International Journal on Data Science and Technology
2024, Vol. 10, No. 2, pp. 38–44
https://doi.org/10.11648/j.ijdst.20241002.13

Research Article

A Sentiment Analysis of News Articles Published Before and During the COVID-19 Pandemic

Hugo Moisés Montesinos-Yufa*, Emily Musgrove

Department of Mathematics, Computer Science, and Statistics, Ursinus College, Collegeville, United States

Abstract

This study investigates the impact of the COVID-19 pandemic on the connotative language used in news articles, leveraging sentiment analysis to gauge shifts in societal attitudes and potential implications for mental health. Utilizing the statistical programming language R, we extracted and analyzed texts from 645 articles published before and during the pandemic by nine authors across three major U.S. newspapers: The Wall Street Journal, New York Times, and The Washington Post. Employing the AFINN and NRC sentiment lexicons, we observed a statistically significant decrease in sentiment during the pandemic period ($p < 0.0001$), suggesting a pervasive shift in media discourse. This decline, consistent across newspapers and journalists, highlights the profound impact of the pandemic on societal attitudes, reflecting the pain and stress experienced by many. Such a decline in sentiment can create a negative feedback loop that exacerbates the already significant health and behavioral challenges triggered by the pandemic and its associated mitigation measures. Our findings underscore the value of sentiment analysis and text mining in assessing the effects of high-stress, long-term events on global public health while identifying a gap in the existing literature that prioritizes disease-focused research over holistic well-being. This study highlights the critical role of journalists and leaders in shaping public sentiment during crises, advocating for early recognition of concerning trends. It also offers a valuable framework for future research connecting major events with the overall media sentiment and their subsequent effects on public health.

Keywords

COVID-19, Sentiment Analysis, Mental Health, News Media, Text Mining, Pandemics, Social Attitudes, Linguistics

1. Introduction

In our globally interconnected world, news media shapes our understanding of events far beyond our immediate experience. Journalistic integrity – the pursuit of truth and the minimization of bias – is thus essential. It influences what matters to us, how we perceive our place in the world and our actions. During times of crisis like the COVID-19 pandemic, biased or emotionally heightened reporting can skew our perception of the world. This can damage relationships, deepen anxieties, and create a sense of hopelessness.

This paper examines how news reporting shifted in sentiment before and during the COVID-19 pandemic. Using sentiment analysis of text-mined articles in both periods, we uncover changes in connotative language use reflecting the journalist's own attitudes that transmit to their readers. This research aims to help us understand how the news shapes our priorities, our sense of community, and our willingness to act.

*Corresponding author: hmontesinosyufa@ursinus.edu (Hugo Moisés Montesinos-Yufa)

Received: 4 July 2024; **Accepted:** 1 August 2024; **Published:** 27 August 2024

Understanding these patterns empowers a greater demand for balanced journalism and promotes critical media literacy.

Literature Review

Text analysis is a fundamental aspect of detecting the sentiment of words or phrases. By analyzing the word content of a sentence, researchers may either search for specific words with known sentiments from a dataset specific to the project, use a non-specific lexicon of words with previously determined sentiments, or use artificial intelligence to assign the sentiment of a word based on the unique context in which it is used [1-8]. While analyzing newspaper articles from before and during the COVID-19 pandemic is the primary focus of our research, researchers have analyzed texts from a number of mediums, contexts, and periods to detect the sentiments, opinions, and emotions of the authors. Sentiment analysis, in particular, has been used in various contexts unrelated to the COVID-19 pandemic, such as detecting hate speech on social media, and exploring the sentiment of media reviews [1, 3, 9]. Many texts published before the pandemic focus on the best practices in sentiment analysis rather than having defined real-world implications [3]. An exception to this is in the financial sector, where sentiment analysis has been used to optimize financial and stock trading decisions since before the pandemic [5, 10]. However, information derived from texts written during a health crisis could have direct real-world implications and could promote policy changes that influence how we approach our daily life [2, 4]. For instance, sentiment analysis of texts written during the COVID-19 pandemic could help officials and academics figure out how best to alleviate health concerns within the population. The mitigation efforts to curb the COVID-19 pandemic also forced widespread behavioral shifts across the populations, affecting the immune system, with documented repercussions for both physical and mental well-being [11]. However, while an extensive body of literature exists on diagnosing, preventing, and treating specific diseases, research focusing on prolonged health and its causes in the absence of disease remains relatively limited [12]. Similarly, research connecting the overall media sentiment with the population health including mental health remains a big gap in the literature. Some studies have used sentiment to assess the underlying opinions, emotions, and mental state of the writer [3-5, 13]. But our research has much broader implications. Unlike previous articles, we uncover a systematic and widespread sentiment decline in journal articles written during the COVID-19 pandemic. Such a decline in sentiments can trigger a systematic, widespread, and potentially long-term deterioration of global public health, the full extent of which remains to be seen.

