Behavior Research Methods
https://doi.org/10.3758/s13428-022-01840-5

Understanding the landscape of web-based medical misinformation about vaccination

Christopher R. Wolfe[1] · Andrew A. Eylem[1] · Mitchell Dandignac[1] · Savannah R. Lowe[1] · Margo L. Weber[1] · Laura Scudiere[2] · Valerie F. Reyna[3]

Accepted: 17 March 2022

Abstract
Given the high rates of vaccine hesitancy, web-based medical misinformation about vaccination is a serious issue. We sought to understand the nature of Google searches leading to medical misinformation about vaccination, and guided by fuzzy-trace theory, the characteristics of misinformation pages related to comprehension, inference-making, and medical decision-making. We collected data from web pages presenting vaccination information. We assessed whether web pages presented medical misinformation, had an overarching gist, used narrative, and employed emotional appeals. We used Search Engine Optimization tools to determine the number of backlinks from other web pages, monthly Google traffic, and Google Keywords. We used Coh-Metrix to measure readability and Gist Inference Scores (GIS). For medical misinformation web pages, Google traffic and backlinks were heavily skewed with means of 138.8 visitors/month and 805 backlinks per page. Medical misinformation pages were significantly more likely than other vaccine pages to have backlinks from other pages, and significantly less likely to receive at least one visitor from Google searches per month. The top Google searches leading to medical misinformation were "the truth about vaccinations," "dangers of vaccination," and "pro con vaccines." Most frequently, pages challenged vaccine safety, with 32.7% having an overarching gist, 7.7% presenting narratives, and 17.3% making emotional appeals. Emotional appeals were significantly more common with medical misinformation than other high-traffic vaccination pages. Misinformation pages had a mean readability grade level of 11.5, and a mean GIS of – 0.234. Low GIS scores are a likely barrier to understanding gist, and are the "Achilles' heel" of misinformation pages.

Keywords Inoculation information · Google search terms · Fuzzy trace theory · Disinformation · Health information · Psycholinguistic properties of websites

Introduction

The politicized response to the COVID-19 pandemic presents a stark reminder of growing vaccine hesitancy in the United States and around the world (Stahl et al., 2016). Although there are a variety of reasons why individuals may be reluctant to have themselves and their children immunized against COVID-19 and a host of other viruses (Jaiswal & Halkitis, 2019), medical misinformation about vaccination has been recognized as a large and growing problem. Researchers have made headway understanding the way social media has been used to spread medical misinformation and anti-vaccination arguments (Broniatowski & Reyna, 2020; Hussain et al., 2019) as well as the way anti-vaccination messages spread via YouTube videos (Tang et al., 2021). Another source of medical misinformation about COVID-19 and vaccination is Twitter (Rosenberg et al., 2020), a phenomenon crossing many national borders and languages (Madraki et al., 2021). Caldarelli et al. (2021) found that the impact of disreputable Twitter posts about COVID-19 in Italy exceeded 20% in conservative communities, with 96% of all disreputable URLs shared by political Twitter groups coming from right- and center right-wing communities (Caldarelli et al., 2021). There are also left-wing biases, albeit of different kinds. The approach taken in the present research was to investigate Google searches that yield medical misinformation and the cognitive and psycholinguistic characteristics of web pages presenting medical

✉ Christopher R. Wolfe
wolfecr@miamioh.edu

[1] Department of Psychology, Miami University, Oxford, OH 45056, USA

[2] Marathon County Health Department, Wausau, WI, USA

[3] Cornell University, Ithaca, NY 14850, USA

Published online: 05 April 2022

misinformation about vaccination. To understand the landscape of web-based medical misinformation about vaccination we used Search Engine Optimization tools and discourse technologies, guided by fuzzy-trace theory (Reyna, 2008, 2012) to assess the characteristics of web pages most likely to affect medical decision-making.

Search Engine Optimization (SEO) tools are designed to help businesses develop digital marketing strategies and understand which Google search terms lead customers to their competitors' web sites. We used an SEO website called Ahrefs (Ahrefs, 2021) to assess aspects of search behavior related to pages presenting medical misinformation about vaccination. Ahrefs uses proprietary algorithms and information from Google Keyword Planner, Clickstream (Clickstream, 2021), and other sources. Ahrefs claims to have indexed 5.02 billion web pages, 9.5 billion keywords in the US, and created an index of 160 million keywords in the US (Ahrefs, 2021). Keywords are the words and phrases that searchers enter into Google, collapsing over typos, spelling mistakes, and minor variations in wording. We cataloged the number of organic (unpaid) keywords, and up to 50 ordered keywords people Googled to arrive at each web page presenting medical misinformation about vaccination. Organic traffic is the monthly estimated average number of visits to a site from organic Google searches (as distinct from AdWords). We further assessed backlinks, the number of links from other websites leading users to each vaccination web page. Search volume represents how many times different keywords are used in Google searches. For every keyword, we also assessed monthly volume, and monthly traffic to each web page presenting medical misinformation about vaccination.

In addition to SEO tools to provide insights into search behaviors, emerging discourse technologies make it possible to automatically evaluate the psycholinguistic properties of text—including those found on web pages. Coh-Metrix (McNamara et al., 2014) is a discourse technology that automatically computes 106 linguistic variables. Coh-Metrix is "data rich" (Wolfe & Dandignac, 2021) drawing upon research with meaningful texts and human research participants (e.g., Coltheart, 1981). The current research examines the psycholinguistic characteristics of pages presenting medical misinformation about vaccination to understand the web-based information environment. This is useful because many strategies to correct vaccine misinformation have backfired (Pluviano et al., 2017).

Recent advances in discourse technologies have enabled researchers to analyze web sites about vaccines and viruses automatically. To illustrate, Wolfe et al. (under review) found that web-page searches reflecting concerns about vaccination scored significantly higher on Gist Inference Score, and were significantly more likely to have verified medical misinformation than other vaccine web pages. Wolfe et al. (under review) found that websites in search clusters expressing vaccination concerns ($N = 531$) had significantly more pages with medical misinformation, 7.2% than other clusters ($N = 667$,) with 2.1%. Chin, Su, and Chin et al. (2020) found that misinformation about the HPV vaccine scored higher on narrativity and word familiarity compared to accurate information, suggesting the misinformation is easier to comprehend. Web-based medical information about influenza has been found to be more difficult to comprehend based on the most widely used metric, Flesch–Kincaid grade level (FKGL), and similar readability measures (Basch et al., 2019). Additionally, research on HPV vaccines and clinical cancer trials has found that the readability of these websites has been rated at a higher level of difficulty than is appropriate for many readers in the U.S. (Hillyer et al., 2020; MacLean et al., 2018). However, despite its wide use and utility, there is reason to believe that FKGL misses important features of text difficulty.

Fuzzy-trace theory (Reyna, 2008, 2012) posits that people represent information along a continuum from superficial verbatim representations that capture precise wording but not necessarily the underlying meaning, to gist representations capturing the bottom-line meaning of events without necessarily retaining specific words. Gist and verbatim representations are independent of one another and in reasoning, problem solving, and decision-making people exhibit a strong preference to utilize the most gist-like mental representation permissible for a given task. Moreover, experts are more likely to form useful gist representations while novices are more likely to rely on superficial verbatim representations. Thus, forming useful gist representations is essential in making decisions such as whether or not to be vaccinated. Gist Inference Scores (GIS; Wolfe, Dandignac, & Reyna, 2019a; Wolfe et al., 2019; Dandignac & Wolfe, 2020; Wolfe et al., 2021) use Coh-Metrix to automatically assess written materials for the extent to which they help readers to form inferences about their bottom-line meaning. Building on fuzzy-trace theory (Reyna, 2008, 2012) and psycholinguistic research, cohesion and abstractly conceptual language aid the encoding of meaningful gist representations, and the lack of cohesion, and greater word concreteness facilitate verbatim representations. The GIS formula combines six Coh-Metrix psycholinguistic variables assessing the likelihood that readers will generate meaningful inferences from texts. Referential cohesion (words and ideas overlap), deep cohesion (logical connectives), and latent semantic analysis (LSA) verb overlap (actions interrelate across the text assessed using latent semantic analysis) are weighted positively, and word concreteness, imagability, and hypernymy (specificity) for nouns and verbs are weighted negatively. We have found that reading higher GIS texts predicts better outcomes related to inferential comprehension (necessary to understand messages), gist reasoning (necessary to use

