How did we get here? A framing and source analysis of early COVID-19 media coverage

Austin Hubner

Published online: 11 Mar 2021.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

How did we get here? A framing and source analysis of early COVID-19 media coverage

Austin Hubner ⓘ

ABSTRACT

This study examines how two major news outlets framed COVID-19 in the months leading up to COVID-19 being declared a pandemic. A combination of computational and manual coding found that the dominant media frames were outbreak, economic consequences, and social consequences. A secondary component of the analysis examined the sources quoted in early media coverage and found that citizens, interest group individuals, and politicians were quoted the most. This suggests that early news coverage of COVID-19 might have contributed to the current politicization of the virus in the United States.

"The story of COVID-19 in the United States is one of the strangest paradoxes of the whole pandemic. No other country has the concentration of scientific skill, technical knowledge, and productive capacity possessed by the US. It is the world's scientific superpower bar none. And yet this colossus of science utterly failed to bring its expertise successfully to bear on the policy and politics of the nation's response." (Horton, 2020, p. 46–47).

At the time of this writing, SARS-Co-V-2, otherwise known as COVID-19 has killed more than 1 million people worldwide and over 400,000 people in the United States (Mazzei, 2021). Despite the devastation that COVID-19 has generated, beliefs about the virus and its prevention measures are polarized in the United States. Indeed, 60% of conservatives believe that the common flu poses a greater risk to human life than COVID-19 (Ritter, 2020), and wearing a mask—an important preventative behavior—has become a partisan act (Lipsitz & Pop-Eleches, 2020). In order to disentangle how COVID-19 evolved from a public health crisis to a major political debate, we must investigate early news coverage of the outbreak as the media helps individuals contextualize unfamiliar risks through their framing of the issue (Abeysinghe & White, 2010; Dudo, Dahlstrom, & Brossard, 2007; Lee & Basnyat, 2013).

The news media plays a critical role in shaping public perceptions about a risk through their initial framing (Lee & Basnyat, 2013); where framing is the act of making a particular aspect of reality salient in a communication text (Entman, 1993). In the research reported here, communication texts are media

CONTACT Austin Hubner ✉ Hubner.9@osu.edu School of Communication, The Ohio State University.

frames defined as a "central organizing idea or story line that provides meaning to an unfolding strip of events [...] The frame suggests what the controversy is about, the essence of the issue" (Gamson & Modigliani, 1989, p. 143). Media frames are especially important for communicating complex, unfamiliar topics as they help individuals understand the threat the risk poses (Lin & Lagoe, 2013).

During an infectious disease outbreak, the news media serve as a mediator between public health officials and members of the public (Lee & Basnyat, 2013). As such, journalists are tasked with both delivering factual information to the public *and* framing it in a way that drives traffic to their articles (Schwitzer et al., 2005). This task is further complicated by the nature of outbreaks, where new information is being quickly uncovered by the expert community. As such, journalists—who often have neither a background in science nor health—must translate complicated, changing information to their audiences (Schwitzer et al., 2005). Unsurprisingly, journalists have been criticized for their coverage of past pandemics where they over-sensationalized H1N1 (Vasterman & Ruigrok, 2013) and politicized both Zika and Ebola virus (Singer, Willison, & Greer, 2020). Given the tendencies to politicize and over sensationalize infectious disease outbreaks, the following research question is posed:

RQ1: How was COVID-19 framed in early coverage of the outbreak?

When developing a story, journalists generally talk to a range of sources in order to help them make sense of the issue (Conrad, 1999). This is especially true for science and health information as journalists generally lack specialized training in the area and thus rely on their sources to help them contextualize the information (Schwitzer et al., 2005). During infectious disease outbreaks, scientific experts play a vital role as sources in the news (Shih et al., 2011). For example, Shih and colleagues (2011) found that public health experts from government organizations and the WHO were the most predominant sources in coverage about the West Nile virus. One might assume, therefore, that stories about COVID-19 would have largely relied on quotes from public health experts.