Prior articles have also supported the idea that high-stress events, such as the COVID-19 pandemic, will adversely affect one's mental health by detrimentally changing their routine and interpersonal relationships as well as the economy and politics of the world around them [14]. The COVID-19 pandemic and the subsequent mitigation measures have impacted the world on both an individual level, affecting individuals' interpersonal relationships and mental health, and a global level, such as impacting the entire financial market [14-16]. These changes caused many individuals to develop new or worsened mental illnesses that have the potential to affect them for the rest of their lives [17-19]. In addition, this volatility leads to many individuals seeking out information and reassurance from the news. This desire to stay informed often leads to doomscrolling, in which a person purposefully engages with negative news and media [19]. An increased consumption of COVID-19 and negative news is associated with feelings of helplessness and psychological distress [20, 21]. Thus, the increase in negative news can cause individuals to seek out more negative news as a self-sabotaging method of staying informed [19]. Moreover, the decreased sentiment of news articles is also associated with decreases in stock market indices and increases in realized volatility [22, 23]. Negative sentiments lead to financial concerns, influence risk aversion, and alter financial decision-making of individuals [24-27]. Financial concerns also affect human behavior and health directly, further worsening the already negative impact that the Covid-19 pandemic and the mitigation efforts had on the economy and on the health of the population [28-32]. In other words, a declining sentiment in times of crisis promotes a negative feedback loop of health, sentiment, behavior, and poor decision making. Early identification of such a declining pattern of sentiments of new media articles can help break this vicious cycle and curb the negative shift in mental health of the U.S. population, their overall health, their behavior, and on the economy. Identifying these concerning changes early can promote not only timely remedial measures but also a better understanding of journalists' and leaders' responsibility in times of crisis.

2. Data and Methods

Data Collection

To gather our data, we collected pdf copies of the articles we would be examining the sentiment of. We chose articles from the New York Times, the Washington Post, and the Wall Street Journal. These journals are the largest daily newspapers in the U.S. as of June 2023 [33]. For each newspaper, we chose nine (9) authors and 24 works per author, 12 written before and 12 written during the pandemic. The combinations result on a total potential sample size of up to N = 648 articles (24 articles per journalist × 9 journalists per journal × 3 journals).[1] Table 1 lists the 27 journalists (9 journalists per journal × 3 journals) showing their general writing topic and their profile links.

For each article, we followed a multi-step procedure to clean, convert, analyze, and visualize the data. First, we scraped and cleaned the text from the articles and tabulated the words and a tally of their usage. We kept the text's body and removed punctuation marks, newline characters,

[1] The exception was Lawrence Ulrich, who had 9 articles pre-Covid and 12 articles during Covid, for a total of 21 articles (instead of 24), leading to a total sample size of 645 (instead of 648) to be analyzed.

on-screen popups, URL elements (for example, "https"), and words that were accidentally combined (such as "terribleand" rather than "terrible and"). We transformed our data into a corpus in a tidy format and confirmed it was clean of any stop words, punctuation, or numbers. We merged the data with the AFINN sentiment lexicon to depict the sentiment of each word. We performed parametric, non-parametric, robust, and Bayesian statistical tests to compare the probability distribution of sentiments across categories, especially over time. For visualization, the data are graphed with the y-axis representing the sentiment value word or author, and the x-axis representing the period. All files are kept in directories specific to their author, newspaper, and period. This allows us the flexibility to create new tables and graphs that compare the sentiment value with alternative categorical variables (such as the newspaper or author) for the x-axis.