information productively in medical decision-making), and declarative knowledge (explicit knowledge of facts; Wolfe, Dandignac, & Reyna, 2019a; Wolfe et al., 2019; Dandignac & Wolfe, 2020; Wolfe et al., 2021). Participants randomly assigned to authentic high GIS texts about breast cancer scored significantly better than low-GIS participants on knowledge and reasoning (Wolfe, Dandignac, & Reyna, 2019a). Another study used GIS to predict the outcome of completing a fill-in-the-blanks comprehension task (Wolfe et al., 2021). Participants were randomly assigned to read texts from the National Cancer Institute website. Participants performed better on the high-GIS texts, while also showing a greater word variation to fill-in-the-blanks with semantically similar words (Wolfe et al., 2019). Dandignac and Wolfe (2020) found comparable outcomes in a national survey of women, testing the effect of GIS on recall of an article about soft tissue sarcoma. They found that a greater percentage of what was recalled was decision-relevant information for high-GIS texts compared to low-GIS texts (Dandignac & Wolfe, 2020). Therefore, our research applies Gist Inference Scores (GIS), a measure of the likelihood that readers will extract meaningful ideas, or form gist representations, from text (Wolfe, Wolfe, Dandignac, & Reyna, 2019a; Wolfe et al., 2019; Dandignac & Wolfe, 2020; Wolfe et al., 2021).

Research suggests that narratives may be more compelling and easier to understand than other approaches to presenting medical information such as statistics (Fagerlin et al., 2005), and narratives have been found to be a factor in vaccine hesitancy (Duchsherer et al., 2020). Thus, we assessed whether each page presenting medical misinformation presents a narrative. Emotional appeals have also been found to influence medical decisions in domains including bone marrow donation (Studts et al., 2010). Hence, we also assessed whether web pages presenting verified medical misinformation used emotional appeals.

As more people get health information online, health-related topics increasingly become the target of misinformation (Krishna & Thompson, 2021), with some going so far as to call medical disinformation "the next phase of bioware-fare" (Bernard et al., 2021). Wolfe et al. (2021) notes that the anti-vaccination movement has gone mainstream. One good explanation for the widespread effectiveness of on-line misinformation about vaccines stems from fuzzy-trace theory (Reyna et al., 2021). In essence, people receiving anti-vaccination messages through social media and web searches may reason that COVID-19 may be a serious threat or the threat may be exaggerated, and vaccination may be safe or it may be dangerous; thus the best course of action is to do nothing. Moreover, people may reason that they are okay now, and if they get vaccinated, they could either be okay or not okay, prompting inaction. Importantly, this suggests that medical misinformation may have negative consequences even when people don't believe it.

Although the mere presence of anti-vaccination information from like-minded sources may increase vaccine hesitancy, rational arguments in favor of vaccination may be ineffective due the "myside bias" against information supporting another side of an argument. It appears that in a politically "polarized environment, people exhibit a strong 'myside bias' (Wolfe, 2012; Wolfe et al., 2009) with other side arguments [about vaccination] apparently having little effect on gist representations" (Wolfe, 2021, p. 529). Weil and Wolfe (2022) found that participants judged arguments about COVID-19 measures largely based on the claim rather than supporting reasons, and that politically conservative participants reported higher levels of acceptable risk, lower risk estimates of activities such as eating in a restaurant, and endorsed more misinformation (however, Jimenez et al., 2020, found that perceptions of disease severity and recall of symptoms did not interact with attitudes towards vaccines, failing to support the confirmation bias).

Our research questions were: (a) what is the nature of Google searches that lead people to medical misinformation about vaccination, and (b) what are the cognitive and psycholinguistic characteristics of pages presenting medical misinformation compared to other high-traffic pages about vaccination that do not present medical misinformation, particularly as they relate to comprehension, inference-making, and medical decision-making.

Methods

We manually searched the web for information about vaccination on computers void of cookies in private browsing mode to ensure that previous searches would not influence subsequent results. Searches were conducted in 2020 from five U.S. States (in the East, Midwest, and West) on the Bing, DuckDuckGo, and Google search engines. Most were Google searches, and for these we used search terms starting with every letter of the alphabet. The rational for these procedures was to help ensure that we did not miss any high-traffic web sites that people used to access information about vaccination that differed markedly from terms that we might use, and to ensure that biases of location and search engine did not cause us to miss high-traffic sites. Google is by far the most popular and influential search engine. In 2021 Google accounted for over 86% of searches worldwide (Statistica, 2022a), and over 93% of searches in the United States (Statistica, 2022b), which is why we used it most extensively. We collected misinformation pages by sometimes using search terms different from those used by the medical community. We took a conservative approach to defining medical misinformation. For example, simply claiming that vaccines are unsafe was not coded as medical misinformation since different people

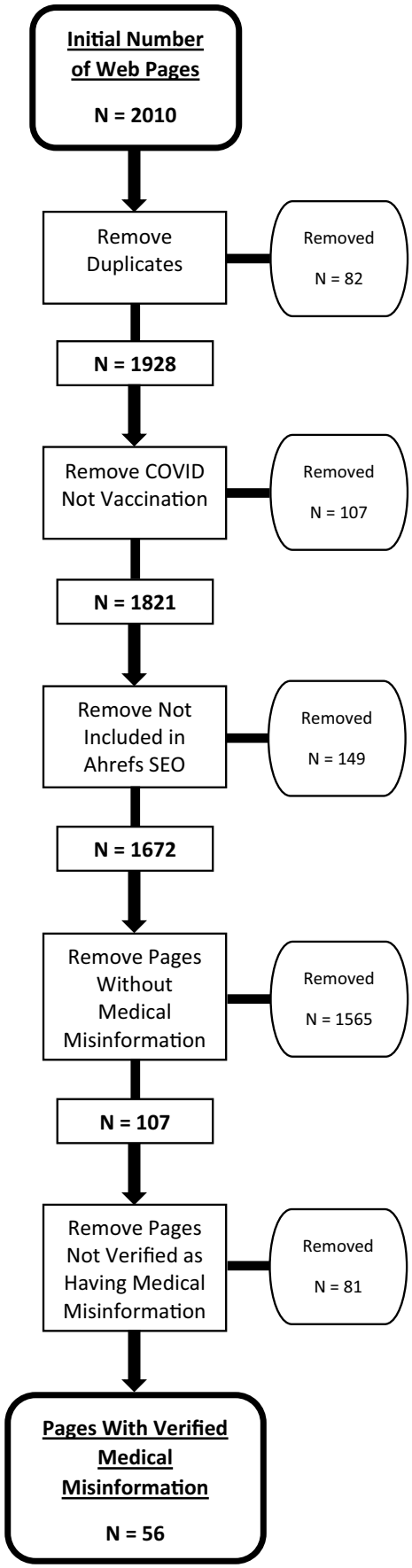

Fig. 1 Number of web pages with verified medical misinformation included in analyses

might have different standards for whether something is safe. Falsely claiming that pharmaceutical companies make most of their money from vaccines was not coded as medical misinformation since the misinformation is of a financial rather than medical nature. However, claiming that vaccines are unsafe because they are full of toxins or that vaccines cause autism were classified as medical misinformation. Coders were instructed that medical misinformation can take several forms. Some types of medical misinformation about vaccinations may include gross exaggerations, conspiracy theories, and outrageous claims (note that the authors may or may not believe what they write, some outrageous information online is clearly "click bait"). Judgments of possible medical misinformation were submitted to a public health expert for verification.

