Communication

Prevalence in News Media of Two Competing Hypotheses about COVID-19 Origins

David Rozado

Otago Polytechnic, Dunedin 9054, New Zealand; david.rozado@op.ac.nz

Abstract: The COVID-19 pandemic has been one of the most disruptive and painful phenomena of the last few decades. As of July 2021, the origins of the SARS-CoV-2 virus that caused the outbreak remain a mystery. This work analyzes the prevalence in news media articles of two popular hypotheses about SARS-CoV-2 virus origins: the natural emergence and the lab-leak hypotheses. Our results show that for most of 2020, the natural emergence hypothesis was favored in news media content while the lab-leak hypothesis was largely absent. However, something changed around May 2021 that caused the prevalence of the lab-leak hypothesis to substantially increase in news media discourse. This shift has not been uniformed across media organizations but instead has manifested itself more acutely in some outlets than others. Our structural break analysis of daily news media usage of terms related to the laboratory escape hypothesis provides hints about potential sources for this sudden shift in the prevalence of the lab-leak hypothesis in prestigious news media.

Keywords: COVID-19; coronavirus; SARS-CoV-2; lab-leak; news media

1. Introduction

The COVID-19 pandemic has caused enormous amounts of human suffering worldwide. As of July 2021, the origins of the SARS-CoV-2 virus that caused the outbreak remain unknown. There are two popular competing hypotheses about such origin. One asserts that the virus probably leaped naturally from wildlife to people (Calisher et al. 2020; Andersen et al. 2020). The other proposes that the virus might have accidentally leaked from a biolab (Bloom et al. 2021; Wade 2021). No conclusive evidence for either hypothesis has yet been uncovered. Determining which hypothesis is correct is critical to prevent a similar outbreak reoccurring again in the future with its associated catastrophic loss of human life.

The author of this work noticed a sudden increase in media mentions of the lab-leak hypothesis around mid-2021. Obviously, anecdotal evidence based on individual subjective perceptions could be the result of cognitive biases. Thus, the author carried out a quantitative analysis of news media content aimed at comparing the temporal dynamics of the two competing SARS-CoV-2 origin hypotheses. This work reports the results of that analysis and provides convenient visualizations about news media chronological coverage of the two alternative conjectures.

Computational content analysis of large bodies of text can be illuminating to elucidate the semantic associations embedded in the text (Rozado and Al-Gharbi 2021; Rozado 2020b). Simply charting word frequencies in a diachronic corpus of news content tracks the time course of historical events and highlights the dynamics of social trends within the cultural context in which the texts were produced (Rozado 2020a; Rozado et al. 2021).

Figure 1 illustrates the validity of our method by tracking sociocultural phenomena related to the COVID-19 pandemic. The first row shows the temporal dynamics of the different terms used to refer to the virus that caused the pandemic. The second row displays the shifting attention paid by news media towards several techniques attempted at alleviating the havoc caused by the virus (Jorge 2021). The third row illustrates how

the media reflected the climate of social *fear*, peaking around March 2020, as the severity of the virus became clear. The subsequent concern with the mounting *death* toll likely prompted, media echoed, mandates for *lockdowns, masks*, and *social distancing* protocols. The last row of Figure 1 shows how in the critical time around February–March 2020, the theme of *freedoms* and *civil liberties* became less prominent in news media written articles, probably due to the urgency of containing the virus. This theme however, rebounded in prevalence between May and July of the same year, perhaps prompted by concerns about overreaching mandates to stop the virus. Figure 1 o illustrates how the media reflected the *unemployment* damage caused by the pandemic and subplot p tracks the occurrence of the *anti-vaxxers* theme.

News Media Aggregate Monthly Prevalence of Words Related to the COVID-19 Pandemic
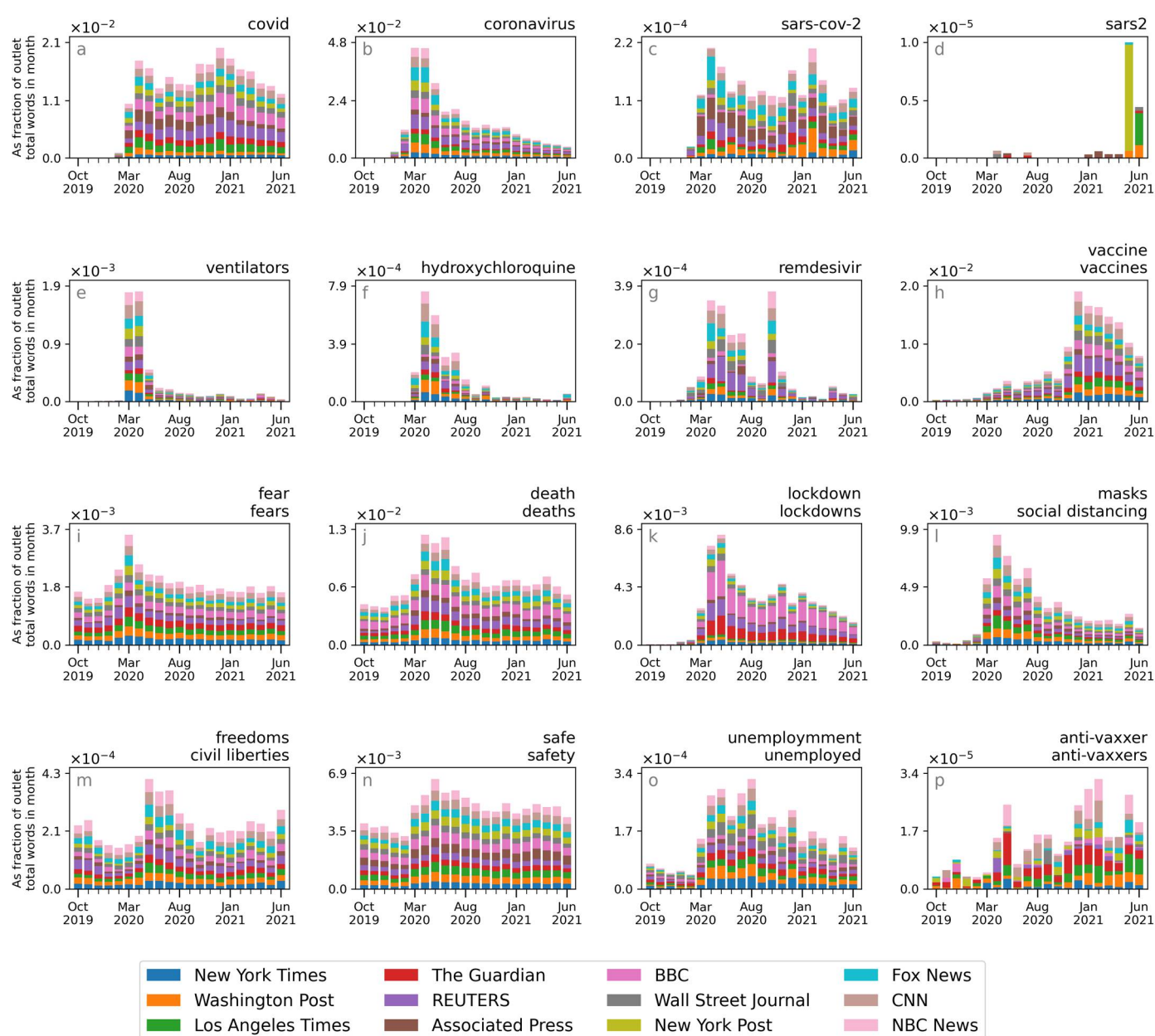### October 1, 2019 - June 30, 2021

