The following paper was researched and written as part of LSE GROUPS 2023.

LSE GROUPS takes place during the final fortnight of the LSE summer term. Undergraduate students are placed in small groups; these are cross-year, interdisciplinary, and group members do not know one another in advance. Each group must then devise its own research question, and carry out every stage of a small-scale research project in less than two weeks.

LSE GROUPS is part of the LSE commitment to students learning through enquiry, and developing the skills for knowledge creation.

The overall theme of LSE GROUPS 2023 was *Connections.*

This paper was submitted on the final Thursday afternoon of the project.

Students then presented their work at a conference, on the closing Friday.

More information on LSE GROUPS, and other papers.

Papers are presented as submitted by the students, without corrections.

London School of Economics and Political Science

Eden.GROUPS@lse.ac.uk

Web of Lies: An experimental exploration of the effect of in-person social connections on misinformation

Hagan Alford, Weixiao Guo, Max Rosenfeld, Daniel Seymour and Khushi Vajpeyi

Group 8

June 2023

Abstract

Background: With the rise of digital connectivity, society risks undervaluing in-person social connections. Research has suggested that the consumption of misinformation in the digital ecosystem is a potential threat to both individuals and institutions regarding disasters, health, and politics (Muhammed and Matthew, 2020).

Objectives: Empirical research has identified drivers of misinformation such as national economic outcomes, anxiety, and education levels. The literature is underdeveloped in two areas: limited examination of social connections and a bias towards researching the social media segment of the misinformation ecosystem. Our study explores how in-person social connections are related to the consumption of misinformation.

Methodology: Utilising secondary website traffic data for misinformation websites, aggregated from the UK and Sweden, over a 29-month period of the COVID-19 pandemic to conduct experimental research. We use UK lockdowns as a treatment, and Sweden's non-lockdown policy as a control, to conduct a difference-in-difference analysis examining the effect of lockdown implementation on web traffic.

Results: Our model finds the lockdown in the UK has a statistically significant relation to an increase in web traffic to misinformation websites. Due to the observational nature of our data, we explored the effects of several confounders according to previous literature, such as national economic outcomes and internet penetration.

Keywords: misinformation, social connection, COVID-19, experimental methodology, isolation, website traffic

1. Introduction

This paper seeks to contribute to existing literature by utilizing the COVID-19 pandemic as a natural experiment to examine the research question: How does an absence of in-person social connections affect the consumption of misinformation? While misinformation was first recorded in the time of Ramesses II in 1213 B.C. (Cline, 2021), it was in 2014 when Craig Silverman at Columbia University coined the now viral term "fake news", which reinforced the great impact that false information can have in a highly interconnected world. Silverman had no idea that it would transform into an umbrella term used by media watchdogs, interest groups and researchers in equal measure. "Fake news" can imply a range of meanings, and contemporary studies classify false information as 'disinformation' or 'misinformation' (House of Commons, 2019). Simply put, disinformation is the deliberate spread of falsehoods while misinformation refers to incorrect information that is spread un-intentionally. This paper will focus on the spread and consumption of misinformation only and will utilize a broad definition of the term as defined by the House of Commons: "the inadvertent sharing of false information" (ibid.). The main channel through which misinformation spreads and gains credibility is through social connections, whether in-person or online. Existing research on misinformation has focused mainly on identifying trends in online misinformation through social media, largely neglecting the role of other channels such as news websites and, more fundamentally, the relationship between in-person social interaction and consumption of misinformation. The Community Life Survey measuring the wellbeing of people aged 16+ in England showed that 48% of the correspondents experienced loneliness often, sometimes, or occasionally from 2020 to 2021 (Wellbeing and Loneliness - Community Life Survey 2020/21, 2021). The onset of the COVID-19 pandemic and subsequent implementation of 'lockdown' policies provides a suitable setting to measure how the consumption of misinformation varies with restrictions on in-person social contact. This paper is structured as follows: first, we will situate our research within existing literature, second, we will describe and justify our methodological approach, third, we will consider the ethics of our study and subsequently present the data analysis. Finally, we will discuss the key research findings and present our concluding remarks.