Yet, Tanner and Friedman (2010) found that online news about health rarely quoted health experts, instead it relied on quotes from nonpublic health sector government officials. An analysis of news coverage about the 2009 H1N1 outbreak, furthermore, found that news coverage included a variety of sources beyond public health officials and tended to favor politicians (Lee & Basnyat, 2013). Even more surprising, a recent analysis of media coverage following the WHO's declaration that COVID-19 had reached pandemic status found that a large number of politicians were quoted in the stories (Hart et al., 2020). This suggests that COVID-19 had already become politicized as the news media was relying on quotes from politicians rather than public health experts. This leads to an important

question of whether initial coverage of the virus was supported by quotes from public health officials *or* was it politicized from the start?

RQ2: Who was quoted in early media coverage of COVID-19?

Method

The results reported here are based on a combination of manual and computational coding. The sample consists of two national news outlets, *The New York Times* and *The Washington Post* from when the news of the virus first broke on January 8, 2020, to when the WHO declared COVID-19 a pandemic on March 11, 2020. The two news outlets were selected due to their large circulation within the United States (Agility, 2020). Relevant articles were collected through Lexis-Nexis by using the search terms "COVID-19" OR "coronavirus." The initial search yielded 3148 articles. Because this research focuses on COVID-19, only articles where the virus was mentioned in the headline and/or lead were included. The final analysis excluded editorials, opinion columns, and advertisements. After removing duplicate and irrelevant articles, the final sample consisted of 1034 articles.

Frame sponsor

Frame sponsors were operationalized as quoted sources. Every individual quoted in the article was recorded, along with their position and organization. The frequency of which the source was quoted in an article was not recorded. The source's organization affiliations broadly fell into the following categories: universities, government agencies (local, federal, and international), interest groups, stakeholders, industry research groups, celebrities, religious organizations, hospitals, public health agencies, and private citizens. The sources were aggregated into 10 categories: Academic, citizens, entertainment industry, government, industry researcher, interest group individual (e.g., interest group, think tank), medical industry (doctors, nurses, hospitals), public health official, politicians, and other.

Frame detection

Frames were identified through a two-step, inductive-deductive approach. Topic modeling was used in the first step to split the corpus of articles into separate "topics." Topic modeling uses a "bag of words" approach to split a corpus of text into individual words and then splits the corpus into topics based upon the frequency in which words appear in one topic exclusive to another topic (Blei, Ng, & Jordan, 2003). More specifically, this study uses Latent Dirichlet Allocation (LDA; for a complete description of LDA and its assumptions see Blei et al., 2003) using the "LDA" and "LDAtuning"

packages for R (Murzintcev & Chaney, 2020). LDA is guided by two assumptions: every document is a mixture of topics and every topic is a mixture of words (Blei et al., 2003).

Following others, only the headline, lead, and first paragraph of each article were included in the model (e.g., Nicholls & Culpepper, 2020). Before running the model, the text was pre-processed using the guidelines set by Maier et al. (2018)—removing stop words, punctuation, numbers, and words that appeared in more than 99% of the documents and in less than 5% of the documents—with one notable exception, the text was not stemmed as it threatens topic interpretation (Maier et al., 2018). After preprocessing, the model was estimated for k-values (topics) ranging from 5 to 300. After visually inspecting the output, the model was rerun for k-values 5 to 75. The final k-value was chosen where the indicators of accuracy and density were minimized, and the Bayesian Markov chain Monte Carlo algorithm was maximized.

Next, the topics were interpreted by examining two types of information: the words that are both the most prevalent and exclusive to each topic (FREX words) and the entire corpus of documents for each topic. Walter and Ophir (2019) contend that while topic modeling is a good starting point for frame detection, other methods should be used to group the topics into frame categories. Therefore, the topics were grouped into frames based upon those detected in previous framing studies. In order to validate the frame assignment, the author developed a codebook and an independent coder coded 10% of the articles to compare against the frame assignment based upon the topic modeling. The Cohen's Kappa score ($\alpha=.76$) indicated substantial agreement.

Results

Based on the statistical indicators described above, a model of 30 topics was selected. After inspection of the FREX words and reading the headlines of the articles assigned to each topic, 10 topics were dropped from further interpretation as their topic assignment was uninterpretable. Thus, 20 topics were labeled (see Table 1). After assigning each topic a label, the topics were manually grouped into clusters which were assigned a frame name based upon previous scholarship. This process resulted in the identification of seven frames. An independent coder assigned articles from the uninterpretable topics to a frame.