Figure 1. Monthly aggregate frequency of terms related to the COVID-19 pandemic in popular news media outlets.

News media is supposed to provide their audiences with the facts and information that said audiences need to understand current events. In a recent UK survey, most people felt that news media organizations helped them respond to the COVID-19 crisis, but a third of respondents believed that news coverage made the crisis worse (Nielsen 2020). Previous work has investigated the role of news media on COVID-19 misperceptions (Bridgman et al. 2020). Other work has reported declining public trust in news media reporting about COVID-19 (Fletcher et al. 2020). The role of news media in terms of agenda-setting with respect to COVID-19 vaccination has also been studied (Medina et al. 2021).

Agenda-setting theory (McCombs and Shaw 1972) studies how news coverage of events shapes the formation of public opinion (McCombs 2018). Previous research has investigated whether news media influences consumers of said media by establishing a hierarchy of news thematic prevalence that filters information streams and shapes audiences perceptions of current events (Mrogers and Wdearing 1988). This work explores the prevalence in news media content of two competing hypotheses about COVID-19 origins. Interpreting the results through the lens of agenda-setting theory can provide insight into how thematic prominence in news outlets of a given hypothesis about SARS-CoV-2 origins has the potential to shape audiences' perceptions about the causal roots of a major and dramatic event such as the COVID-19 pandemic.

2. Methods

The textual content of news and opinion articles from the outlets listed in Figure 1 is available in the outlets online domains and/or public cache repositories such as Google cache, The Internet Wayback Machine (Notess 2002) and Common Crawl (Mehmood et al. 2017). Textual content included in our analysis is circumscribed to the articles' headlines and main text and does not include other article elements such as figure captions. This work has not analyzed video or audio content of news media organizations, except when an outlet explicitly provides a transcript of such content in article form. Targeted textual content was located in HTML raw data using outlet-specific XPath expressions. Tokens were lowercased prior to estimating frequency counts.

Frequency usage of a target word or n-gram in an outlet for any given temporal interval (monthly, weekly or daily) was estimated by dividing the number of occurrences of the target word/n-gram in all articles within a given interval by the total number of all words in all articles within that interval. This method of estimating frequency accounts for variable volume of total article output over time.

Latent associations in news media content were measured using embedding models. Reliable word embeddings require substantial amounts of textual data to produce robust results. Thus, we derived word embedding models for each month between October 2020 and June 2021 from all the combined outlets monthly content. The gensim (Řehůřek and Sojka 2010) implementation of word2vec with the continuous bag of words (CBOW) architecture setting was employed to train the embedding models.

Prior to estimating word embedding models, tokens were lowercased. Markup language tags, URLs, non-alphanumeric characters, punctuation, and multiple spaces were removed before training the embeddings models.

For training the word embedding models, the following parameters were used: vector dimensions = 300, window size = 10, negative sampling = 10, down sampling frequent words = 0.0001, minimum frequency count = 5 (only terms that appear more than 5 times in the corpus were included into the word embedding model vocabulary), and number of training iterations (epochs) through the corpus = 5. The exponent used to shape the negative sampling distribution was the default 0.75.

3. Results

3.1. Prevalence of Two Alternative COVID-19-Origins Hypotheses in News Media

We now focus on analyzing news media treatment of the two competing hypotheses regarding the pandemic origin. The first officially acknowledged signs of COVID-19

surfaced around December 2019 in Wuhan, China (Sohrabi et al. 2020). Chinese authorities initially reported that many early cases had been traced back to the Wuhan wet market. This was reminiscent of the 2003 SARS1 outbreak in which a bat virus first jumped to *civets*, some of which were sold in *wet markets*, and from there the virus again leaped the species barrier to infect people (LeDuc and Barry 2004). Many scientists and Chinese government officials proposed that a similar event could have happened again, perhaps with the intermediate host this time being *pangolins* (Zhang et al. 2020) or a direct transmission from bats to humans (Zhou et al. 2020). The Wuhan wet market was signaled as perhaps the breeding ground of the outbreak (China Daily 2020). News media at the time echoed this first plausible explanation (see first row of Figure 2) and largely ignored the possibility of a lab-leak, as evidenced in the second, third, and fourth rows of Figure 2. Over time however, conclusive supportive evidence for the natural emergence hypothesis has not yet materialized, as no signs of prior intermediate host infection with COVID-19 have been found despite an intensive search (Wade 2021). Initial cases of COVID-19 not linked to the Wuhan wet market were also eventually reported (Chan et al. 2020). This perhaps explains the decreasing prevalence of the intermediate host hypothesis in news media content since its peak around February-April of 2020; see Figure 2a–d.

The possibility of a of lab-leak was largely absent in news media discourse during most of 2020 and has only gained prominence in mid-2021, as shown in the second row of Figure 2. A compelling reason to not rule out the lab-leak hypothesis was due to Wuhan hosting a virology laboratory that conducted research work on coronaviruses, the Wuhan Institute of Virology (WIV). Media interest about the lab has however, only peaked recently; as reflected in Figure 2i.

The WIV is also China's only maximum biosafety level-4 (BSL-4) laboratory, meaning it is authorized and equipped to work on the most dangerous viral pathogens; see Figure 2j. Critically, since at least 2015, this research lab had been working on gain-of-function experiments to make coronavirus strains more infectious of cells lining the human respiratory tract (Daszak 2014; Wade 2021), allegedly under inadequate safety conditions (*Washington Post* 2020). The rationale for such research being that the insights gained from it could be useful to prevent natural spillovers. Media mentions of the WIV engaging in this type of research have only picked up in May–June of 2021, see Figure 2k.