2. Literature Review

2.1 Developing the research question

Muhammed and Mathew's (2020) paper reviewed and synthesized 28 findings on social media information, showing characteristics and consequences of misinformation spread in the fields of politics, health and disasters. This motivated us to further navigate the causes of misinformation spread and hence contribute to preventing severe losses. Inspired by the research of Marin et al. (2021) on how social isolation and psychological impairment increase one's vulnerability to misinformation during the Covid-19 pandemic, we developed an interest in associating in-person social connections with misinformation spread. Lewandowsky et. al. (2017) identified the rise in prominence of alternative truths following large societal trends in declining social capital, increases economic inequality, and greater polarization. Lee, Agrawal and Rao (2015) found in the aftermath of the 2013 Boston Bombing, tweets from accounts with larger follower counts had increased legitimacy as well as a faster dispersion rate after the immediate attack. Identifying the conditions and rate of misinformation spreads is directly related to our research in reviewing website traffic.

2.2 Refining our methodology

Existing studies have developed reliable methods of compiling and analyzing social media data: Chen et al. (2020) developed a dataset of over 123 million multilingual tweets focusing on Twitter responses to COVID-19 related events while Lenti et. al. (2023) studied global misinformation flows within no-vax communities using a similar dataset. Our initial approach, thus, was to use Twitter IDs to study retweets. However, changes to the accessibility of the Twitter API led us to shift from this approach and turn to other components of digital information system as suggested by Ofcom (2021). Ofcom's (2021) paper also allowed us to identify other channels, such as news websites and video platforms, of online misinformation spread.

Moreover, existing studies analyzing the relationship between social connection and misinformation only draw upon limited sample sizes. Marin et al. (2022) included under 200 participants from the US and Italy in their study, which largely reduced the generalizability of their findings, driving us to search a wide range of websites.

A comparative study of France and Italy by Fletcher et al. (2018) described people's consumption of misinformation on social media and news websites, indicated by their interactions on Facebook and time spent on major news sites. This guided us to better filter websites and focus on data that could speak about our research question. After contextualising the research question in the UK, we referred to Ofcom's (2021) paper to identify reliable analytics providers including NewsGuard and SimilarWeb, given that the study focused on the UK and analysed 14 million visits from 177 false information websites from 2018 to 2020.

3. Methodology

3.1 Collecting website data

We analyzed secondary data to address our research question. In particular, our methodological approach consisted of country selection, misinformation website identification and collection of website traffic data. The website traffic data consisted of monthly observations from the online data analytics platform Semrush over a 23-month period from January 2019 to November 2020.

Preliminarily, as it will be subsequently covered in more detail, we employed a Difference-in-Difference experimental design to examine misinformation in the UK, requiring the selection of a control variable. Sweden was chosen as the UK had strict Covid-19 restrictions, moving from no interaction, 6-person, and 30-person restrictions (Baker et al., 2023) while Sweden had no restrictions (Covid-19 In Sweden, 2023) for our observation period. Furthermore, Sweden and the UK are both geographically proximal, high-income countries.

There were 34 unique websites with 32 websites in the British dataset and 17 in the Swedish dataset. This means both countries share 15 English language misinformation websites, a reflection of the fact that misinformation during the Covid-19 pandemic disproportionately arose from the United States and other English-speaking countries (Lenti et al., 2023). The websites were selected from a collection of factchecking, university library guides, and government websites to produce a list of commonly identifiable misinformation sites. To verify this claim, we used the NewsGuard website nutrition label tool which analyzes news sources and evaluates them based on several criteria to produce a score. The exact details of NewsGuard's methodology are covered in Appendix A. The inclusion criteria for websites were if they were primarily English-speaking (100% for UK and 88% for Sweden). The two uniquely Swedish websites were Newsvoice and Nyadagbladet and they were only included in the Swedish dataset to strengthen our data since English speaking websites would limit the representativeness of our sample.

Utilizing Ofcom (2021) as a methodological guide, we reviewed a range of website traffic analytics service providers. We selected Semrush as it gave the most comprehensive historical website data, and we could isolate organic traffic which bypassed the problem of idiosyncratic advertising campaigns from the websites. An initial screening of the websites was conducted to eliminate those deemed inactive and those that had activity in only one country. Using Python, the data was reformatted into median monthly data due to the high variance amongst website traffic.