For each search, we cataloged vaccination results five Google results pages deep, with a stop rule per search set when no new URLs were found on the first results page. We reasoned that it is unlikely for a high-traffic web site to appear beyond the 5th page of any Google search, particularly because most users do not search beyond the first page of results. We manually cataloged URL, title, author, our search term, medical misinformation judgments, focus, mentioning COVID-19, domain, search engine, and depth in search results pages. As noted in Fig. 1, we cataloged a total of 2010 web pages, however, 82 were removed as duplicates (between coders), and 107 were removed because they addressed COVID-19 but not vaccination. We submitted 1821 URLs for SEO analysis, concentrating on data within the US. Of these, 149 provided no data about the specific web page, thus we were able to catalog data for 1672 web pages about vaccination in the U.S. Of these pages, 107 were identified as possibly presenting strictly medical misinformation (other kinds of misinformation such as exaggerating the profits pharmaceutical companies make from vaccines or equating vaccination with the biblical "mark of the beast" were not included) and evaluated by a public health expert experienced with vaccine health education. Of these, 56 were confirmed as presenting medical misinformation. We also used SEO data about 1565 vaccine web pages without medical misinformation for comparison, as well as psycholinguistic and cognitive data about 82 high-traffic vaccine web pages without medical misinformation identified by Wolfe et al. (under review). We also present published psycholinguistic norms for social studies texts for 11th grade through adult readers for comparison (McNamara et al., 2014). Our interest is in generalizable knowledge rather than specific web pages. We note that not all misinformation is intentional disinformation, for example, when the PBS Frontline page presents multiple perspectives, and

unrelated pages sometimes share the same title. Thus, inferences about specific pages should be made with caution.

For these 56 pages, we recorded monthly organic traffic from Google (number of visitors not directed by advertising—none were the result of advertising), number of backlinks, and number of organic keywords. In order of hits per page, we collected data on up to 50 keywords, recording keyword (e.g. "are vaccines safe"), position in search results, volume (average monthly searches using the keyword), and traffic (how many visitors the keyword brought to each specific web page). In addition to judgments about whether a site presented medical misinformation, coders made judgments of whether it provided readers with an overarching gist, whether it presented a narrative, and whether it made an emotional appeal. Web pages were coded as making emotional appeals when the authors appeared to be trying to influence readers by manipulating their feelings of anger, sorrow, fear etc. Pages were coded as narrative when at least part of the text constituted storytelling with characters, goals, and actions. Overarching gist was operationalized as a pithy statement about the bottom line meaning of the entire text. Overarching gist, emotional appeal, and narrative were operationalized holistically. Coders were also trained on web sites exemplifying each construct at the begging of the training period, with the list of exemplars expanding through each round of rating and establishing inter-rater reliability and consensus. Initially, two team leaders and the PI rated blocks of 20 then 30 then 50 web sites discussing whether each exemplified overarching gist, narrative, or emotional appeal. The team leaders then trained coders first by providing examples of texts exhibiting and not exhibiting each trait and then leading coders in practice and feedback on separate blocks of 20 web sites. For example, a web site called CDC Admits in Federal Court It Has No Evidence 'Vaccines Don't Cause Autism' (Martino, 2020) was used to exemplify both an overarching gist (the CDC claims that vaccines don't cause autism, but they are liars), and emotional appeal (this is a bombshell story and fact-checkers have been handing out false claims about our vaccine content, leading to the near-complete destruction of our business due to demonetization). The site Jenny McCarthy: "We're Not An Anti-Vaccine Movement… We're Pro-Safe Vaccine" (Frontline, 2015) presents a narrative about her son turning blue and being rushed to the hospital, and an emotional appeal including a statement that the doctors "blew it off" as a febrile seizure. A site called "The Cold Hard Truth About Vaccines" exemplifies an overarching gist, (there is a connection between vaccinations and miscarriages/still births), an emotional appeal, and presents a narrative. The Scientific American site Fact or Fiction?: Vaccines Are Dangerous is an example of a site that does not have an overarching gist, use narrative, or make an emotional appeal. All judgments were made by two coders with differences settled by consensus. After a training period, for the current study and research reported by Wolfe et al. (under review), two researchers made each judgment separately, again with differences being settled by consensus. Inter-rater reliability (IRR) calculated as the conditional probability that one judge made an affirmative judgment given that the other had done so (Cedillos-Whynott et al., 2016; Gwet, 2014) and absolute agreement. For overarching gist, conditional IRR = 0.76 (absolute agreement 0.93); for narrative, conditional IRR = 0.72 (absolute agreement 0.96); and for emotional appeal, conditional IRR = 0.81 (absolute agreement 0.98).

In preparation for Coh-Metrix analysis, for each page, we copied and pasted the text into a Word document, removing all images, ads, links and references. Four pages were unsuitable for Coh-Metrix analysis (e.g., presenting labeled links rather than expository text). Paragraphs were separated by two line breaks. The cleaned text was then submitted to a Coh-Metrix 3.0 analysis online (Coh-Metrix, 2021). For each text, we cataloged measures of readability and the ease of making inferences including text length in words, FKGL, Flesch reading ease, GIS, syntactic simplicity, syntactic similarity, and verb overlap LSA. To assess the accessibility and ease of processing web pages conveying medical misinformation about vaccination, we examined word-level and whole text-level psycholinguistic variables for each text. Word-level Coh-Metrix measures recorded include word concreteness, word frequency, age of word acquisition, familiarity of content words, concreteness of content words, imagability, and Colorado meaningful word norm. Coh-Metrix whole text variables cataloged include narrativity percentile, referential cohesion percentile, deep cohesion percentile, verb cohesion percentile, connectivity percentile, temporality percentile, and causal verb indices.

Results

Organic traffic from Google to pages presenting verified medical misinformation (i.e., visitors connecting via web search rather than ads) ranged from 2400/month to less than 1 per month according to Ahrefs analytics estimates based on data from Google and proprietary algorithms. Table 1 presents monthly traffic, backlinks from other web sites, number of organic keywords and search cluster based on clustering of web pages by the Google keywords used to reach them for pages with more than one visitor per day.

Table 1 demonstrates that the distribution of pages presenting medical misinformation is heavily skewed with just a few pages accounting for the majority of traffic. Because traffic was also skewed for vaccination web pages without medical misinformation, the differences in traffic between vaccination misinformation pages (mean = 135.8, standard deviation = 425) and non-misinformation vaccination pages

Table 1 Web traffic, backlinks, keywords, and cluster for medical misinformation web pages with summary and comparison

Title of Web Page	Monthly Organic Traffic	Backlinks	Organic Keywords	Cluster
Truth Will Prevail	2400	174	1666	Dangers of Vaccination
Should Any Vaccines Be Required for Children	1400	13,900	2737	Arguments Against Vaccination/ Anti-Vax
Vaccines - Pros & Cons	1300	4250	1109	Arguments Against Vaccination/ Anti-Vax
Jenny McCarthy: "We're Not An Anti-Vaccine Movement, We're Pro-Safe Vaccine"	933	427	536	Arguments Against Vaccination/ Anti-Vax
Dangers of Vaccinations	797	2	48	Dangers of Vaccination
The Pros & Cons of Immunization	137	0	66	Arguments Against Vaccination/ Anti-Vax
How Can Vaccines Cause Damage	98	24	66	Dangers of Vaccination
Vaccinations: Know the Risks and Failures	87	852	127	Vaccine Facts
8 Reasons to Delay Vaccines for Kids	81	55	70	Alternatives to Standard Vaccination
The Truth About Vaccines	50	0	47	The Truth About Vaccines
The World Needs COVID-19 Vaccines. It May Also be Overestimating Their Power	44	468	141	Arguments Against Vaccination/ Anti-Vax
The Plausible Connection Between Vaccines and SIDS	41	57	58	SIDS and Vaccines
Vaccine Papers: An Objective Look at Vaccine Dangers	39	165	44	Dangers of Vaccination
All Vaccination Web Pages Without Medical Misinformation Mean (Standard Deviation) [Median]‡	1062.8 (7742.8) [17]	13,000.1 (303,927.3) [78]	308.2 (938.7) [77.5]	
Vaccination Medical Misinformation Web Pages Mean (Standard Deviation) [Median]	135.8 (425.0) [3]	805.0 (2375.1) [57]	144.1 (444.9) [19]	

‡ Adapted from Wolfe et al. (under review) excluding data from pages with medical misinformation.