3. Results

Using the robust Yuen's trimmed-mean and Mann-Whitney non-parametric tests as well as graphical methods, we were able to assess the statistical impact of the period (before or during COVID-19) on the probability distribution of the sentiment of words/tokens used in the media articles. The sentiment was calculated using the AFINN lexicon, ranging from -5 to 5, for each of the 307,832 words. Before the COVID-19 pandemic, the trimmed sentiment mean was 0.478 (with a median of +1 for the non-parametric test). During the pandemic, the trimmed sentiment mean declined to -0.188 (with a median of -1). The probability of this decline occurring by chance, under the null hypothesis of equal distribution of word sentiments before and during COVID-19, is negligible ($p<0.0001$). This finding is statistically significant regardless of the method used. The t-statistic was 92.4 using the robust method ($p<0.0001$), and the W-statistic from the non-parametric Mann-Whitney test was 9.48e+09 ($p<0.0001$) with a rank biserial correlation of -0.19 [99% CI: -0.19, -0.18]. Figure 1 shows the distribution of words used (or more precisely, tokens) by their sentiment value before and during the COVID-19 pandemic and a statistical summary of the results of the Mann-Whitney non-parametric test.

Table 1. *List of 27 Journalist Studied, Their Three Corresponding Journals (NYT, WSJ, and WP) and Main Topics, Lending 648 Articles.*

Journalist	The New York Times	The Wall Street Journal	The Washington Post	General Topic	Profile Link
Andrea Petersen		✓		Consumer Health and Mental Health	https://www.wsj.com/news/author/andrea-petersen
Andrea Sachs			✓	Travel reporting	https://www.washingtonpost.com/people/andrea-sachs/
Andy Kessler		✓		Tech, Markets, and Culture	https://www.wsj.com/news/author/andy-kessler
Carl Zimmer	✓			Science, Life, Biology, and Evolution	https://www.nytimes.com/by/carl-zimmer
Cathy Free			✓	Stories About Humanity for Inspired Life	https://www.washingtonpost.com/people/cathy-free/
Christopher Mims		✓		Technology. Various Subjects	https://www.wsj.com/news/author/christopher-mims
David Ignatius			✓	Foreign Affairs	https://www.washingtonpost.com/people/david-ignatius/
Emily Wax-Thibodeaux			✓	Breaking News and Social Movements	https://www.washingtonpost.com/people/emily-wax-thibodeaux/
Eric Sylvers		✓		Economics and Politics	https://www.wsj.com/news/author/eric-sylvers
Erik Wemple			✓	Cable News Industry	https://www.washingtonpost.com/people/erik-wemple/
Ivan Nechepurenko	✓			Russia, Ukraine, Belarus, and wider Region	https://www.nytimes.com/by/ivan-nechepurenko
Jason Gay		✓		Sports and Humor	https://www.wsj.com/news/author/jason-gay
Jennifer Schuessler	✓			Culture, Intellectual Life, and Ideas	https://www.nytimes.com/by/jennifer-schuessler

Journalist	The New York Times	The Wall Street Journal	The Washing-ton Post	General Topic	Profile Link
Jesse Wegman	✓			Supreme Court, Law and Politics.	https://www.nytimes.com/by/jesse-wegman
Karin Brulliard			✓	American West	https://www.washingtonpost.com/people/karin-brulliard/
Kwame Anthony Appiah	✓			Ethical Dilemmas and Moral Issues	https://www.nytimes.com/by/kwame-anthony-appiah
Lawrence Ulrich	✓			Automotive Industry and Electric Vehicles	https://www.nytimes.com/by/lawrence-ulrich
Linda Greenhouse	✓			Supreme Court, Law and Justice.	https://www.nytimes.com/by/linda-greenhouse
Marc A. Thiessen			✓	Foreign and Domestic Policy	https://www.washingtonpost.com/people/marc-a-thiessen/
Min Joo Kim			✓	South and North Korea	https://www.washingtonpost.com/people/min-joo-kim/
Newley Purnell		✓		Global Technology	https://www.wsj.com/news/author/newley-purnell
Rolfe Winkler		✓		Digital Health and Technology	https://www.wsj.com/news/author/rolfe-winkler
Saabira Chaudhuri		✓		Consumer Goods and Sustainability	https://www.wsj.com/news/author/saabira-chaudhuri
Spencer S. Hsu			✓	Homeland Security and Immigration	https://www.washingtonpost.com/people/spencer-s-hsu/
Stephanie Armour		✓		Healthcare Policy	https://www.wsj.com/news/author/stephanie-armour
Steve Lohr	✓			Technology, the Economy and Work	https://www.nytimes.com/by/steve-lohr
Tejal Rao	✓			Food, Food Culture, Restaurants	https://www.nytimes.com/by/tejal-rao