3.2. Operationalization of in-person social connections

We operationalized in-person social interaction as consisting of social gatherings and working in a group setting. Due to the stickiness of the work-from-home trend, we focused on the more measurable size of social gatherings allowed by the British and Swedish governments. For Britain, we sorted each month in the 23-month period from January 2019 to November 2020 into either a lockdown or a non-lockdown month. We defined a lockdown as any month that had a restriction on 6 people or under. Hence, we classified the three months after the First National Lockdown in the UK as a non-lockdown month which we accounted for in the model. In comparison, Sweden pursued, throughout the start of the COVID-19 pandemic, a non-lockdown policy only implementing a restriction on gatherings up to eight people in November 2020, so all months in the 23-month period are classified as non-lockdown.

3.3 Model Design

We selected difference-in-difference as the most appropriate statistical model to run a natural experiment using the UK as a treatment group and Sweden as a control group. This method relies on the parallel trends assumption which we assume to hold given the apparent parallel trend we see when the data is plotted (see figure 3).

To construct our initial model as follows:

$$web\ traffic = \beta_0 + \beta_1(UK) + \beta_2(post - lockdown) + \beta_3(UK \cdot \text{post} - \text{lockdown})$$

We now consider potential confounders. The existing literature identified the following key confounders: economic inequality, national economic outcomes and education (Lewandowsky et al., 2017; Milan and Treré, 2020; Zrnec et al, 2022). We then identified internet penetration as a further potential confounder. We were unable to include controls for a number of these confounders due to a lack of frequent data so have compared national annual data for 2020.

Figure 1

Characteristic (2020)	UK	Sweden
Population	67,081,000	10,353,442
% of Population Using Internet	95%	95%
Enrolment in Tertiary Education	69%	85%
Gini Coefficient	32.6	28.9

Figure 1 shows that the UK and Sweden are relatively comparable in terms of internet penetration and income inequality, measured by the Gini coefficient. However, they vary in Tertiary education levels which, according to existing literature, could account for high misinformation consumption in the UK. We were able to control for national economic outcomes intrinsically by including quarterly GDP growth in our model, giving us the final difference-in-difference model:

$$web\ traffic = \beta_0 + \beta_1(UK) + \beta_2(post - lockdown) + \beta_3(UK \cdot \mathrm{post - lockdown}) + \beta_4(GDP\ growth)$$

Figure 2

Variable	Definition
UK	A Dummy for the treatment area. It's coded 1 for UK data
Post-lockdown	A Dummy for post-treatment periods coded 1 for post-lockdown
UK * post-lockdown	Interaction Term
GDP growth	Quarterly GDP growth rate

4. Ethical Considerations

Due to our reliance on secondary data, in terms of ethics we were primarily concerned about the appropriate collection and storage of all information. We have agreed to store all relevant information on LSE authenticated servers and have followed all procedures for collecting data. We have also followed both Semrush and NewsGuard’s terms of service when using their free trials. We considered potential hazards in this experiment as being financial costs due to service subscriptions and computer viruses. To reduce these hazards, we have deselected automated payments for either service.

5. Findings

We begin by defining our statistical hypotheses:

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0$$

Here β_3 is the interaction coefficient.

We ran our difference-in-difference model with three variations. Firstly, we set the treatment time as the beginning of the first lockdown in the UK, 2020-03, and dummy coded every subsequent period as post-lockdown. The black line in figure 3 indicates the treatment event.