(mean = 1062.8, standard deviation = 7742.8) was not statistically significant $F(1,1560) = 0.802$, $p = 0.37$. However, the proportion of pages presenting medical misinformation about vaccination with at least one visitor from an organic Google search per month, 54.5%, was significantly lower than for pages that did not present medical misinformation, 77.4%, $X^2(1) = 7.05$, $p = 0.008$. Backlinks from other web pages was also heavily skewed with the number of backlinks for high-traffic misinformation sites ranging from 13,900 to 0. Vaccination pages with and without medical misinformation did not differ significantly in the number of backlinks, $F(1,1544) = 0.089$, $p = 0.74$. However, the proportion of pages presenting medical misinformation about vaccination with at least one backlink from another page, 22.2%, was significantly higher than for pages that did not present medical misinformation, 8.1%, $X^2(1) = 11.87$, $p = 0.0006$. Misinformation pages did not differ from non-misinformation pages in the number of organic Google keywords bringing users to the pages, $F(1,1562) = 1.698$, $p = 0.193$. The number of organic keywords leading to high-traffic misinformation pages ranged from 2737 to 44 with a mean of 144.1 (standard deviation = 444.9). Most searches were in clusters of Google searches pertaining to dangers of vaccination and arguments against vaccination/anti-vaccination information.

The most frequent top Google search terms for medical misinformation pages were "the truth about vaccinations" (five pages), "dangers of vaccination" (three pages), "pro con vaccines" (two pages), and "Bill Gates vaccines…" (two pages). Table 2 presents to top three Google keywords for each vaccination web page presenting medical misinformation with Google traffic exceeding one hit per day, along with search volume per keyword, and how much organic traffic to each particular page was the result of searches with that keyword. For example, the top three keywords leading to a web page called *Truth Will Prevail* were "dangers of vaccination" used 2900 times/month and bringing 228 users/month to that page; "why vaccines are bad" used 2200 times/month and bringing 205 users/month to *Truth Will Prevail*, and "vaccines are bad" used 1100 times and bringing users to *Truth Will Prevail* 107 times/month. It appears that many visitors to *Truth Will Prevail* were searching for "ammunition" supporting anti-vaccination positions whereas searches using the keyword "are vaccines dangerous" leading to the web page *Dangers of Vaccinations* appear to be inquiries based on concerns about danger. Because keyword volume and traffic are heavily skewed, there were no significant differences between vaccination pages with and without medical misinformation, $Fs < 1$.

Table 2 Top three keywords with keyword volume and keyword page traffic for medical misinformation web pages with summary and comparison

Title of Web Page	Keyword 1 (Volume)	Traffic Keyword 1	Keyword 2 (Volume)	Traffic Keyword 2	Keyword 3 (Volume)	Traffic Keyword 3
Truth Will Prevail	dangers of vaccinations (2900)	228	why vaccines are bad (2200)	205	vaccines are bad (1100)	107
Should Any Vaccines Be Required for Children	pro con vaccines (300)	122	vaccines pro con (150)	70	cons of vaccines (600)	68
Vaccines - Pros & Cons	pro con vaccines (250)	86	vaccines pro con (200)	61	procon vaccines (200)	46
Jenny McCarthy: "We're Not An Anti-Vaccine Movement, We're Pro-Safe Vaccine"	Jenny McCarthy vaccines (1700)	103	Jenny McCarthy autism (700)	58	Jenny McCarthy anti vaccine (600)	46
Dangers of Vaccinations	dangers of vaccinations (1300)	188	dangers of vaccines (900)	74	are vaccines dangerous (1000)	44
The Pros & Cons of Immunization	vaccination pros and cons (400)	18	vaccine pros and cons (250)	11	vaccinations pros and cons (600)	11
How Can Vaccines Cause Damage	vaccine damaged (150)	19	vaccine damage (150)	12	do vaccines cause brain damage (30)	10
Vaccinations: Know the Risks and Failures	national vaccine information center (4900)	41	nvic (3600)	29	know the risks (100)	3.1
8 Reasons to Delay Vaccines for Kids	reasons to delay vaccinations (100)	44	getting vaccines while sick (100)	3.4	can vaccination be given during cold (50)	2.5
The Truth About Vaccines	Ty Bollinger vaccines (60)	22	truth about vaccines documentary (20)	9	the truth about vaccines documentary (200)	6
The World Needs COVID-19 Vaccines. It May Also be Overestimating Their Power	vaccines don't work (100)	17	the world may have power (150)	11	sterilizing immunity (100)	5
The Plausible Connection Between Vaccines and SIDS	SIDS and vaccines (600)	12	SIDS vaccines (300)	6	vaccines and SIDS (200)	5
Vaccine papers: An Objective Look at Vaccine Dangers	dangers of vaccinations (2900)	13	vaccination papers (60)	3.7	are vaccines dangerous (1100)	3.3
All Vaccination Web Pages Without Medical Misinformation Mean Number [Median]‡	5504.4 [150]	192.0 [6]	2039.9 [100]	72.6 [3.6]	1647.8 [100]	44.4 [2.6]
Vaccination Medical Misinformation Web Pages Mean Number [Median]	381.6 [100]	22.1 [2.7]	269.5 [100]	13.6 [0.5]	227.7 [50]	9.6 [0.6]

‡ Adapted from Wolfe et al. (under review) excluding data from pages with medical misinformation.

Table 3 examines high-traffic medical misinformation about vaccine web sites with respect to their focus, and whether they present an overarching gist, a narrative or story, and whether they try to influence the reader with an emotional appeal. The focus of high-traffic pages presenting medical misinformation was on vaccine safety and dangers, pros and cons of vaccination, and vaccines and autism. Misinformation pages were somewhat, but not significantly

Table 3 Focus, overarching gist, narrative, and emotional appeal of medical misinformation pages with summary and comparison

Title of Web Page	Focus	Overarching Gist	Narrative	Emotional Appeal
Truth Will Prevail	safety, prevention, conspiracy	-	Narrative	Emotional Appeal
Should Any Vaccines Be Required for Children	pros and cons of vaccinations	-	-	-
Vaccines - Pros & Cons	spread of info	-	-	-
Jenny McCarthy: "We're Not An Anti-Vaccine Movement, We're Pro-Safe Vaccine"	info on vaccines and autism	-	Narrative	Emotional Appeal
Dangers of Vaccinations	Vaccine safety	Gist	-	-
The Pros & Cons of Immunization	prevention, safety	-	-	-
How Can Vaccines Cause Damage	safety	-	-	-
Vaccinations: Know the Risks and Failures	vaccine risk and failure	-	-	-
8 Reasons to Delay Vaccines for Kids	Delaying vaccination	-	-	-
The Truth About Vaccines	vaccines are not as safe	-	-	-
The World Needs Covid-19 Vaccines. It May Also be Overestimating Their Power	COVID-19 vaccine	-	-	-
The Plausible Connection Between Vaccines and SIDS	Vaccine safety	-	-	-
Vaccine Papers: An Objective Look at Vaccine Dangers	Vaccine dangers	-	-	-
High Traffic Vaccination Web Pages Without Medical Misinformation: Percent of Total‡		23.2%	12.1%	6.0%
Vaccination Medical Misinformation Web Pages: Percent of Total		32.7%	7.7%	17.3%

‡ Adapted from Wolfe et al. (under review) excluding data from pages with medical misinformation.

more likely to present an overarching gist (32.7%) compared to high-traffic vaccination web sites without medical misinformation, (23.2%), $X^2(1) = 1.99$, $p = 0.159$. There were no significant difference between vaccination pages presenting medical misinformation and those without medical misinformation with respect to whether they present a narrative, $X^2(1) = 0.8$, $p = 0.37$. However, pages presenting medical misinformation were significantly more likely to make an emotional appeal to the reader (17.3%) compared to high-traffic pages that do not present medical misinformation (6.0%), $X^2(1)=6.63$, p=0.01. It can also be seen that of the top five medical misinformation pages by traffic, one has an overarching gist, two present narratives, and two make emotional appeals.

