Front Psychol*. 2019, 10:205. [10.3389/fpsyg.2019.00205](https://doi.org/10.3389/fpsyg.2019.00205)
15. Flanagan BE, Gregory EW, Hallisey EJ, Heitgerd JL, Lewis B: A social vulnerability index for disaster management. *J Homel Secur Emerg Manag*. 2011, 8:1-22.
16. Yale Climate Opinion Maps 2020. (2020). Accessed: March 2, 2021: <https://climatecommunication.yale.edu/visualizations-data/ycom-us/>.
17. Olive JK, Hotez PJ, Damania A, Nolan MS: The state of the antivaccine movement in the United States: a focused examination of nonmedical exemptions in states and counties. *PLoS Med*. 2018, 15:e1002578. [10.1371/journal.pmed.1002578](https://doi.org/10.1371/journal.pmed.1002578)
18. A Detailed Map of Who Is Wearing Masks in the U.S. . (2020). Accessed: March 2, 2021: <https://www.nytimes.com/interactive/2020/07/17/upshot/coronavirus-face-mask-map.html>.
19. Vosoughi S, Roy D, Aral S: The spread of true and false news online . *Science*. 2018, 359:1146-51. [10.1126/science.aap9559](https://doi.org/10.1126/science.aap9559)
20. NPR/PBS NewsHour/Marist Poll: September 2018. (2018). Accessed: March 2, 2021: <https://ropercenter.cornell.edu/ipoll/study/31115585>.
21. Global Disinformation Index. Time to defund the coronavirus infodemic . (2021). Accessed: April 9, 2021: <https://disinformationindex.org/>.