A closer look at the author-specific statistical analysis yields similar results to those in Figure 1, but from a different perspective that uses words within each author's articles and their corresponding sentiment. The median sentiment for the years 2018-2019 was 0.69, and it declined to -0.17 in the period 2020-2023. The probability of this decline occurring by chance was almost zero ($p = 6.65e-06$ using the non-parametric Wilcoxon test and $p = 5.26e-10$ using a standard parametric t-test). Moreover, the decline in sentiment was large in magnitude. The 99% confidence interval for the median decline is [-1.00, -0.98], and the mean decline was even larger, with a 99% CI given by [-2.39, -1.18]. Figure 2 illustrates these results. Each dashed line represents an author, showing their average sentiment pre-COVID (left violin/boxplot) and during Covid (right violin/boxplot). All authors under consideration experienced a significant decline in their average token sentiment with the exception of Jesse Wegman, who stayed neutral. The solid line illustrates the median decline from 0.69 to -0.17. A summary of the non-parametric Wilcoxon sign test results is also provided in Figure 2.

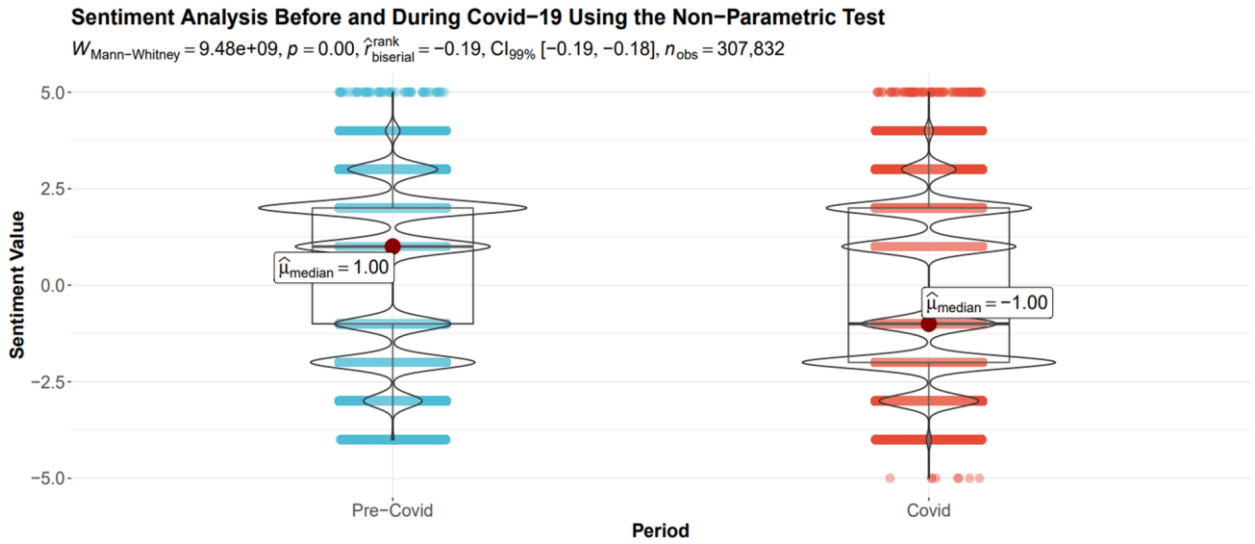

Figure 1. *Sentiment of Words Before and During COVID-19, and Statistical Analysis.*