Having investigated search behavior our next step was to analyze the contents of pages presenting medical misinformation about vaccination for readability, ease of understanding, and ease of making inferences using discourse technologies. Coh-Metrix measures of readability, simplicity, gist inferences, and syntax are presented in Table 4 along with comparisons to high-traffic vaccine web pages without misinformation, and norms for social studies texts for 11th grade through adult published by McNamara et al. (2014). Here it can be seen that the most visited pages vary greatly on measures of readability and comprehension. Pages generally scored high on syntactic simplicity, arguably aiding comprehension but differed sharply on Flesch–Kincaid grade level and Gist Inference Score. For example, the page titled *Jenny McCarthy: "We're Not An Anti-Vaccine Movement: We're Pro-Safe Vaccine"* scored at a 6th grade reading level (6.246) with a relatively high GIS of 0.484. This suggests that even relatively unsophisticated readers are likely to comprehend the text. Moreover, at least for readers comparable to those participating in previous studies (e.g., Wolfe, et al., 2019; Wolfe et al., 2021), including those with low levels of health literacy and those with a high school education or less (Dandignac & Wolfe, 2020), we predict that they will be able to draw inferences about the bottom line meaning of that page. Conversely, the page *Vaccines - Pros & Cons* was written at a college freshman level (FKGL = 12.902) with a low GIS of – 0.467. Comparing pages with medical misinformation to high-traffic vaccine websites without misinformation, there were not significant differences for syntactic simplicity, Flesch reading ease, or Flesch–Kincaid grade level, Fs < 1. Misinformation pages were significantly lower on LSA verb overlap $F(1,132) = 4.65$, $p = 0.033$ indicating that actions are repeated less in pages with misinformation. Misinformation pages also scored at a lower level (easier to read) on the Coh-Metrix variable second language readability (RDL2), $F(1,132) = 8.093$, $p = .005$, suggesting that they are easier for non-native English readers to comprehend. Nevertheless, pages with medical misinformation scored significantly lower on GIS, $F(1,132) = 9.29$, $p = 0.003$, indicating that it is likely to be harder for readers to make inferences about the bottom line meaning of misinformation pages than high-traffic pages without medical misinformation.

Table 5 presents word level Coh-Metrix measures for pages presenting medical misinformation about vaccination.

Table 4 Coh-Metrix measures of readability, simplicity, gist inferences, and syntax for medical misinformation web pages with summary and comparisons

Title of Web Page	Number of Words Per Page (DESWC)	Syntactic Simplicity Percentile (PCSYNp)	Verb Overlap LSA (SMCAUSlsa)	Flesch Reading Ease (RDFRE)	Flesch-Kincade Grade Level (RDFKGL)	Second Language Readability (RDL2)	Gist Inference Score (GIS)
Truth Will Prevail	2380 analyzed of longer text	61.41	0.105	51.364	9.837	13.863	0.026
Should Any Vaccines Be Required for Children	1504	68.79	0.113	41.324	11.997	8.038	-0.282
Vaccines - Pros & Cons	3266	59.87	0.085	39.616	12.902	7.156	– 0.467
Jenny McCarthy: "We're Not An Anti-Vaccine Movement: We're Pro-Safe Vaccine"	5402	64.8	0.084	75.205	6.246	19.156	0.484
Dangers of Vaccinations	336	79.39	0.096	55.127	9.364	12.467	– 0.124
The Pros & Cons of Immunization	1224	54.38	0.099	40.665	12.83	6.385	0.051
How Can Vaccines Cause Damage	1747	59.87	0.089	29.046	14.437	7.777	– 0.060
Vaccinations: Know the Risks and Failures	3623	60.26	0.056	29.307	14.627	5.336	– 0.759
8 Reasons to Delay Vaccines for Kids	862	68.08	0.057	53.638	10.061	6.978	– 0.326
The Truth About Vaccines	1119	54.38	0.06	56.38	9.484	15.203	– 0.161
The World Needs COVID-19 Vaccines. It May Also be Overestimating Their Power	1180	68.08	0.039	49.068	10.223	6.908	– 0.225
The Plausible Connection Between Vaccines and SIDS	1641	53.59	0.055	36.594	12.841	2.975	– 0.381
Vaccine Papers: An Objective Look at Vaccine Dangers	1875	79.1	0.068	42.628	10.44	13.118	0.183
Norms for 11th Grade-Adult Social Studies Texts (Standard Deviation)†	300 (23.085)	47.311 (22.974)	0.097 (0.040)	49.059 (9.598)	11.430 (2.240)	14.039 (4.552)	NA

Table 4 (continued)

Title of Web Page	Number of Words Per Page (DESWC)	Syntactic Simplicity Percentile (PCSYNp)	Verb Overlap LSA (SMCAUSlsa)	Flesch Reading Ease (RDFRE)	Flesch-Kincade Grade Level (RDFKGL)	Second Language Readability (RDL2)	Gist Inference Score (GIS)
High-Traffic Vaccination Web Pages Without Medical Misinformation Mean (Standard Deviation)‡	1493 (1484)	59.628 (18.005)	0.079 (0.032)	45.661 (10.986)	11.322 (2.295)	10.970 (4.980)	– 0.030 (0.378)
Vaccination Medical Misinformation Web Pages Mean (Standard Deviation)	1820 (1885)	58.113 (15.292)	0.069 (0.022)	44.291 (10.915)	11.494 (2.083)	8.684 (4.183)	– 0.234 (0.389)

† From McNamara et al. (2014) Appendix B.

‡ Adapted from Wolfe et al. (under review) excluding data from pages with medical misinformation.

These measures were less variable and more consistent with high-traffic vaccination web pages without misinformation and in comparison to social studies texts for 11th grade through adult. Medical misinformation pages scored significantly higher on imagability of content words, $F(1,132) = 7.37$, $p = 0.008$ indicating that these texts bring images more readily to mind. However, they did not differ significantly for word concreteness, word familiarity, $Fs < 1$; or concreteness of content words, $F(1,132) = 2.25$, $p = 0.135$ or Colorado meaningful word norms, $F(1,132) = 2.301$, $p = 0.132$.

Coh-Metrix whole text-level measures for medical misinformation web pages are presented in Table 6. Compared to high-traffic vaccination web pages without medical misinformation, misinformation pages scored significantly lower on referential cohesion, at just the 23rd percentile, $F(1,132) = 18.99$, $p < 0.0001$, suggesting less overlap than most texts in words between sentences and paragraphs forming related threads. Misinformation pages scored significantly higher on temporality percentile, $F(1,132) = 8.52$, $p = 0.004$, indicating that misinformation pages score high on time-related coherence. There were not significant differences for narrativity, deep cohesion, verb cohesion, and connectivity, $Fs < 1$.

Discussion

Web pages presenting verified medical misinformation about vaccination differ markedly from one another in their overall popularity. Visitors per month and backlinks from other pages are heavily skewed with just a handful of pages accounting for most cases of reading medical misinformation. However, pages with medical misinformation about vaccination are significantly less likely than other vaccine pages to receive at least one visitor per month as a result of a Google search, and significantly more likely to have at least one other web page link to it. Thus, it appears that medical misinformation is relatively more likely to be shared on another page than the result of a Google search. These findings are consistent with research about the sharing of medical misinformation on Twitter (Caldarelli et al., 2021; Rosenberg et al., 2020), finding high rates of sharing disreputable URLs. Although there is evidence that political conservatives are more likely to share misinformation about vaccination (Caldarelli et al., 2021) and endorsed more misinformation (Weil & Wolfe, 2022), the political ideology of people linking to medical misinformation revealed by analyzing SEO data in the current study remains unknown, including biases of the left and right. At the level of methodology, this suggests that SEO tools such as Ahrefs (2021) are valuable because they allow researchers and public health educators to concentrate on the most influential bogus claims, rather than being side tracked by outlandish claims that are only read by a few people, whether as a result of search or sharing. Unfortunately, it also suggests that sweeping generalizations about all medical misinformation web pages should be treated with caution. However, taking organic traffic into account, it is clear that most of the medical misinformation about vaccination on the web read by people is in Google searches related to dangers of vaccination, arguments against vaccination and anti-vaccination information and, to a lesser extent, vaccine facts, alternatives to standard vaccination, the truth about vaccines (also the name of a series of anti-vaccination documentaries) and SIDS and vaccines. Some pages avoid the "myside bias" (Wolfe, 2012) presenting both arguments in favor of