Figure 3

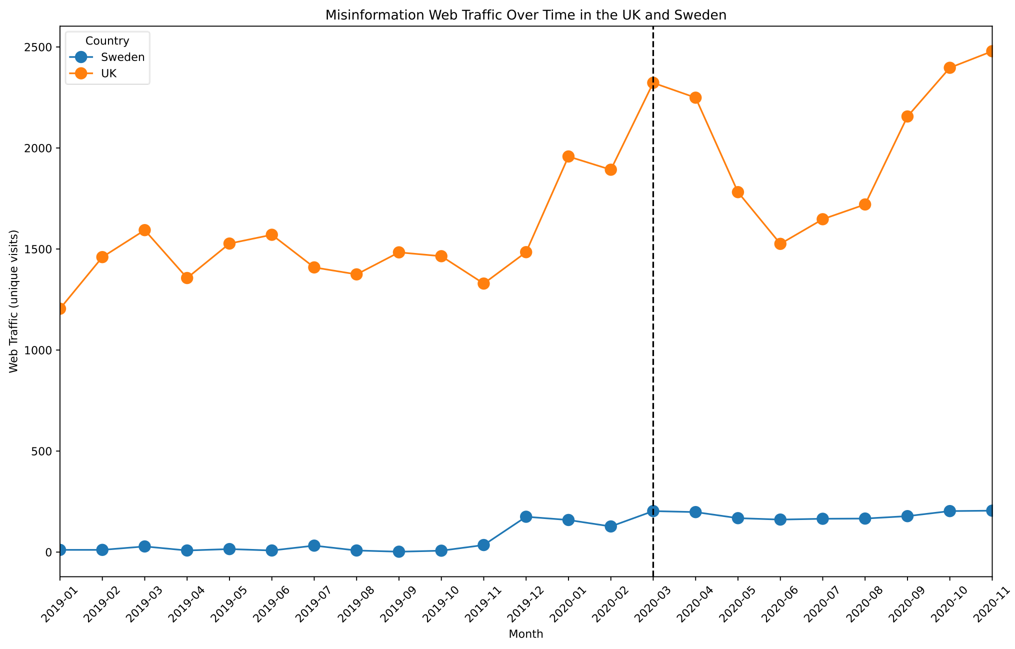

Figure 4

	Coefficient	Std Err	**t**	**p > |t|**	95% conf. interval	
Constant	44.7143	53.398	0.837	0.407	-63.125	152.554
UK	1462.9757	75.526	19.371	0.000	1310.448	1615.503
Post-lockdown	138.2990	85.374	1.620	0.113	-34.117	310.715
UK · post-lockdown	385.0087	120.758	3.188	0.003	141.113	628.884
GDP growth	0.0400	4.163	0.010	0.992	-8.368	8.448

The positive UK coefficient is statistically significant ($p < 0.001$) and indicates larger consumption of misinformation in the UK as compared to Sweden on average. According to the model, an absence of in-person social connection was associated with on average around 1462 extra visitors per month over the lockdown period for this sample. The post-lockdown coefficient is statistically insignificant and so we are unable to ascertain if consumption increased or decreased on average during the post-lockdown period. The interaction term has a statistically significant ($p < 0.01$) positive coefficient which allows us to reject the null hypothesis, indicating that the lockdown in the UK resulted in a greater increase in misinformation consumption than would have occurred had the lockdown not been implemented. The GDP growth coefficient is highly insignificant eliminating this confounder and reinforcing our finding.

The second variation of our model was to dummy code the three months from 2020-07 to 2020-09 inclusive to not be considered in the post-lockdown period. This is done to account for the temporary easing of restrictions on social gatherings to allow gatherings up to thirty people. The red lines on figure 5 indicate the exclusion period.