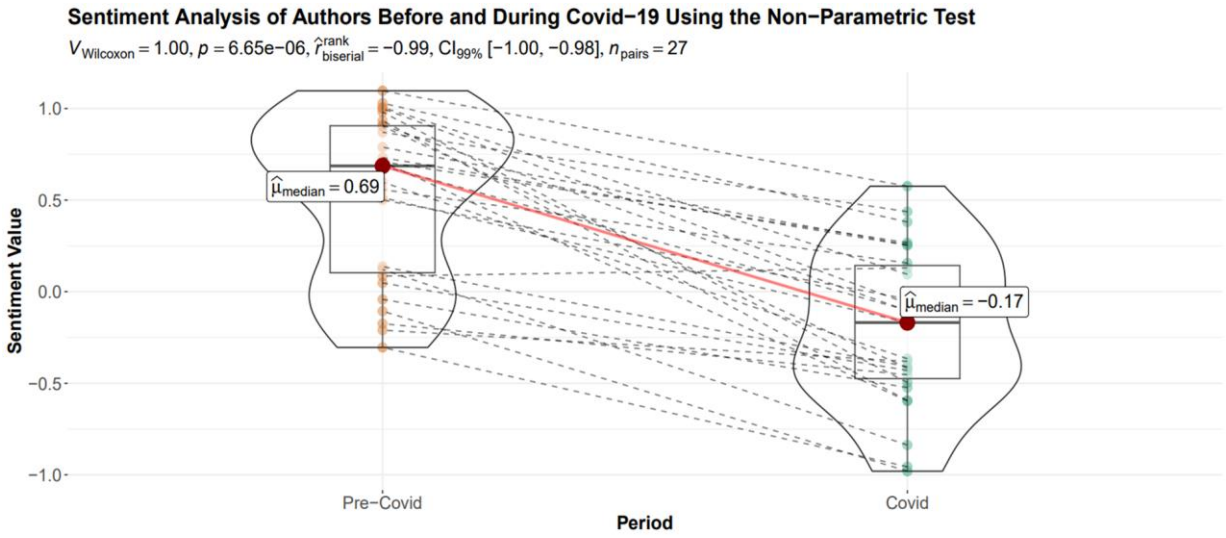

Figure 2. *Statistical Analysis of Sentiment Change Within Authors' Article Before and During COVID-19.*

4. Discussion and Limitations

This study documents a shifting language pattern in news reporting during the COVID-19 pandemic. While effort was made to keep the analysis objective and unbiased, the data may still reflect our choice of authors, topics, and articles included in the analysis. The technique selection could also introduce subjectivity based on our research expertise and the perceived suitability of the methods. Given that human choices are inherently non-random, it is difficult to ensure that our choice of data and methods are exempt from biases and to represent accurately the entire population of authors, topics, and articles. The availability of a larger and more comprehensive dataset of widespread articles and topics could increase the scope and representation of this study. In addition, our sentiment analysis is applied to articles as the unit of measurement without consideration of the reach of those articles or the proportion of the population who read them. Finally, our sentiment analysis does not consider tokens with unknown or missing sentiments or spelling variations, limiting the number of usable words in our analysis to those with explicitly known sentiments.

5. Conclusions and Future Research

The results of this study show a statistically significant decrease in the average sentiment of articles published by three popular U.S. newspapers during the COVID-19 pandemic. The decline was systematic among newspapers and journalists studied. A more negative narrative inevitably affects how readers see the world around us, fostering a more pessimistic lens. This potential change in widespread sentiment and attitude could be linked to the rise in mental illnesses, such as depression and anxiety, recorded during the pandemic [1]. Journalists are not

exempt from society's attitudes, which may permeate their work. However, they must be aware of the negative impacts of perpetuating these sentiments. Their failure to do so should at least call into discussion their moral obligations and social responsibilities, especially in times of need.