The hypothesis of a lab-leak was dismissed early in 2020 by some prominent members of the scientific community as a conspiracy theory and their opinions were published in prestigious scientific journals such as *The Lancet* (Calisher et al. 2020) and *Nature Medicine* (Andersen et al. 2020). This perhaps could explain why mainstream news media mostly echoed the natural emergence hypothesis and largely ignored the lab-leak alternative hypothesis during most of 2020 as the world suffered the thrust of the pandemic.

At least one signatory of *The Lancet* letter (Calisher et al. 2020) was a member and president of the EcoHealth Alliance, an organization that had funded coronavirus gain-of-function research at the Wuhan Institute of Virology with U.S. government grants from the National Institute of Allergy and Infectious Diseases (NIAIDS) (Daszak 2014; Wade 2021). The NIAIDS coronavirus gain-of-function research grant to the Wuhan Institute of Virology through EcoHealth has only recently attracted substantial media attention; see subplot *l* in Figure 2.

The most similar public genome to SARS-CoV-2 is a bat coronavirus known as *RaTG13*, with a genome similarity to SARS-CoV-2 of 96% (Zhou et al. 2020). Media attention to this virus that was retrieved from a cave in the Yunnan province (1800 km away from Wuhan), sequenced and published by staff from the Wuhan Institute of Virology (Zhou et al. 2020), has also only recently become prominent; see Figure 2m.

Several relevant molecular features of SARS-CoV-2 were also largely underreported by mainstream news media during 2020. In the middle of the SARS2 spike protein, a motif called the furin cleavage site is critical for the subunits of the spike protein (S1 and S2) to be cut apart by a protein cutting tool on the surface of human cells known as furin (Johnson et al. 2020). Such cleavage allows the virus to fuse with the target cells' membrane,

inject its genetic material into the cell and cause the cell to generate new copies of the virus. The human furin protein will cut any protein chain that carries the motif amino acid sequence proline-arginine-arginine-alanine (PRRA). SARS2 is the only SARS-related beta-coronavirus with a furin cleavage site, making it particularly optimized to target human cells (Wade 2021; Peacock et al. 2021). Yet, news media outlets have largely overlooked this molecular feature of the virus until recently; see Figure 2n.

News Media Aggregate Monthly Prevalence of Words Related to Covid-Origins Hypotheses
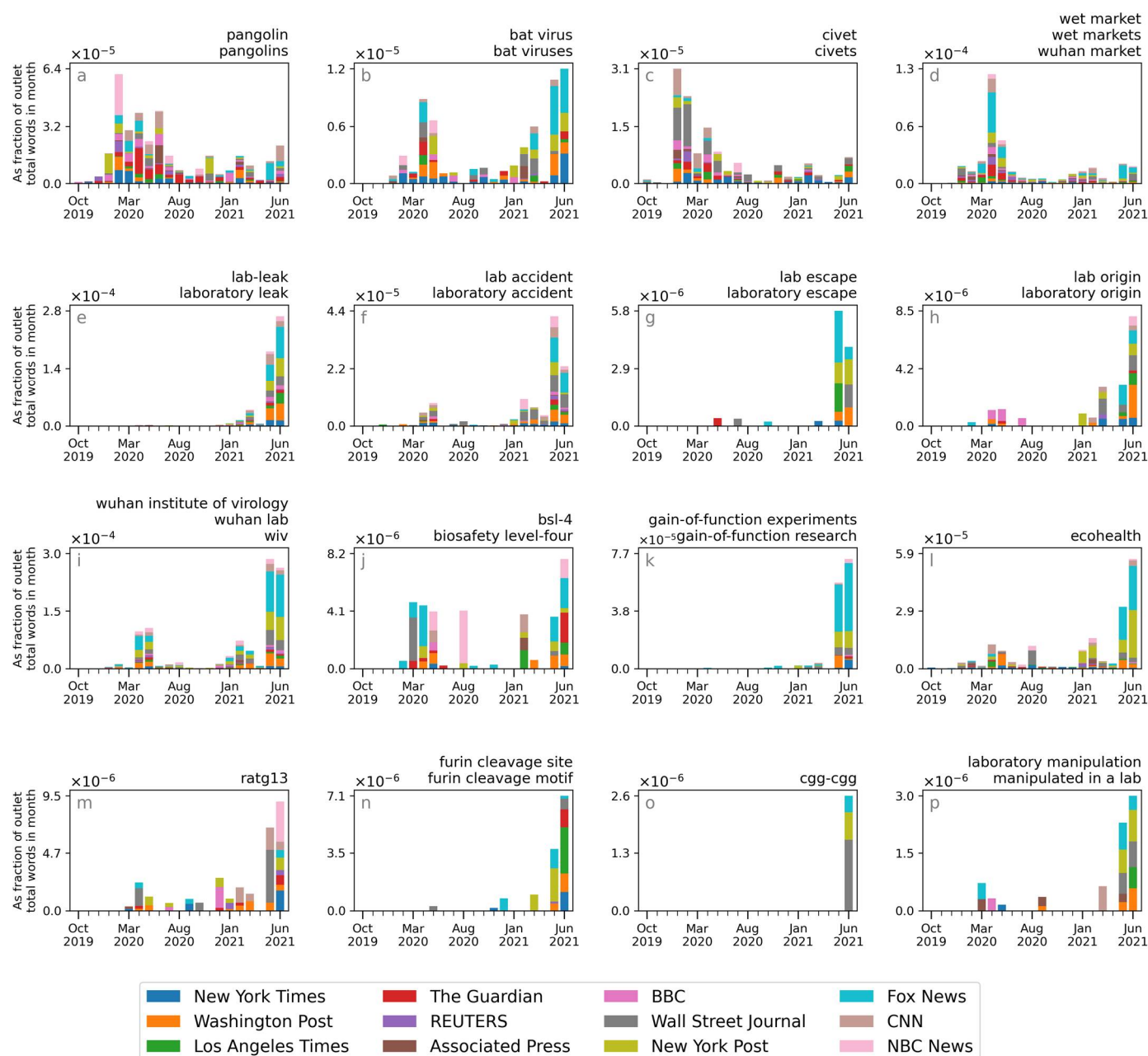### October 1, 2019 - June 30, 2021

Figure 2. Monthly frequency of terms related to the two competing hypotheses about COVID-19 origins: natural emergence (first row) and lab-leak (second, third and fourth rows).