Figure 5

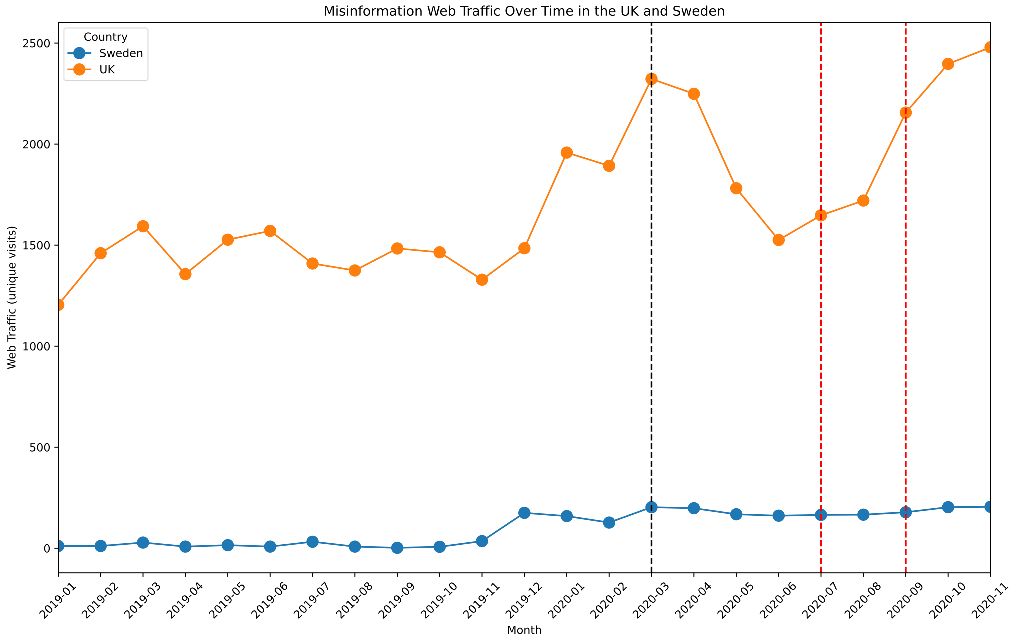

Figure 6

	Coefficient	Std Err	**t**	**p > |t|**	95% conf. interval	
Constant	41.0575	40.875	1.004	0.321	-41.490	123.606
UK	1471.6387	57.607	25.546	0.000	1355.300	1587.977
Post-lockdown	231.8514	82.521	2.810	0.008	65.197	398.506
UK · post-lockdown	589.3080	116.642	5.052	0.000	353.745	824.871
GDP growth	20.8106	4.335	4.800	0.000	12.055	29.566

In this variation we observe a statistically significant ($p < 0.001$) increase misinformation consumption in the post-lockdown period. We also observe a larger, and more significant ($p < 0.001$), interaction coefficient, allowing us to reject the null hypothesis, strengthening our initial observation of an increase in consumption as a result of the lockdown. However, in this variation the GDP growth coefficient is positive and statistically significant ($p < 0.001$) and therefore explains some of the increase in misinformation consumption.

We were curious about the apparent drop in UK traffic in the first three months of the lockdown. To explore this, our third variation reduces the time in the post period to only include dates up until the first easing of restrictions.