Abbreviations

COVID-19	Coronavirus Disease 2019
AFINN	English Sentiment Lexicon Developed by Finn Årup Nielsen
NRC	Emotion Lexicon Developed at the National Research Council Canada (NRC) by Saif Mohammad and Peter Turney

Supplementary Material

The supplementary material can be accessed at https://doi.org/10.11648/j.ijdst.20241002.13

Acknowledgments

The authors gratefully acknowledge the support of Ursinus College, its Department of Mathematics, Computer Science and Statistics, the Summer 2023 Ursinus Research Experience with Undergraduates (REU) and Summer Fellows programs, and the Inclusive Data Science and Innovative Research Initiative (IDSIRI). We also extend our thanks to the Mathematical Association of America (MAA) community, particularly those at MathFest 2023, for their valuable feedback.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Z. Zhang, D. Robinson, and J. Tepper, "Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network," in The Semantic Web, A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, and M. Alam, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 745–760. https://doi.org/10.1007/978-3-319-93417-4_48

[2] S. Sharma and A. Sharma, "Twitter Sentiment Analysis During Unlock Period of COVID-19," in 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Nov. 2020, pp. 221–224. https://doi.org/10.1109/PDGC50313.2020.9315773

[3] P. Tiwari, B. K. Mishra, S. Kumar, and V. Kumar, "Implementation of n-gram Methodology for Rotten Tomatoes Review Dataset Sentiment Analysis," Int. J. Knowl. Discov. Bioinforma. IJKDB, vol. 7, no. 1, pp. 30–41, 2017, https://doi.org/10.4018/IJKDB.2017010103

[4] T. Wang, K. Lu, K. P. Chow, and Q. Zhu, "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model," IEEE Access Pract. Innov. Open Solut., vol. 8, pp. 138162–138169, 2020, https://doi.org/10.1109/ACCESS.2020.3012595

[5] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning," Entropy, vol. 21, no. 6, p. 589, 2019, https://doi.org/10.3390/e21060589

[6] M. Ojeda-Hernández, D. López-Rodríguez, and Á. Mora, "Lexicon-based sentiment analysis in texts using Formal Concept Analysis," International Journal of Approximate Reasoning, vol. 155, pp. 104–112, 2023, https://doi.org/10.1016/j.ijar.2023.02.001

[7] O. Alsemaree, A. Alam, S. Gill, and S. Uhlig, "An analysis of customer perception using lexicon-based sentiment analysis of Arabic Texts framework," Heliyon, vol. 10, no. 11, p. e30320, 2024, https://doi.org/10.1016/j.heliyon.2024.e30320

[8] V. Singh, G. Singh, P. Rastogi, and D. Deswal, "Sentiment Analysis Using Lexicon Based Approach," Fifth International Conference on Parallel Distributed and Grid Computing (PDGC), pp. 13–18, 2018, https://doi.org/10.1109/PDGC.2018.8745971

[9] F. Xing and J. Zhan, "Sentiment analysis using product review data," Journal of Big Data, vol. 2, no. 5, 2015, https://doi.org/10.1186/s40537-015-0015-2

[10] T. Rao and S. Srivastava, "Analyzing stock market movements using Twitter sentiment analysis," *IEEE Computer Society*, 2012, https://doi.org/10.1109/ASONAM.2012.30

[11] Fisher E., Fitzgibbon M., Glasgow R., Haire-Joshu D., Hayman L., Kaplan R., Nanney MS, Ockene JK. Behavior matters. Am J Prev Med. 2011, 40 (5), 15-30. https://doi.org/10.1016/j.amepre.2010.12.031

[12] Cutler W, Kolter J, Chambliss C, O'Neill H, Montesinos-Yufa HM. Long term absence of invasive breast cancer diagnosis in 2,402,672 pre and postmenopausal women: A systematic review and meta-analysis. PLoS One. 2020, Sep 10; 15 (9): e0237925. https://doi.org/10.1371/journal.pone.0237925 PMID: 21496745; PMCID: PMC3137947.