At the S1/S2 junction, the 12-nucleotide sequence codifying the PRRA motif that renders the protein chain susceptible to be cleaved by furin and allow viral particles to fuse with human cells membrane is T-CCT-CGG-CGG-GC. This sequence contains the

unusual feature that the double arginine codons pattern, *CGG-CGG*, has never been found in any other beta coronavirus (Wade 2021). This molecular characteristic also appears to have been absent from news media discourse until recently; see Figure 2o. Perhaps as a result of the above discussed unusual molecular features of SARS-CoV-2, some news media outlets have only recently started to mention the possibility that the virus might have been *manipulated in a lab*; see Figure 2p.

3.2. High Frequency Analysis of the COVID-19 Lab-Leak Hypothesis Prevalence in News Media

Figure 2 only allows us to observe that the prevalence of the lab-leak hypothesis in news media content markedly increased in May and June of 2021. To visualize higher-resolution dynamics around this period, we replicate the previous analysis using weekly frequency counts for a set of key target words denoting the lab-leak hypothesis theme. Figure 3 shows that the prevalence in news media of the lab-leak hypothesis theme increased during the month of May to spike in the last week of that month, and then decreased gradually as the month of June progressed. There also appears to be milder peaks of this topic prevalence in mid-February and in the week at the end of March/beginning of April.

Figure 3 also illustrates that not all news media outlets have manifested a spike in the prevalence of the lab-leak hypothesis theme in their textual content. Instead, the increased prevalence has been driven mainly by just some outlets such as Fox News, The New York Post, The Wall Street Journal, and The Washington Post.

To achieve even higher granularity temporal dynamics of the lab-leak hypothesis thematic prevalence in news media, we next analyze daily frequency counts of target words in news media content from 1 January 2021 to 30 June 2021; see Figure 4. We leverage the usage peaks identified in Figure 3 to guide a search for potentially relevant events around those dates that could have plausibly influenced media coverage of the lab-leak hypothesis.

We have identified and highlighted in Figure 4 six such potentially relevant events. The first three are the beginning of the World Health Organization's (WHO) field visit to China to investigate the origins of the pandemic, their visit to the Wuhan Institute of Virology, and the end of their field visit to China. The next event corresponds to the publication of the WHO report on its Wuhan field visit investigation that recommended a call for further studies and reiterated that all hypotheses about COVID-19 origins remain open (World Health Organization 2021).

The next relevant event concerns Nicholas Wade, a former science reporter at the New York Times, and his publication of "The origin of COVID: Did people or nature open Pandora's box at Wuhan?" on 5 May 2021 (Wade 2021). In his article, Wade enumerated what he considered substantial evidence pointing in the direction of the lab-leak hypothesis, although he acknowledged that no definite proof existed yet for either the natural emergence or the lab-leak hypotheses.

The final highly likely influential event in press coverage of the lab-leak hypothesis concerns the U.S. president, Joe Biden, ordering to its intelligence community on 26 May 2021 to further investigate the origins of the COVID-19 virus and provide a report back to him within 90 days (REUTERS 2021).

News Media Aggregate Weekly Prevalence of Words Related to the Lab-Leak Hypothesis
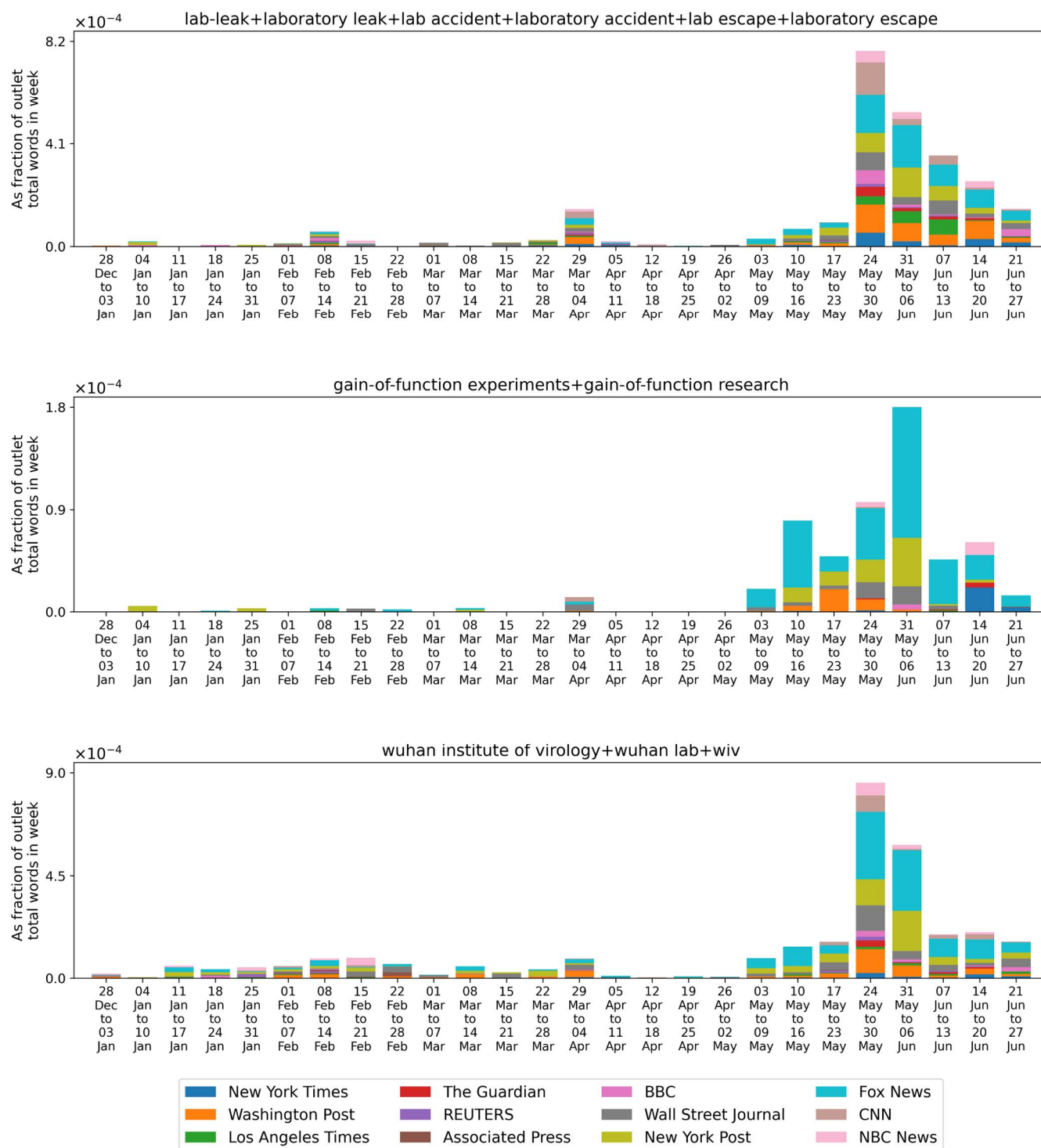December 28, 2020 - June 27, 2021