Figure 7

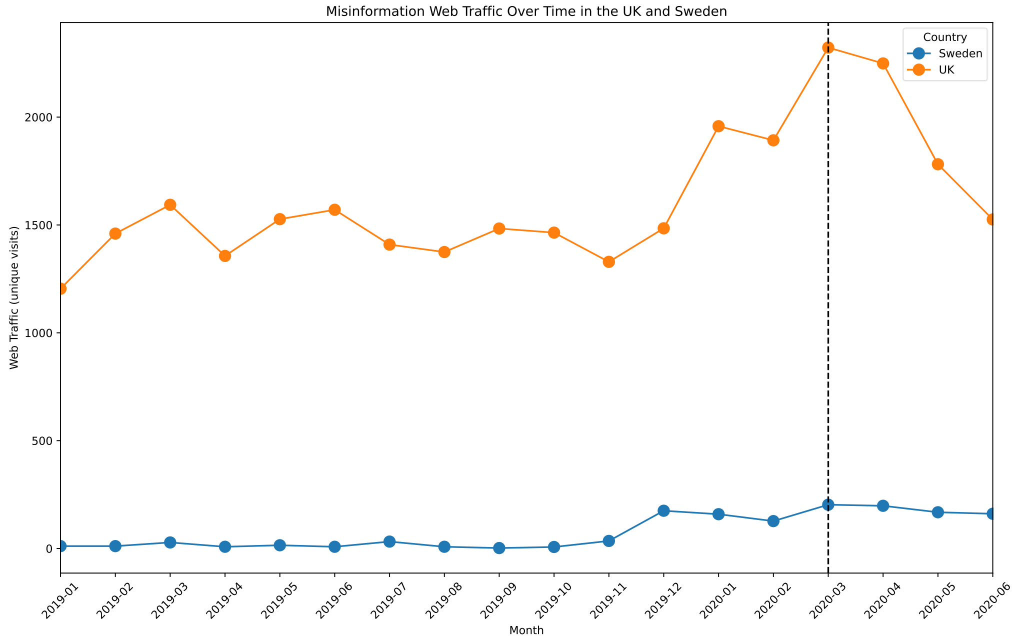

Figure 8

	Coefficient	Std Err	**t**	**p** > |**t**|	95% conf. interval	
Constant	55.2667	35.362	1.563	0.128	-16.855	127.388
UK	1382.8167	53.043	26.070	0.000	1274.634	1490.999
Post-lockdown	-2357.9333	364.075	-6.477	0.000	-3100.469	-1615.397
UK · post-lockdown	-3423.9833	527.180	-6.495	0.000	-4499.174	-2348.793
GDP growth	-309.1917	44.203	-7.008	0.000	-399.944	-219.640

From this variation we find all coefficients are statistically significant with $p < 0.001$. We find a similar result for the UK variable as we did in the two previous variations. However, we now find the opposite outcomes for the other variables. We find that consumption falls in the post-lockdown period and that the lockdown made this effect greater in the UK than it would have been without lockdown. We also find that GDP growth is negatively related to misinformation consumption which agrees with existing research (Milan and Treré, 2020).

6. Discussion:

Overall, our findings indicate that a lack of in-person social connections results in an increase in consumption of misinformation. This supports Marin et al.'s (2022) findings that social isolation is associated with misinformation. Through our approach we have expanded on this literature by showing the effect on a national level and by considering misinformation in general as opposed to COVID-19 specific.

The results of the first and second variations of our model work to reinforce our overall finding. By initially coding all post-lockdown months as the treatment period, our model shows that social isolation increases the consumption of misinformation. The increase in significance and size of the effect of lockdown when we remove the period of lockdown easing in the second variation, further reinforces that the effect we are observing is the result of the absence of in-person social connections. Additionally, the finding in the second variation that GDP growth per quarter is positively related to misinformation consumption contradicts existing research which suggests that negative national economic outcomes are related to an increase in misinformation (Milan and Treré, (2020). However, the relationship between negative economic outcomes and the COVID-19 pandemic and lockdowns calls the validity of this result into question.

Finally, a potential explanation for our finding from the third variation of our model is the availability and frequency of official information. Lee et al.'s (2015) study of misinformation, following the 2013 Boston bombing, found that misinformation after the beginning of a crisis spreads as a result of people searching for information quickly and that if official sources produce information fast enough then they can prevent misinformation from spreading. At the beginning of the COVID-19 pandemic, the UK government made daily press briefings. These briefings ended in June 2020, the exact point on the graph where we start to see an increase in misinformation consumption. When restrictions of social gatherings are then reimplemented we see a significant spike in consumption.

7. Conclusion

In conclusion, our findings indicate a positive relationship between the absence of in-person social connections and the consumption of misinformation. We do acknowledge; however, the methodological limitations of our study and we interpret our findings with these methodological challenges in mind.

Firstly, our sample of websites is small relative to the potential population for two main reasons: availability of data and the nature of misinformation websites often being small and temporary. Additionally, our model was unable to intrinsically control for several confounders owing to the availability of data. However, our findings are consistent with existing literature (see Lee et al., 2015), and we control for extrinsic confounders in the comparison between the UK and Sweden, acknowledging any differences, both of which work to reinforce the validity of our findings.

For future research, we recommend repeating our approach with a stronger sample and extending the study to contain other areas of the misinformation ecosystem, such as social media, to test the external validity and causal inference of our findings.

Acknowledgements

We would like to thank Dr Ioanna Gouseti, Department of Sociology, LSE for her constant support and feedback. We would also like to thank Dr Ellis Saxey and the LSE Eden Centre for Education Enhancement for giving us this opportunity. We are grateful for the assistance of Mr Haoran Shi, Dr Sara Geneletti, and Dr Michael Gmeiner to this project.

Bibliography

Baker, C., Brown, J., Kirk-Wade, E. and Barber, S. (2023) 'Coronavirus: A history of English lockdown laws'. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-9068/ (Accessed: 14 June 2023).

Chen, E., Lerman, K. and Ferrara, E. (2020) 'Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set', *JMIR Public Health and Surveillance*, 6(2), p. e19273. Available at: https://doi.org/10.2196/19273.