The frame with the highest prevalence was outbreak ($n= 397$) which incorporated articles tracking the initial outbreak in China, its subsequent spread to other countries, the case count and death toll, and reports of epicenters of outbreak (e.g., cruise ships, New York City, Washington State). The frame economic consequences ($n= 149$) included articles about stock market projections, the Federal Relief bill, and the global supply chain.

Table 1. Topic classification and keywords.

Frame, topic, FREX words, *n*
Frame 1: Economic consequences
• Topic 14: Federal Reserve (*n*= 48)
• Economic, Federal Reserve, global, central, economy
• Topic 27: Economic Effects (*n*= 37)
• Economy, oil, supply, world's, China, companies
• Topic 28: Stock Market (*n*= 52)
• Fears, markets, global, investors, markets
Frame 2: Social consequences
• Topic 16: Canceled events (*n*= 45)
• Canceled, games, events, school, concerts
• Topic 17: Education effects (*n*= 40)
• Schools, students, close, announced, university
• Topic 19: Workforce (*n*= 20)
• Home, workers, sick, many, take
Frame 3: Medical/Expert Response
• Topic 3: Controlling spread (*n*= 16)
• Disease, can, say, control, experts
• Topic 4: Health officials (*n*= 33)
• Health officials, public health officials, growing, halt, medical
• Topic 15: Fighting COVID-19 (*n*= 45)
• Patients, hospital, medical, center, care
• Topic 38: Open questions (*n*= 28)
• People, infected, least, illness, virus
Frame 4: Medical supplies
• Topic 24: Testing (*n*= 30)
• Testing, test, positives, California, days
• Topic 37: Prevention equipment (*n*= 31)
• Face masks, empty, hand sanitizer, supplies, show
• Topic 40: Shortages (*n*= 8)
• Virus, coronavirus, risk, drug, stop
Frame 5: Political response
• Topic 30 (*n*= 55)
• Trump, president, administrative, response, Pence
• Topic 23: Leader's response (*n*= 30)
• China, epidemic, crisis, government, leader
• Topic 20: Relief Bill (*n*= 26)
• White House, emergency, response, billion
Frame 6: Misinformation
• Coronavirus, spread, media, social, misinformation (*n*= 17)
Frame 7: Outbreak
• Topic 2: Restrictions (*n*= 29)
• Travel, countries, across, Europe, across
• Topic 5: Tracking outbreak (*n*= 35)
• Cases, new, confirmed, reported, infections
• Topic 6: Cruise ship (*n*= 56)
• Cruise ship, passengers, Japan, quarantine, Americans
• Topic 8: Preparing for a global outbreak (*n*= 29)
• World, around, now, pandemic, organization
• Topic 10: Cases and deaths (*n*= 39)
• First, virus, case, death, man, patient
• Topic 11: Reacting to China (*n*= 45)
• China, Chinese, Wuhan, authorities, city
• Topic 12: Chinese spread (*n*= 8)
• Outbreak, country, way, Province, spread
• Topic 13: New York outbreak (*n*= 30)
• New York, city, day, woman, Cuomo
• Topic 18: Global spread (*n*= 40)
• South Korea, North Korea, Italy, Milan, country's
• Topic 21: Hong Kong response (*n*= 26)
• China, Hong Kong, outbreak, coronavirus, trade
• Topic 25: Outbreak (*n*= 11)
• Spread, coronavirus, million, even, outbreak
• Topic 26: Washington State outbreak (*n*= 15)
• State, coronavirus, Washington, department, amid
• Topic 31: US response (*n*= 15)
• United States, American, novel, flights, vast

Similarly, the social consequence frame ($n = 137$) discussed the social impacts on human life such as school shutdowns, lockdowns, working from home, and anxiety surrounding COVID-19. This was followed by the frame medical/ expert response ($n = 131$) which included articles with public health official's advice, reports of frontline medical workers, and experts searching for answers about the virus. The political response frame ($n = 123$) included articles about world leaders' responses to COVID-19. The medical supply frame ($n = 71$) dealt with articles about shortages in supplies (e.g., masks and testing) as well as questions about the efficacy of preventative behaviors. Finally, a small subset of articles framed COVID-19 in terms of the spread of false information ($n = 24$).