Figure 3. Weekly frequency of terms related to the lab-leak hypothesis about COVID-19 origins.
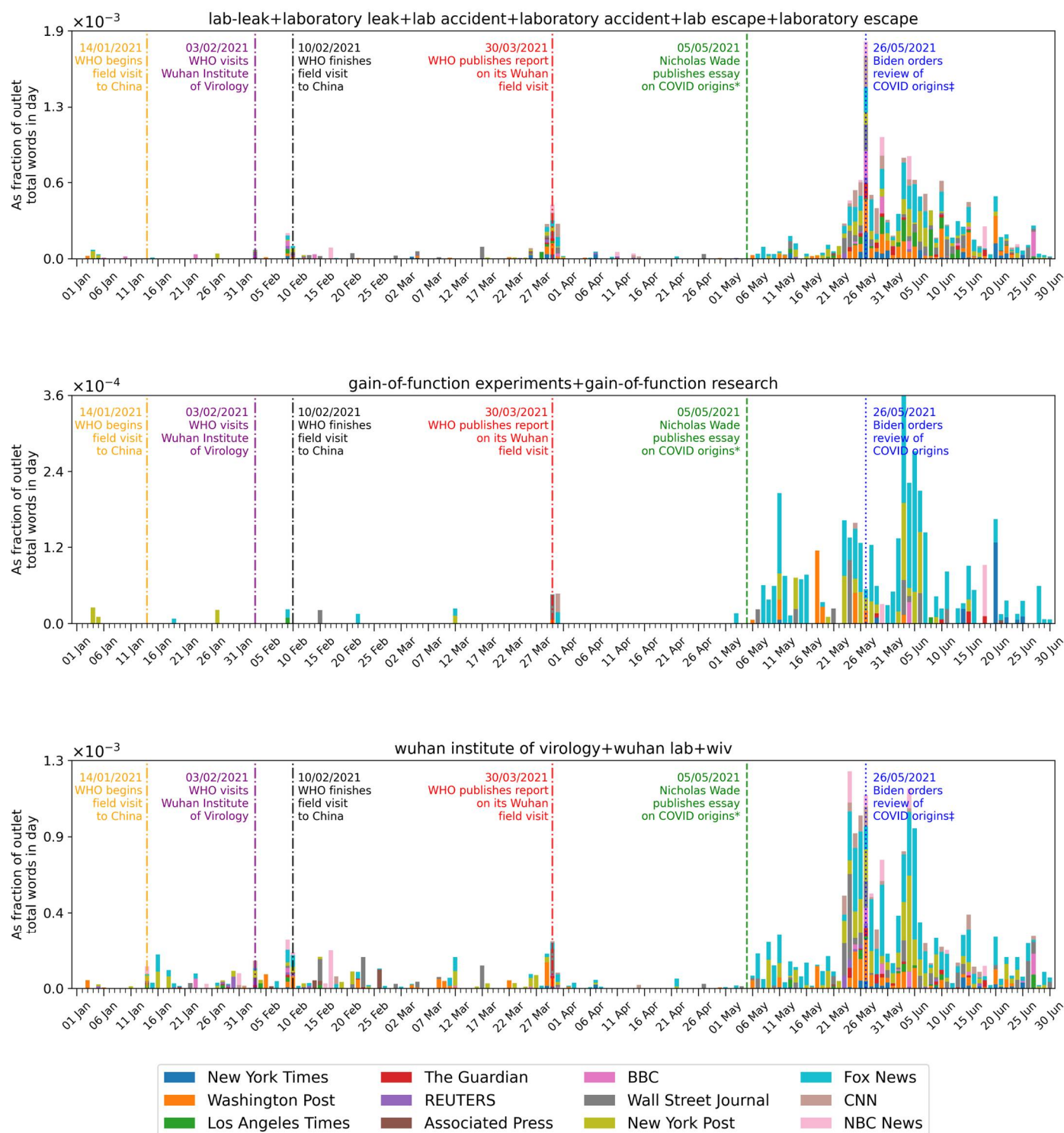
Figure 4. Daily frequency of terms related to the lab-leak hypothesis about COVID-19 origins. Statistically significant Chow tests and paired *t*-tests of overall prevalence ($p < 0.05$, Bonferroni corrected for multiple comparisons) prior to and post each vertical line event are indicated with the ‡ and * symbols, respectively.

A structural break analysis using the Chow test ([Chow 1960](#)) (Bonferroni adjusted for multiple comparisons) to determine whether regression coefficients prior to each event highlighted in Figure [4](#) were different from regression coefficients after each event (window size = 14) were statistically significant ($p < 0.05$) for the Biden event on 26 May 2021 for two sets of words (‡ markers in Figure [4](#)). Paired *t*-tests (Bonferroni adjusted for multiple comparisons) of overall prevalence prior to and after (window size = 14) each highlighted event in Figure [4](#) reached statistical significance ($p < 0.05$) for Nicholas Wade's article of 5 May 2021 for the three sets of words analyzed (* markers in Figure [4](#)). The largely absent prevalence of the lab-leak hypothesis theme in the days prior to Nicholas Wade's publication and the subsequent gradual pickup in media interest provides suggestive, but ultimately circumstantial, evidence about whether this particular event could have triggered increased media coverage of the lab-leak hypothesis.

3.3. Latent Associations about COVID-19 Origins in News Media Content

While frequency analysis of a corpus of text can be informative about the thematic prevalence of certain topics, the technique is also limited in that it does not analyze the context in which words are being used. To overcome this limitation, we next performed an analysis of news media articles using word2vec embedding models ([Mikolov et al. 2013](#)) to measure the frequency with which sets of words are associated. We built embedding models for each month between October 2019 and June 2021 using news media articles published in the corresponding month. This allows for chronological measurements of the strength with which sets of words are associated (i.e., appear in the vicinity of each other or in similar contexts) in news media articles.