Table 5 Coh-Metrix word-level measures for medical misinformation web pages with summary and comparisons

Title of Web Page	Word Concrete Percentile (PCCNCp)	Content Word Frequency (WRDFRQc)	Age of Word Acquisition (WRDAOAc)	Familiarity of Content Words (WRDFAMc)	Concreteness Content Words (WRDCNCc)	Imagability Content Words (WRDIMGc)	Colorado Meaningful Word Norm (WRDMEAc)
Truth Will Prevail	24.83	2.129	370.112	568.552	374.361	405.302	425.376
Should Any Vaccines Be Required for Children	37.45	1.924	409.124	554.053	387.732	419.877	429.727
Vaccines - Pros & Cons	59.48	1.878	391.116	554.397	397.281	429.094	437.691
Jenny McCarthy: "We're Not An Anti-Vaccine Movement" We're Pro-Safe Vaccine"	15.15	2.414	334.166	578.112	361.391	389.579	415.715
Dangers of Vaccinations	37.83	2.056	381.48	550.198	383.788	414.432	412.275
The Pros & Cons Of Immunization	42.86	1.955	396.404	558.719	369.376	406.133	431.537
How Can Vaccines Cause Damage	20.9	1.959	393.913	562.341	367.869	391.714	416.38
Vaccinations: Know the Risks and Failures	65.91	1.811	398.953	555.695	388.046	420.904	432.556
8 Reasons to Delay Vaccines for Kids	47.61	2.037	369.421	567.491	384.535	417.007	423.551
The Truth About Vaccines	43.25	2.257	384.3	574.763	372.296	401.998	432.801
The World Needs COVID-19 Vaccines. It May Also be Overestimating Their Power	19.22	2.054	383.398	567.981	368.14	396.868	427.133
The Plausible Connection Between Vaccines and SIDS	36.69	1.898	389.841	562.404	379.513	410.766	438.184
Vaccine papers: An Objective Look at Vaccine Dangers	13.79	2.071	391.529	565.724	361.501	393.57	416.795

Table 5 (continued)

Title of Web Page	Word Concrete Percentile (PCCNCp)	Content Word Frequency (WRDFRQc)	Age of Word Acquisition (WRDAOAc)	Familiarity of Content Words (WRDFAMc)	Concreteness Content Words (WRDCNCc)	Imagability Content Words (WRDIMGc)	Colorado Meaningful Word Norm (WRDMEAc)
Norms for 11th Grade-Adult Social Studies Texts (Standard Deviation)†	51.251 (27.792)	2.993 (0.106)	381.515 (31.295)	378.074 (26.879)	378.074 (26.879)	410.346 (24.994)	430.164 (17.090)
Mean for High-Traffic Vaccination Pages (Standard Deviation)‡	32.122 (17.166)	2.055 (0.173)	382.756 (18.972)	562.384 (8.292)	374.258 (13.835)	401.055 (12.373)	421.451 (11.810)
Mean for Medical Misinformation Vaccination Pages (Standard Deviation)	37.714 (20.524)	2.013 (0.149)	382.165 (18.017)	562.547 (6.966)	378.551 (18.860)	407.895 (16.486)	424.625 (11.169)

† From McNamara et al. (2014) Appendix B.

‡ Adapted from Wolfe et al. (under review) excluding data from pages with medical misinformation.

vaccination and arguments against—including con-side argument with medical misinformation—which has been found to increase the persuasiveness of arguments (Wolfe et al., 2009).

An examination of keywords people used to get to the top pages presenting medical misinformation about vaccines suggests that many are searching for arguments supporting anti-vaccination positions (e.g., "vaccines are bad"), while some such as "do vaccines cause brain damage" appear to be actual inquiries. Knowing these keywords may be an asset for researchers wishing to conduct research on different "communities of inquiry" (Wolfe et al., under review) by targeting groups for recruitment with selected Google AdWords. Select AdWords may further be used to for targeted messaging to counteract medical misinformation, for example by providing different interventions to people searching for information about delaying a child's vaccination than to those seeking information about vaccines and autism.

We found that almost a third of the pages presenting medical misinformation about vaccines had an overarching gist. Whereas GIS assesses the likelihood that readers can make inferences from a text about bottom line meaning, overarching gist judgments are about whether the entire text supports a single coherent message that can be distilled into simple meaningful mental representation. In coding, this was operationalized as a pithy statement. We did not find significant differences in overarching gist between pages presenting medical misinformation about vaccination and other vaccine web pages. However, fuzzy-trace theory suggests that understanding the bottom-line meaning of pages with medically valid information is important to counteracting misinformation. Indeed an "Achilles' heel" of misinformation pages is that they score relatively low on Gist Inference Scores (Wolfe, Dandignac, & Reyna, 2019a; Wolfe et al., 2019; Dandignac & Wolfe, 2020; Wolfe et al., 2021), suggesting that gist-based interventions rooted in fuzzy-trace theory may be particularly effective (Blalock & Reyna, 2016; Reyna, 2020).

The finding that misinformation pages are more likely to make an emotional appeal to the reader (17.3% compared to 6.0% for high-traffic pages that do not present medical misinformation) further suggests that they may disrupt the cognitive elaboration of message content (Rosselli et al., 1995). Fuzzy-trace theory suggests that gist-based interventions that target core values and affect may be useful antidotes to such misinformation (Reyna, 2020). In contrast to traditional expected utility and standard dual process theories, which have sometimes contrasted emotion with rational thought, fuzzy-trace theory suggests that health communicators may be well served by linking messages that communicate the bottom line meaning of factually accurate medical information and evoking core values to elicit emotions in the service of health-promoting decisions and actions.

Although less than 8% of pages presenting medical misinformation utilized narratives, which was not significantly different than high-traffic vaccination sites without medical misinformation, of the top five medical misinformation

Table 6 Coh-Metrix whole text-level measures for medical misinformation web pages with summary and comparisons

Title of Web Page	Narrativity Percentile (PCNARp)	Referential Cohesion Percentile (PCREFp)	Deep Cohesion Percentile (PCDCp)	Verb Cohesion Percentile (PCVERBp)	Connectivity Percentile (PCCONNp)	Temporality Percentile (PCTEMPp)
Truth Will Prevail	27.76	11.9	59.87	47.21	11.7	55.57
Should Any Vaccines Be Required for Children	8.85	24.2	49.6	4.46	0.94	34.09
Vaccines - Pros & Cons	8.23	37.83	68.08	2.22	1.39	52.39
Jenny McCarthy: "We're Not An Anti-Vaccine Movement, We're Pro-Safe Vaccine"	84.13	35.94	71.57	53.59	3.14	65.54
Dangers of Vaccinations	23.89	13.79	77.04	5.59	0.12	31.21
The Pros & Cons of Immunization	17.88	30.85	88.88	12.71	0.45	79.95
How Can Vaccines Cause Damage	8.38	20.61	55.17	17.62	0.99	44.43
Vaccinations: Know the Risks and Failures	8.38	51.99	55.96	0.68	0	55.17
8 Reasons to Delay Vaccines for Kids	22.96	21.77	95.99	1.79	0.1	57.53
The Truth About Vaccines	37.45	28.1	80.51	37.07	5.48	29.81
The World Needs COVID-19 Vaccines. It May Also be Overestimating Their Power	36.32	11.51	35.94	4.18	10.93	52.79
The Plausible Connection Between Vaccines and SIDS	14.23	12.3	73.57	8.53	2.28	87.29
Vaccine papers: An Objective Look at Vaccine Dangers	28.43	18.41	88.1	16.6	1.66	75.49
Norms for 11th Grade-Adult Social Studies Texts (Standard Deviation)†	25.892 (17.196)	39.602 (25.268)	60.029 (25.962)	41.409 (27.942)	7.839 (14.163)	47.270 (30.021)
Mean for High-Traffic Vaccination Pages (Standard Deviation)‡	26.197 (13.608)	40.564 (25.325)	66.913 (20.200)	16.029 (18.439)	5.668 (11.777)	45.017 (22.870)
Mean for Medical Misinformation Vaccination Pages (Standard Deviation)	24.713 (16.189)	22.888 (17.939)	66.559 (17.690)	16.906 (17.756)	5.573 (8.338)	56.616 (21.650)

† From McNamara et al. (2014) Appendix B.