Moving to the second research question, there were 3410 unique sources quoted in early media coverage. After aggregation, citizens were quoted the most ($n = 927$), followed by interest group individuals ($n = 722$), politicians ($n = 712$), and academics ($n = 649$; see Table 2). Of each unique source, President Trump was quoted in the most articles ($n = 137$), followed by the director of the WHO, Tedros Adhanom Ghebreyesus ($n = 53$), Anthony Fauci ($n = 45$), and Vice President Mike Pence ($n = 44$).

In order to gain a more nuanced understanding of the sources quoted, a secondary examination was conducted which looked at the number of times the news outlets quoted representatives from the CDC and the WHO. The CDC was quoted in 143 articles while the WHO was quoted in 124 articles. Following the same line of inquiry, the analysis examined what fields the academics quoted resided in. Unsurprisingly, the majority of experts represented individuals in virology, epidemiology, and public health ($n = 382$), but this was followed by political scientists ($n = 113$) and social scientists ($n = 34$).

Discussion

Individuals often turn to the news media to make sense of new, unfamiliar risks. As such, the news media's initial framing of a risk is vital as it likely influences individual attitudes and risk-perceptions. Given that COVID-19 became a highly polarized topic within the U.S. this research sought to understand how the news media initially framed COVID-19. The framing analysis revealed that early coverage framed it largely in terms of tracking the outbreak as it spread from Wuhan, China to other parts of the world. Indeed, a lot of attention was spent delving into case counts and growing death tolls as major epicenters of outbreak cropped up. This was followed by the economic consequences of the virus, as experts projected a crashing stock market and major impacts on the world's supply chain. The third most prevalent frame was the societal consequences of COVID-19, which delved into shutdowns and the effects of those shutdowns on everyday life.

Table 2. Source type and frequency of quoted statements.

Source type	Frequency
Citizen	927
Interest group individual	722
Politician	712
Academic	649
Government official	588
Industry researcher	586
Public health official	573
Medical industry	238
Entertainment industry	137
Other	163

Taken together, this shows that early news coverage was focused on two aspects of COVID-19: its spread and subsequent detrimental effect on society. The focus on societal effects, rather than information regarding the health risks and expert community response, might be because the expert community was still largely uncertain about the cause, transmission, and subsequent health effects of COVID-19. Additionally, the WHO was slow to act on COVID-19, not wanting to spark panic amongst the public (Horton, 2020). As such, news media attention turned to individual consequences, such as business closures, rather than the health effects of COVID-19. This focus might have grounded COVID-19 as a threat to one's way of life *rather* than a serious threat to one's health.

The pattern of framing COVID-19 in terms of societal impacts was reiterated when examining quoted sources since citizens prevailed over all other source types. This pattern was further cemented as interest group individuals and politicians were quoted more frequently than academics and public health experts. The reliance on politicians rather than scientific experts and public health officials continued in the months *after* COVID-19 was declared a pandemic, suggesting a troubling pattern where the health effects of COVID-19 had yet to reach the forefront of news coverage (Hart et al., 2020). This finding is markedly different than previous outbreaks where public health and government officials were the dominant sources used by the news media.

While this study is important for our understanding of the foundation of current risk perceptions related to COVID-19, there are three limitations to note. First, this analysis only examined the first 3 months of COVID-19 news coverage. Therefore, more research is needed to understand how media frames have evolved during the course of the pandemic. Second, the framing analysis utilized topic modeling to identify topics within the sample of news coverage. The topics were then grouped into seven frames. Although previous framing scholarship was consulted and descriptions of previously identified frames were adapted to fit the current context, this could have introduced potential researcher-bias into the frame detection process. This limitation was partially

mitigated by the development of a codebook and having an independent coder manually assign frames to a subset of the news articles in order to test the validity of the computation finding. Finally, this study only examined two news outlets that are slightly left leaning. While both outlets have a large number of subscribers within the US, the inclusion of other outlets might have changed the findings of this study. Therefore, future research should examine potential framing differences between news outlets.