[13] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, "Natural language processing applied to mental illness detection: a narrative review," Npj Digit. Med., vol. 5, no. 1, Art. no. 1, Apr. 2022, https://doi.org/10.1038/s41746-022-00589-7

[14] J. Gruber et al., "Mental health and clinical psychological science in the time of COVID-19: Challenges, opportunities, and a call to action.," Am. Psychol., vol. 76, no. 3, p. 409, 20200810, https://doi.org/10.1037/amp0000707

[15] R. Weger et al., "Trends in Language Use During the COVID-19 Pandemic and Relationship Between Language Use and Mental Health: Text Analysis Based on Free Responses From a Longitudinal Study," JMIR Mental Health, 2023, https://doi.org/10.2196/40899

[16] C. Vargas-Sierra and M. Á. Orts, "Sentiment and emotion in financial journalism: a corpus-based, cross-linguistic analysis of the effects of COVID," Humanities and Social Sciences Communications, vol. 10, no. 219, 2023, https://doi.org/10.1057/s41599-023-01725-8

[17] N. Kathirvel, "Post COVID-19 pandemic mental health challenges," Asian J. Psychiatry, vol. 53, p. 102430, Oct. 2020, https://doi.org/10.1016/j.ajp.2020.102430

[18] C. K. Ettman, S. M. Abdalla, G. H. Cohen, L. Sampson, P. M. Vivier, and S. Galea, "Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic," JAMA Netw. Open, vol. 3, no. 9, p. e2019686, Sep. 2020, https://doi.org/10.1001/jamanetworkopen.2020.19686

[19] R. Blades, "Protecting the brain against bad news," CMAJ, vol. 193, no. 12, pp. E428–E429, 2021, https://doi.org/10.1503/cmaj.1095928

[20] J. Kellerman, R. Shalaby, and S. Chokshi, "The Mental Health Impact of Daily News Exposure During the COVID-19 Pandemic: Ecological Momentary Assessment Study," JMIR Mental Health, vol. 9, no. 5, 2022, https://doi.org/10.2196/36966

[21] K. Stainback, B. Hearne, and M. Trieu, "COVID-19 and the 24/7 News Cycle: Does COVID-19 News Exposure Affect Mental Health?," Socius, vol. 6, 2020, https://doi.org/10.1177/2378023120969339

[22] M. Costola, O. Hinz, M. Nofer, and L. Pelizzon, "Machine learning sentiment analysis, COVID-19 news and stock market reactions," Research in International Business and Finance, vol. 64, 2023, https://doi.org/10.1016/j.ribaf.2023.101881

[23] C. Bai, Y. Duan, X. Fan, and S. Tang, "Financial market sentiment and stock return during the COVID-19 pandemic," Finance Research Letters, vol. 54, 2023, https://doi.org/10.1016/j.frl.2023.103709

[24] Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291. https://doi.org/10.2307/1914185

[25] Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as Feelings. *Psychological Bulletin*, 127(2), 267-286. https://doi.org/10.1037/0033-2909.127.2.267

[26] Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance*, 61(4), 1645-1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

[27] Nofsinger, J. R. (2005). Social mood and financial economics. *Journal of Behavioral Finance*, 6(3), 144-160. https://doi.org/10.1207/s15427579jpfm0603_4

[28] Lim, H., Kim, J., & Lee, J. (2017). Financial worries and financial behaviors among college students: The mediating role of self-control. *Journal of Financial Counseling and Planning*, 28(1), 70-80.

[29] Xiao, J. J., & O'Neill, B. (2012). The impact of financial stress on the academic achievement of young adults: Evidence from a natural experiment. *Journal of Family and Economic Issues*, 33(2), 189-201.

[30] Coleman, E., Innocent, J., Kircher, S., Montesinos-Yufa, H., & Trauger, M. (2024). *A Pandemic of Mental Health* (Unpublished working paper). Ursinus College.

[31] Montesinos-Yufa, H., & Nagasuru (2024). Gender-Specific Mental Health Outcomes in Central America: A Natural Experiment. International Journal of Data Science and Technology (Forthcoming).

[32] Brandt EN, Baird MA, Berkman LF, et al. Health and behavior: The interplay of biological, behavioral, and societal influences. Health. 2001.

[33] "Leading print newspapers U.S. by circulation 2022," Statista. https://www.statista.com/statistics/272790/circulation-of-the-biggest-daily-newspapers-in-the-us/ (accessed Jun. 15, 2023).