‡ Adapted from Wolfe et al. (under review) excluding data from pages with medical misinformation.

pages by traffic, two present narratives. Shaffer et al. (2013) found important differences between process-focused narratives and experience-focused narratives in decisions about cancer treatment. Thus, the use of narrative in communication about vaccination warrants further research attention.

Automated psycholinguistic analyses with Coh-Matrix suggest that there are few distinguishing characteristics of medical misinformation pages. Indeed, pages presenting medical misinformation about vaccines typically look like other pages presenting information about vaccination, with some even associated with health care providers (i.e., chiropractors, "naturopathic physicians," homeopaths) and children's and women's health advocates.

It is unlikely that Coh-Metrix could be used to automatically screen for medical misinformation without the judgment of expert human judges, though machine learning and deep learning (LeCun et al., 2015) may offer a promising avenue. Unfortunately, these data also show that some highly visited pages presenting medical misinformation are accessible to less skilled readers. A question for further research is how a person's reading level, the FKGL of texts, and text GIS interact. The impact of medical misinformation on readers for whom English is a second language (ESL) also deserves more research attention (Wolfe et al., 2021). The finding that misinformation pages scored significantly lower on the Coh-Metrix Second Language Readability measure (McNamara et al., 2014) than other vaccine pages suggests that such readers may be more vulnerable to medical misinformation than others, which is concerning.

Medical misinformation pages scored significantly higher on imagability of content words, and significantly lower on referential cohesion. These appear to be the drivers of the significant difference on GIS between medical misinformation pages and others highly visited pages about vaccination. In developing written interventions to counteract medical misinformation, the authors would be wise to express accurate medical information and arguments for vaccination at an appropriate level of conceptual abstraction with relatively higher cohesion to help readers form coherent gist representations of key concepts.

Limitations

This research is limited in the following ways: first, the scope of this research is on text-based web pages and it excludes dynamic social media and on-line videos. Although our study did not consider video and other dynamic content, web pages often serve as a gateway to other content in other formats (video, social media, etc.), and as such, this study adds a piece of the larger puzzle, illuminating one vital component of a larger infrastructure in the online medical misinformation ecosystem. Notably, an article suggesting that the coronavirus vaccine could lead to death was the most visited Facebook link in the U.S. from January through March 2021 (Dwoskin, 2021). YouTube videos are also a source of medical misinformation, and people are more likely to reach antivaccine videos through direct navigation from another antivaccine video rather than through goal-oriented browsing (Tang et al., 2021). A shortcoming of Coh-Metrix (2021) is that it cannot be used to assess infographics or video, only text. Our procedure for identifying medical misinformation is conservative in the classic sense. All judgments about medical misinformation were made by trained coders who were instructed to flag pages that could possibly contain medical misinformation. Those initial judgments were confirmed by a public health expert. Thus, although some coders produced initial "false alarms" that were identified by public health expert, none of the misinformation pages analyzed here was included without two judgments. Thus we have confidence that the pages identified as presenting medical misinformation do so, we cannot rule out the possibility that we missed some pages with medical misinformation. More broadly, collecting the diverse collection of vaccine web pages "by hand" as opposed to using a spider or a limited number of search terms selected a priori was labor intensive, and it is not likely that this procedure missed many high-traffic web sites. However, it is possible that the web pages we analyzed and the pages we discovered presenting verified medical misinformation are the result of biases in our initial searches, though we introduced procedures to minimize biases. Finally, this research represents a "snapshot" in time, and some aspects of the landscape of web-based medical misinformation about vaccination are likely to have changed since we collected these data.

Conclusions

In both the number of visitors per month and the number of backlinks from other web sites, web pages presenting verified medical misinformation about vaccination differ sharply from one another with a relative handful of pages having the lion's share of impact. Many Google searches resulting in medical misinformation are related to the safety of vaccines, but the way questions are framed (i.e., vaccine safety vs. dangers of vaccination) is a good predictor of whether or not a search will result in medical misinformation. Additionally, many searches appear to reflect a pre-existing antivaccination position, with the people apparently Googling to find support for their beliefs. Relatively speaking, pages presenting medical misinformation about vaccination are more likely to be shared with web links, but less likely than non-misinformation pages to be found in a Google search. Overall, vaccine hesitancy should not be painted with broad brushstrokes given notable differences in search behavior.

Web pages presenting medical misinformation are in many ways similar to other web sites about vaccination, though they are more likely to appeal to the reader's emotions. Unfortunately, several high-traffic web pages conveying medical misinformation are written at an accessible grade level with characteristics likely to help readers form a mental representation of the bottom-line meaning of the text. Nonetheless, as a whole, many of these pages are relatively weak in promoting gist inferences, and fuzzy-trace theory suggests that gist-based interventions designed to cue core values and help people understand the meaning of complex information might effectively compete with misinformation thus reducing the impact of medical misinformation. This approach should be assessed in future research.

Acknowledgements This research was supported by a Miami University Grant to Improve Research and a Summer Research Appointment. Preparation of this manuscript was supported in part by a grant from the National Science Foundation to Valerie F. Reyna (SES 2029420). We wish to thank Dornu Biragbara, Alyssa Conyers, Lydia Copeland, Alexia Harding, Anne Hardy, Meghana Muvva, Danielle Nabor, Will Wagner, Cynthia Wang, and Tori Ward, for their capable assistance with data collection; and Trung Le, Lydia Copeland, Alexia Harding, Anne Hardy, and Tori Ward for assistance in coding data and establishing inter-rater reliability. The authors do not report any conflicts of interest.

None of the data or materials for the experiments reported here is available, and none of the experiments was preregistered.

References

Ahrefs. Ahrefs by the numbers. https://Ahrefs.com/big-data. Accessed August 23, 2021.

Basch, C. H., Fera, J., & Garcia, P. (2019). Readability of influenza information online: implications for consumer health. *American Journal of Infection Control, 47*(11), 1298–1301. https://doi.org/10.1016/j.ajic.2019.04.178

Bernard, R., Bowsher, G., Sullivan, R., & Gibson-Fall, F. (2021). Disinformation and epidemics: Anticipating the next phase of biowarfare. *Health Security, 19*(1), 3–12. https://doi.org/10.1089/hs.2020.0038

Blalock, S. J., & Reyna, V. F. (2016). Using fuzzy-trace theory to understand and improve health judgments, decisions, and behaviors: A literature review. *Health Psychology, 35*, 781–792.

Broniatowski, D. A., & Reyna, V. F. (2020). To illuminate and motivate: A fuzzy-trace model of the spread of information online. *Computational and Mathematical Organization Theory, 26*(4), 431–464. https://doi.org/10.1007/s10588-019-09297-2

Caldarelli, G., De Nicola, R., Petrocchi, M., Pratelli, M., & Saracco, F. (2021). Flow of online misinformation during the peak of the COVID-19 pandemic in Italy. *EPJ Data Science, 10*(1), 34.

Cedillos-Whynott, E. M., Wolfe, C. R., Widmer, C. L., Brust-Renck, P. G., Weil, A. M., & Reyna, V. F. (2016). The Effectiveness of argumentation in tutorial dialogues with an intelligent tutoring system. *Behavior Research Methods, 48*(1), 857–868. https://doi.org/10.3758/s13428-015-0681-1

Chin, C. L., Su, W. Y., & Chin, J. (2020). Representing the true and false text information about human papillomavirus vaccines. In *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care* (Vol. 9, No. 1, pp. 317–321). SAGE Publications.