Figure [5](#) shows the results of our analysis. The first row of Figure [5](#) contains subplots using a dashed orange line and it is only used to illustrate that the technique produces sensible results, including detecting the temporal occurrences of events such as Donald Trump's infection and subsequent positive testing for COVID-19, or Joe Biden winning the Democratic Party nomination for the U.S. presidency around March/April 2020 and his subsequent electoral victory in the U.S. presidential election of November 2020.

The second row of Figure [5](#) illustrates the decreasing prevalence of the intermediate host hypothesis in news media as shown by the declining association of *coronavirus* with potential intermediate hosts such as *pangolins*, *civets*, and *bats*, as well as peak association of the virus with *wet markets* between January and February of 2020.

The third row of Figure [5](#) shows how news media have recently started to more strongly associate terms such as *covid* or *coronavirus* with a *lab-leak* or the *Wuhan Institute of Virology*. Subplot k in the figure also shows that during 2020, the media mostly did not report on the *gain of function* research experiments being conducted since 2015 at the WIV ([Daszak 2014](#); [Wade 2021](#)). Similarly, associations about the *dangerous* nature of *gain of function* research are stronger in mid-2021 than at any time in 2020. The commonality for all these associations is that their strength of association has peaked around May and June of 2021.

Subplots m and n in Figure [5](#) shows that associations between the research grants from NIAID/NIH for *gain of function* research at the WIV have become more prominently linked in the last few months. Subplot o also illustrates how in recent journalistic discourse, the *lab-leak* hypothesis is often associated with terms denoting *racism*. The embedding method used does not allow discerning whether such associations occur because news media content suggests that it is racist to propose the lab-leak hypothesis or whether some writers are arguing that the lab-leak hypothesis was not properly scrutinized previously because of concerns about accusations of racism or in an attempt to not stir up racist sentiment. The pattern could also be the result of a combination of all the previous possibilities. Finally, associations between the *WIV* and a potential *laboratory accident* have become more prominent in 2021, although the relationship was also briefly common in April and May of 2020; see subplot p in Figure [5](#).

Strength of association between terms *v* and *w* as measured in embedding space* using $cos(\theta) = \frac{v \cdot w}{\|v\| \times \|w\|}$
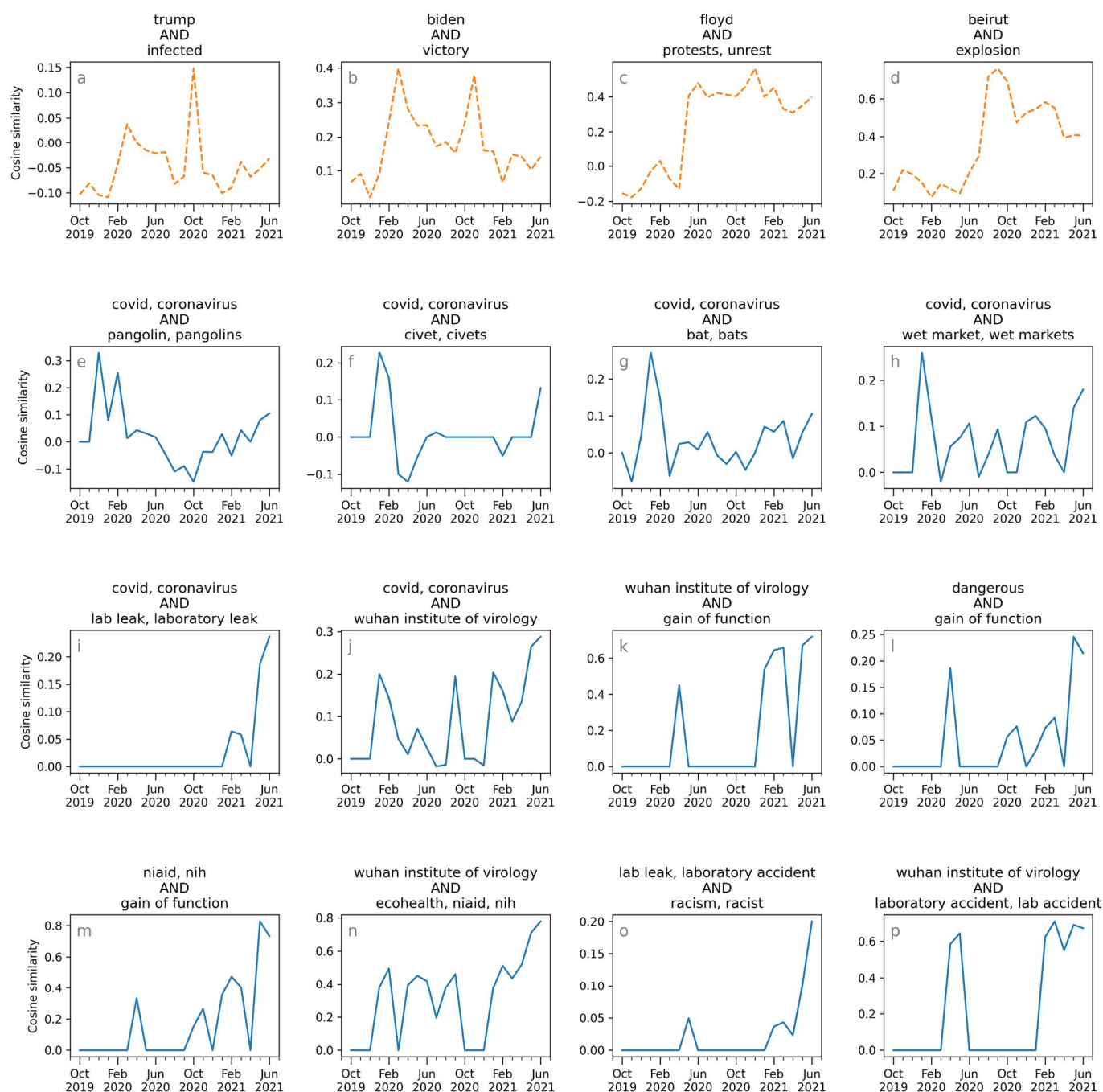
* Embedding models built with monthly article data from:
The New York Times, The Washington Post, Los Angeles Times, The Guardian, REUTERS, Associated Press, BBC, Wall Street Journal, New York Post, Fox News, CNN and NBC News

Figure 5. Chronological plots of monthly association strength between sets of terms in embedding models derived from news media content.