Cline, E.H. (2021) *1177 B.C.: The year civilization collapsed*. Princeton University Press. Available at: https://press.princeton.edu/books/paperback/9780691208015/1177-bc.

Covid-19 in Sweden (2023) *sweden.se*. Available at: https://sweden.se/life/society/sweden-and-corona-in-brief (Accessed: 14 June 2023).

Fletcher, R., Cornia, A., Graves, L. and Nielsen, R.K., 2018. Measuring the reach of" fake news" and online disinformation in Europe. *Australasian Policing*, *10*(2).

House of Commons (2019) 'Disinformation and "fake news"', *House of Commons*, Eighth Report of Session 2017-19, p. 111.

Lee, J., Agrawal, M. and Rao, H.R. (2015) 'Message diffusion through social network service: The case of rumor and non-rumor related tweets during Boston bombing 2013', *Information Systems Frontiers*, 17(5), pp. 997–1005. Available at: https://doi.org/10.1007/s10796-015-9568-z.

Lenti, J., Mejova, Y., Kalimeri, K., Panisson, A., Paolotti, D., Tizzani, M. and Starnini, M., 2023. Global Misinformation Spillovers in the Vaccination Debate Before and During the COVID-19 Pandemic: Multilingual Twitter Study. *JMIR infodemiology*, *3*, p.e44714.

Lewandowsky, S., Ecker, U.K.H. and Cook, J. (2017) 'Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era', *Journal of Applied Research in Memory and Cognition*, 6(4), pp. 353–369. Available at: https://doi.org/10.1016/j.jarmac.2017.07.008.

Marin, A., Vives-Rodriguez, A., DeCaro, R., Schiloski, K.A., Hajos, G.P., Di Crosta, A., Ceccato, I., La Malva, P., Lahdo, N.C., Donnelly, K. and Dong, J., 2021. Associations between misinformation around COVID-19 pandemic, severity of social isolation, and cognitive impairment. *Alzheimer's & Dementia*, *17*, p.e054468. (Accessed: 15 June 2023).

Milan, S. and Treré, E. (2020) 'The Rise of the Data Poor: The COVID-19 Pandemic Seen From the Margins', *Social Media + Society*, 6(3), p. 2056305120948233. Available at: https://doi.org/10.1177/2056305120948233.

Ofcom (2021) 'Understanding online false information in the UK', *Ofcom*, (Economist discussion paper series, Issue number 2), p. 56.

Pennycook, G. *et al.* (2021) 'Shifting attention to accuracy can reduce misinformation online', *Nature*, 592(7855), pp. 590–595. Available at: https://doi.org/10.1038/s41586-021-03344-2.

Wellbeing and Loneliness - Community Life Survey 2020/21 (2021) *GOV.UK*. Available at: https://www.gov.uk/government/statistics/community-life-survey-202021-wellbeing-and-loneliness/wellbeing-and-loneliness-community-life-survey-202021 (Accessed: 15 June 2023).

Zrnec, A., Poženel, M. and Lavbič, D. (2022) 'Users' ability to perceive misinformation: An information quality assessment approach', *Information Processing & Management*, 59(1), p. 102739. Available at: https://doi.org/10.1016/j.ipm.2021.102739.

Appendix A: NewsGuard Rating Criteria

NewsGuard Rating Criteria	100 points
Does not repeatedly publish false content	22 Points
Gathers and presents information responsibly	18 Points
Regularly corrects or clarifies errors	12.5 Points
Handles the difference between news and opinion responsibly	12.5 Points
Avoids deceptive headlines	10 Points
Website discloses ownership and financing	7.5 Points
Clearly labels advertising	5 Points
Reveals who’s in charge, including possible conflicts of interest	10 Points
Website discloses ownership and financing	7.5 Points
The site provides the names of content creators	5 Points

Appendix B: Semrush Methodology

"Traffic Analytics reports are based on petabytes of clickstream data that comes from multiple proprietary and 3rd party data sources, Semrush’s proprietary AI and machine learning algorithms and Big Data technologies. The data is accumulated and approximated from the user behavior of over 200 million real internet users, and over a hundred different apps and browser extensions are used to collect it." (Semrush, 2023)