Clickstream (2021). ClickStream: The smarter SEO agency. https://clickstream.cc/. Accessed August 13, 2021.

Coh-Metrix (2021). Accessed August 29, 2021 from http://141.225.61.35/cohmetrix2017

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*(1), 497–505. https://doi.org/10.1080/14640748108400805

Dandignac, M., & Wolfe, C. R. (2020). Gist Inference Scores predict gist memory for authentic patient education cancer texts. *Patient Education and Counseling, 103*(8), 1562–1567. https://doi.org/10.3758/s13428-015-0681-1

Duchsherer, A., Jason, M., Platt, C. A., & Majdik, Z. P. (2020). Immunized against science: Narrative community building among vaccine refusing/hesitant parents. *Public Understanding of Science (Bristol, England), 29*(4), 419–435. https://doi.org/10.1177/0963662520921537

Dwoskin, E. (2021). Facebook says post that cast doubt on COVID-19 vaccine was most popular on the platform from January through March. The Washington Post. Accessed August 23, 2021 from https://www.washingtonpost.com/technology/2021/08/21/facebook-coronavirus-vaccine/

Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making, 25*(4), 398–405. https://doi.org/10.1177/0272989X05278931

Frontline (2015). *Jenny McCarthy: "We're not an anti-vaccine movement… We're pro-safe vaccine*. Retrieved from the web on January 29, 2022 from https://www.pbs.org/wgbh/frontline/article/jenny-mccarthy-were-not-an-anti-vaccine-movement-were-pro-safe-vaccine/

Gwet, K. L. (2014). Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters. *Advanced Analytics, LLC*.

Hillyer, G. C., Beauchemin, M., Garcia, P., Kelsen, M., Brogan, F. L., Schwartz, G. K., & Basch, C. H. (2020). Readability of cancer clinical trials websites. *Cancer Control, 27*(1), 1–6. https://doi.org/10.1177/1073274819901125

Hussain, A., Ali, S., Ahmed, M., & Hussain, S. (2019). The anti-vaccination movement: a regression in modern medicine. *Cureus, 10*(7), e2919. https://doi.org/10.7759/cureus.2919

Jaiswal, J., & Halkitis, P. N. (2019). Towards a more inclusive and dynamic understanding of medical mistrust informed by science. *Behavioral Medicine, 45*(2), 79–85.

Jimenez, A. . V., Mesoudi, A., & Tehrani, J. J. (2020). No evidence that omission and confirmation biases affect the perception and recall of vaccine-related information. *PLoS ONE, 15*(3), e0228898. https://doi.org/10.1371/journal.pone.0228898

Krishna, A., & Thompson, T. L. (2021). Misinformation about health: A review of health communication and misinformation scholarship. *American Behavioral Scientist, 65*(2), 316–332.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

MacLean, S. A., Basch, C. H., Ethan, D., & Garcia, P. (2018). Readability of online information about HPV Immunization. *Human Vaccines & Immunotherapeutics, 15*(7–8), 1505–1507. https://doi.org/10.1080/21645515.2018.1502518

Madraki, G., Grasso, I.M., Otala, J., Liu, Y., & Matthews, J. (2021). Characterizing and comparing COVID-19 misinformation across languages, countries and platforms. In *Companion Proceedings of the Web Conference 2021* (pp. 213–223).

Martino, J. (2020). CDC admits in federal court they have no evidence "vaccines don't cause autism." Retrieved from the web on January 29, 2022 from https://www.lewrockwell.com/2020/

03/no_author/cdc-admits-in-federal-court-they-have-no-evidence-vaccines-dont-cause-autism/
McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
Pluviano, S., Watt, C., & Della Sala, S. (2017). Misinformation lingers in memory: Failure of three pro-vaccination strategies. *PloS One, 12*(7), 1–15. https://doi.org/10.1371/journal.pone.0181640
Reyna, V. F. (2008). A theory of medical decision making and health: Fuzzy trace theory. *Medical Decision Making, 28*(1), 850–865. https://doi.org/10.1177/0272989X08327066
Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in fuzzy-trace theory. *Judgment and Decision Making, 7*(3), 332–359.
Reyna, V. F. (2020). Of viruses, vaccines, and variability: Qualitative meaning matters. *Trends in Cognitive Science, 24*, 672–675.
Reyna, V. F., Broniatowski, D. A., & Edelson, S. (2021). Viruses, Vaccines, and COVID-19: Explaining and Improving Risky Decision-making. *Journal of Applied Research in Memory and Cognition, 10*(4), 491–509.
Rosenberg, H., Syed, S., & Rezaie, S. (2020). The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Canadian journal of emergency medicine, 22*(4), 418–421.
Rosselli, F., Skelly, J. J., & Mackie, D. M. (1995). Processing rational and emotional messages: The cognitive and affective mediation of persuasion. *Journal of Experimental Social Psychology, 31*, 163–190.
Shaffer, V. A., Hulsey, L., & Zikmund-Fisher, B. J. (2013). The effects of process-focused versus experience-focused narratives in a breast cancer treatment decision task. *Patient Education & Counseling, 93*, 255–264.
Stahl, J.-P., Cohen, R., Denis, F., Gaudelus, J., Martinot, A., Lery, T., & Lepetit, H. (2016). The impact of the web and social networks on vaccination. New challenges and opportunities offered to fight against vaccine hesitancy. *Medecine et maladies infectieuses, 46*, 117–122.
Statistica (2022a). Worldwide desktop market share of leading search engines from January 2010 to December 2021. Retrieved from the web on January 29, 2022 from https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/
Statistica (2022b). Market share of selected leading mobile search providers in the United States from October 2012 to September 2021. Retrieved from the web on January 29, 2022 from https://www.statista.com/statistics/511358/market-share-mobile-search-usa/
Studts, J. L., Ruberg, J. L., McGuffin, S. A., & Roetzer, L. M. (2010). Decisions to register for the National Marrow Donor Program: Rational vs emotional appeals. *Bone Marrow Transplantation, 45*(3), 422–428. https://doi.org/10.1038/bmt.2009.174
Tang, L., Fujimoto, K., Amith, M., Cunningham, R., Costantini, R. A., York, F., et al. (2021). "Down the rabbit hole" of vaccine misinformation on YouTube: Network exposure study. *Journal of Medical Internet Research, 23(1)*, e23262. https://doi.org/10.2196/23262
Weil, M. A., & Wolfe, C. R. (2022). Individual differences in risk perception and misperception of COVID-19 in the context of political ideology. *Applied Cognitive Psychology, 2022(36)*, 19–31. https://doi.org/10.1002/acp.3894
Wolfe, C. R. (2021). Fuzzy-trace theory and the battle for the gist in the public mind. *Journal of Applied Research in Memory and Cognition, 10*, 527–531.
Wolfe, C. R. (2012). Individual differences in the "MySide bias" in reasoning and written argumentation. *Written Communication, 29*, 474–498.
Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication, 26*, 183–209.
Wolfe, C. R., & Dandignac, M. (2021). Revising flash fiction for Coh-Metrix: Experiential learning with discourse technologies. *Journal on Excellence in College Teaching, 32(1)*, 123–140 http://celt.miamioh.edu/ject/issue.php?v=32&n=1
Wolfe, C. R., Dandignac, M., & Reyna, V. F. (2019a). A theoretically motivated method for automatically evaluating texts for gist inferences. *Behavior Research Methods, 51*(6), 2419–2437. https://doi.org/10.3758/s13428-019-01284-4
Wolfe, C. R., Dandignac, M., Sullivan, R., Moleski, T., & Reyna, V. F. (2019b). Automatic evaluation of cancer treatment texts for gist inferences and comprehension. *Medical Decision Making, 39*(8), 939–949. https://doi.org/10.1177/0272989X19874316s
Wolfe, C. R., Dandignac, M., Wang, C., & Lowe, S. R. (2021). Gist Inference Scores predict cloze comprehension "in your own words" for native, not ESL readers. *Health Communication*, early online access, https://doi.org/10.1080/10410236.2021.1920690