4. Discussion

As of July 2021, the origin of the SARS-CoV-2 virus that caused the COVID-19 pandemic remains a mystery. The results presented here suggest that for most of 2020, popular

news media outlets mostly ignored or downplayed the possibility of a lab-leak as a reason for the virus outbreak. Perhaps the publication in prestigious scientific journals, such as *The Lancet* and *Nature Medicine,* of opinion pieces dismissing the lab-leak hypothesis (Andersen et al. 2020; Calisher et al. 2020) played a role in media attitudes towards this hypothesis.

Alternatively, the fact that early on in the pandemic, U.S. president at the time, Donald Trump, advocated for the lab-leak hypothesis without providing explicit evidence (*BBC News* 2020) could have contributed to prominent news outlets avoiding such hypothesis due in part to the notorious mutual animosity between news media organizations and Trump.

If mutual hostility between news media outlets and former U.S. President Donald Trump partly prompted prestigious outlets during 2020 to downplay a plausible hypothesis about COVID-19 origins, the ability of news media institutions to reliably investigate and report on politically-loaded events in an unbiased manner could be raised into question.

In May 2021, however, something caused the prevalence of the lab-leak hypothesis in news media discourse to substantially spike in prominence in some, but not all, of the studied media outlets. Although our analysis cannot provide conclusive evidence about what caused the shift, it provides hints about potential sources for the structural break in the prevalence of the lab-leak hypothesis in news media discourse.

If Nicholas Wade's essay did indeed trigger the sudden increase in attention of at least some outlets to the lab leak hypothesis, it is extraordinarily striking that news media scrutiny about the causal roots of a pandemic that has killed, as of July 2021, more than 4 million people worldwide (COVID-19 Data Repository CSSE—JHU 2021) could be dependent on the investigative reporting of a single individual. It is also noteworthy that most of the interest in the lab-leak hypothesis seems to have emerged in right-leaning news outlets (Fox News, the Wall Street Journal, and the New York Post). Although, the Washington Post, a prominent left-leaning newspaper (AllSides Media Bias Ratings 2019), has also manifested in its content an increasing prevalence of the lab-leak hypothesis.

Interpreting the results of this work through the media agenda-setting theory suggests the potential of news media to shape public perceptions about important current events. A valid criticism of this interpretation is that with the growing influence of the Internet and social media, people can find information through alternative sources other than traditional news media outlets, making it harder for news media to uniquely set agendas. Nonetheless, the majority of the population still trusts news media reporting on the COVID-19 pandemic (De Coninck et al. 2020). Such trust, however, appears to be eroding (Fletcher et al. 2020). Fair and honest reporting on current events is essential to maintain trust between the public and news media organizations. If additional supporting evidence for the lab-leak hypothesis eventually surfaces while supporting evidence for the natural emergence hypothesis fails to materialize, the downplaying of the lab-leak hypothesis in mainstream news outlets during the first 16 months of the pandemic could contribute to further public erosion of trust in news media.

References

AllSides Media Bias Ratings. 2019. AllSides. 2019. Available online: https://www.allsides.com/blog/updated-allsides-media-bias-chart-version-11 (accessed on 31 July 2021).

Andersen, Kristian G., Andrew Rambaut, W. Ian Lipkin, Edward C. Holmes, and Robert F. Garry. 2020. The Proximal Origin of SARS-CoV-2. *Nature Medicine* 26: 450–52. [CrossRef]

BBC News. 2020. Coronavirus: Trump Stands by China Lab Origin Theory for Virus. May 1. sec. US & Canada. Available online: https://www.bbc.com/news/world-us-canada-52496098 (accessed on 31 July 2021).

Bloom, Jesse D., Yujia Alina Chan, Ralph S. Baric, Pamela J. Bjorkman, Sarah Cobey, Benjamin E. Deverman, David N. Fisman, Ravindra Gupta, Akiko Iwasaki, Marc Lipsitch, and et al. 2021. Investigate the Origins of COVID-19. *Science* 372: 694–94. [CrossRef] [PubMed]

Bridgman, Aengus, Eric Merkley, Peter John Loewen, Taylor Owen, Derek Ruths, Lisa Teichmann, and Oleg Zhilin. 2020. The Causes and Consequences of COVID-19 Misperceptions: Understanding the Role of News and Social Media. *Harvard Kennedy School Misinformation Review* 1. [CrossRef]

Calisher, Charles, Dennis Carroll, Rita Colwell, Ronald B. Corley, Peter Daszak, Christian Drosten, Luis Enjuanes, Jeremy Farrar, Hume Field, Josie Golding, and et al. 2020. Statement in Support of the Scientists, Public Health Professionals, and Medical Professionals of China Combatting COVID-19. *The Lancet* 395: e42–e43. [CrossRef]

Chan, Jasper Fuk-Woo, Shuofeng Yuan, Kin-Hang Kok, Kelvin Kai-Wang To, Hin Chu, Jin Yang, Fanfan Xing, Jieling Liu, Cyril Chik-Yan Yip, Rosana Wing-Shan Poon, and et al. 2020. A Familial Cluster of Pneumonia Associated with the 2019 Novel Coronavirus Indicating Person-to-Person Transmission: A Study of a Family Cluster. *The Lancet (London, England)* 395: 514–23. [CrossRef]

China Daily. 2020. Wuhan Wet Market Closes amid Pneumonia Outbreak. January 1. Available online: https://www.chinadaily.com.cn/a/202001/01/WS5e0c6a49a310cf3e35581e30.html (accessed on 31 July 2021).

Chow, Gregory C. 1960. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica* 28: 591–605. [CrossRef]

COVID-19 Data Repository CSSE—JHU. 2021. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Available online: https://github.com/CSSEGISandData/COVID-19 (accessed on 31 July 2021).

Daszak, Peter. 2014. Understanding the Risk of Bat Coronavirus Emergence. June 1. Available online: https://reporter.nih.gov/search/xQW6UJmWfUuOV01ntGvLwQ/project-details/9491676 (accessed on 31 July 2021).

De Coninck, D., L. d'Haenens, and K. Matthijs. 2020. Forgotten Key Players in Public Health: News Media as Agents of Information and Persuasion during the COVID-19 Pandemic. *Public Health* 183: 65–66. [CrossRef] [PubMed]

Fletcher, Richard, Antonis Kalogeropoulos, and Rasmus Kleis Nielsen. 2020. *Trust in UK Government and News Media COVID-19 Information Down, Concerns Over Misinformation from Government and Politicians Up. SSRN Scholarly Paper ID 3633002*. Rochester: Social Science Research Network, Available online: https://papers.ssrn.com/abstract=3633002 (accessed on 31 July 2021).