This study helps provide insight into an ongoing public health crisis by examining how the mass media initially framed COVID-19. More research is needed, however, in order to fully understand the important role that the mass media played in the COVID-19 pandemic. Future work, therefore, should build upon this study by comparing how different media outlets have framed the pandemic as well as the role of different framing devices such as the visuals used in the news coverage.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Austin Hubner (M.A., The Ohio State University) isa doctoral candidate in the School of Communication at The OhioState University.

ORCID

Austin Hubner ⓘD http://orcid.org/0000-0001-6495-1196

References

Abeysinghe, S., & White, K. (2010). Framing disease: The avian influenza pandemic in Australia. *Health Sociology Review*, 19(3), 369–381.

Agility (2020). Top 10 U.S. newspapers by circulation. https://www.agilitypr.com/resources/top-media-outlets/top-10-daily-american-newspapers/

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022.

Conrad, P. (1999). Uses of expertise: Sources, quotes, and voice inreporting of genetics in the news. *Public Understanding of Science*, 8, 285–302.

Dudo, A. D., Dahlstrom, M. F., & Brossard, D. (2007). Reporting a potential pandemic: A risk-related assessment of avian influenza coverage in US newspapers. *Science Communication*, 28(4), 429–454.

Entman, R. M. (1993). Framing: Toward a fractured paradigm. *Journal of Communication*, 43 (4), 51–58.

Gamson, W. A., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, 95(1), 1–37.

Hart, P.S., Chinn, P., & Soroka, S. (2020). Politicization andpolarization in COVID-19 news coverage. Science Communication,52(5), 679–697

Horton, R. (2020). *The COVID-19 catastrophe: What's gone wrong and how to stop it happening again*. Medford, MA: John Wiley & Sons.

Lee, S. T., & Basnyat, I. (2013). From press release to news: Mapping the framing of the 2009 H1N1 A influenza pandemic. *Health Communication*, 28(2), 119–132.

Lin, C. A., & Lagoe, C. (2013). Effects of news media and interpersonal interactions on H1N1 risk perception and vaccination intent. *Communication Research Reports*, 30(2), 127–136.

Lipsitz, K., & Pop-Eleches, G. (2020). The partisan divide in social distancing. *Available at SSRN 3595695*.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., . . . Schmid-Petri, H. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2–3), 93–118.

Mazzei, P. (2021, January 19). The virus death toll in the U.S. has passed 400,000. *The New York Times*. https://www.nytimes.com/2021/01/19/us/coronavirus-deaths-usa-400000.html

Murzintcev, N., & Chaney, N. (2020). ldatuning: Tuning of the latent Dirichlet allocation models parameters (Version 1.0.0) [Computer software]. https://cran.rproject.org/web/packages/ldatuning/index.html

Nicholls, T., & Culpepper, P. D. (2020). Computational identification of media frames: Strengths, weaknesses, and opportunities. *Political Communication*, 1–23.

Ritter, Z. (2020). Republicans still skeptical of COVID-19 lethality. https://news.gallup.com/poll/311408/republicans-skeptical-covid-lethality.aspx

Schwitzer, G., Mudur, G., Henry, D., Wilson, A., Goozner, M., Simbra, M., . . . Baverstock, K. A. (2005). What are the roles and responsibilities of the media in disseminating health information? *PLoS Medicine*, 2(7), e215.

Shi,T.J., Brossard, D., & Wijaya, R. (2011). News coverage of publichealth issues: The role of news sources and the processes of newsconstruction. In Merrick, J. (Ed.), *Public Health Yearbook* (pp. 93-106). Nova Science Publishers, Inc.

Singer, P. M., Willison, C. E., & Greer, S. L. (2020). Infectious disease, public health, and politics: United States response to Ebola and Zika. *Journal of Public Health Policy*, 41, 1–11.

Tanner, A. & Friedman, D.B. (2010). Authorship and informationsourcing for health news on local TV web sites: An exploratoryanalysis. *Science Communication*, 33(1), 3–27

Vasterman, P. L., & Ruigrok, N. (2013). Pandemic alarm in the Dutch media: Media coverage of the 2009 influenza A (H1N1) pandemic and the role of the expert sources. *European Journal of Communication*, 28(4), 436–453.

Walter, D., & Ophir, Y. (2019). News frame analysis: An inductive mixed-method computational approach. *Communication Methods and Measures*, 13(4), 248–266.