Johnson, Bryan A., Xuping Xie, Birte Kalveram, Kumari G. Lokugamage, Antonio Muruato, Jing Zou, Xianwen Zhang, Terry Juelich, Jennifer K. Smith, Lihong Zhang, and et al. 2020. Furin Cleavage Site Is Key to SARS-CoV-2 Pathogenesis. *BioRxiv*. [CrossRef]

Jorge, April. 2021. Hydroxychloroquine in the Prevention of COVID-19 Mortality. *The Lancet Rheumatology* 3: e2–e3. [CrossRef]

LeDuc, James W., and M. Anita Barry. 2004. SARS, the First Pandemic of the 21st Century1. *Emerging Infectious Diseases* 10: e26. [CrossRef]

McCombs, Maxwell. 2018. Agenda-Setting. In *The Blackwell Encyclopedia of Sociology*. Atlanta: American Cancer Society, pp. 1–2. [CrossRef]

McCombs, Maxwell E., and Donald L. Shaw. 1972. The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly* 36: 176–87. [CrossRef]

Medina, Leslie M., Janette R Rodriguez, and Philip Joseph D Sarmiento. 2021. Shaping Public Opinion through the Lens of Agenda Setting in Rolling out COVID-19 Vaccination Program. *Journal of Public Health* 43: e389–e390. [CrossRef]

Mehmood, Muhammad Amir, Hafiz Muhammad Shafiq, and Abdul Waheed. 2017. Understanding Regional Context of World Wide Web Using Common Crawl Corpus. Paper present at 2017 IEEE 13th Malaysia International Conference on Communications (MICC), Johor Bahru, Malaysia, November 28–30; pp. 164–69. [CrossRef]

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *Advances in Neural Information Processing Systems 26*. Edited by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger. Red Hook, NY, USA: Curran Associates, Inc., pp. 3111–19. Available online: http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf (accessed on 31 July 2021).

Mrogers, Everett, and James Wdearing. 1988. Agenda-Setting Research: Where Has It Been, Where Is It Going? *Annals of the International Communication Association* 11: 555–94. [CrossRef]

Nielsen. 2020. Report 'Most in the UK Say News Media Have Helped Them Respond to COVID-19, but a Third Say News Coverage Has Made the Crisis Worse.'. *Salto*. August 25. Available online: https://participationpool.eu/resource/report-most-in-the-uk-say-news-media-have-helped-them-respond-to-covid-19-but-a-third-say-news-coverage-has-made-the-crisis-worse/ (accessed on 31 July 2021).

Notess, Greg R. 2002. The Wayback Machine: The Web's Archive. *Online* 26. Available online: https://www.elibrary.ru/item.asp?id=4370136 (accessed on 31 July 2021).

Peacock, Thomas P., Daniel H. Goldhill, Jie Zhou, Laury Baillon, Rebecca Frise, Olivia C. Swann, Ruthiran Kugathasan, Rebecca Penn, Jonathan C. Brown, Raul Y. Sanchez-David, and et al. 2021. The Furin Cleavage Site in the SARS-CoV-2 Spike Protein Is Required for Transmission in Ferrets. *Nature Microbiology* 6: 899–909. [CrossRef]

Řehůřek, Radim, and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. Paper present at LREC 2010 Workshop on New Challenges for NLP Frameworks, Valletta, Malta, May 20; pp. 45–50.

REUTERS. 2021. Biden Orders Review of COVID Origins as Lab Leak Theory Debated. *Reuters*. May 26. Available online: https://www.reuters.com/business/healthcare-pharmaceuticals/biden-says-us-intelligence-community-divided-covid-origin-2021-05-26/ (accessed on 31 July 2021).

Rozado, David. 2020a. Prejudice and Victimization Themes in New York Times Discourse: A Chronological Analysis. *Academic Questions* 33: 89–100. [CrossRef]

Rozado, David. 2020b. Wide Range Screening of Algorithmic Bias in Word Embedding Models Using Large Sentiment Lexicons Reveals Underreported Bias Types. *PLoS ONE* 15: e0231189. [CrossRef]

Rozado, David, Musa Al-Gharbi, and Jamin Halberstadt. 2021. Prevalence of Prejudice-Denoting Words in News Media Discourse: A Chronological Analysis. *Social Science Computer Review*. [CrossRef]

Rozado, David, and Musa Al-Gharbi. 2021. Using Word Embeddings to Probe Sentiment Associations of Politically Loaded Terms in News and Opinion Articles from News Media Outlets. *Journal of Computational Social Science*. [CrossRef]

Sohrabi, Catrin, Zaid Alsafi, Niamh O'Neill, Mehdi Khan, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, and Riaz Agha. 2020. World Health Organization Declares Global Emergency: A Review of the 2019 Novel Coronavirus (COVID-19). *International Journal of Surgery (London, England)* 76: 71–76. [CrossRef] [PubMed]

Wade, Nicholas. 2021. The Origin of COVID: Did People or Nature Open Pandora's Box at Wuhan? *Bulletin of the Atomic Scientists*. May 5. Available online: https://thebulletin.org/2021/05/the-origin-of-covid-did-people-or-nature-open-pandoras-box-at-wuhan/ (accessed on 31 July 2021).

Washington Post. 2020. OpinionState Department Cables Warned of Safety Issues at Wuhan Lab Studying Bat Coronaviruses. April 14. Available online: https://www.washingtonpost.com/opinions/2020/04/14/state-department-cables-warned-safety-issues-wuhan-lab-studying-bat-coronaviruses/ (accessed on 31 July 2021).

World Health Organization. 2021. WHO Calls for Further Studies, Data on Origin of SARS-CoV-2 Virus, Reiterates That All Hypotheses Remain Open. March 30. Available online: https://www.who.int/news/item/30-03-2021-who-calls-for-further-studies-data-on-origin-of-sars-cov-2-virus-reiterates-that-all-hypotheses-remain-open (accessed on 31 July 2021).

Zhang, Tao, Qunfu Wu, and Zhigang Zhang. 2020. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current Biology: CB* 30: 1578. [CrossRef]

Zhou, Peng, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, and et al. 2020. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* 579: 270–73. [CrossRef] [PubMed]