

**Misinformation and Covid-19
in Popular Health Websites**

Capstone Project Report
Master of Science in Informatics and Analytics

University of North Carolina
at Greensboro

Jeanne Reppert
April 11, 2021

1 Abstract

During the pandemic, websites providing “alternative” health advice and anti-vaccine messages shifted their focus to Covid-19. Many of these websites now advocate against Covid-19 vaccines and promote other Covid-19 myths, thus fueling the problem of vaccine hesitancy. This project investigates content from three websites (mercola.com, childrenshealthdefense.org, and greenmedinfo.com) along with their social media activity. While archives of articles prior to 2020 were scraped, this study more closely investigates articles and posts from January 2020 to December 2020. A variety of criteria were used to select the websites for this study including Newsguard, a service that rates websites for their level of adherence to journalistic principles and their inclusion of misinformation. All Facebook and Twitter posts made by these organizations were collected through Massmine and CrowdTangle. Network traffic data was collected through Semrush.

Word frequency analysis, word correlations, and sentiment analysis were employed to confirm that the body of articles studied contained both innocuous health information and a significant concentration Covid-19 misinformation. It was observed that the strategy embraced by these websites utilizes social media platforms to attract subscribers allowing daily information to be shared through email and SMS messaging. While some social media platforms have taken steps to limit the spread of misinformation, all three websites currently have some presence on at least Twitter or Facebook.

In addition to serving as a primary descriptive tool for investigating Covid-19 misinformation during the pandemic, data collected was further utilized to construct machine learning models to discern between articles that spread misinformation and articles that contain information that is generally accepted by the scientific and medical community. An additional website containing a collection of articles about Covid-19 from a variety of news websites which circulate reliable medical information was scraped for this purpose (medscape.com). After tagging and creating a dataset of more than 1750 observations, a machine learning model was applied. This model was able to correctly identify articles containing misinformation or reliable medical information with a surprisingly high level of accuracy.

2 Introduction

As the Covid-19 pandemic unfolded, the popularity of websites reputed for spreading misinformation has grown significantly. Some attribute this growth to the initial lack of information about the virus and the adjustments that medical professionals have made to public policies as new information was revealed. Distrust has grown as public policies encroached on citizens’ perceived freedoms. Anti-vaxxers, who have strong organizational structures and an existing following, harnessed public distrust to amplify the uncertainty surrounding Covid-19 related policies. Many leaders propagate conspiracy theories and use fear tactics to gain support. Key public health officials and philanthropists such as Anthony Fauci and Bill Gates have been targeted as villains.

Identifying and understanding online medical misinformation is essential for public health governance. In addition to flagging problematic websites and social media posts, public health officials must directly refute conspiracy theories and specific scientific misinformation spread online. Even before the Covid-19 pandemic, the spread of misinformation had been characterized by the utilization of “firehosing”, a strategy often used to describe political propaganda where as much misinformation is disseminated as frequently as possible until it is difficult and even exhausting for the reader to discern between what is true and what is false (Tran, 2019). Since the beginning of the pandemic, misinformation related to Covid-19 has grown significantly, leading concerned researchers to refer to the phenomena as an “infodemic” (Ball & Maxmen, 2020). Currently, with the roll-out of the vaccine over the past few months, vaccine hesitancy, fueled by the persistent flow of medical misinformation has become a significant concern (Alexandra S. Levine, 2020). While some suggest censorship and deplatforming as a solution to the growth of medical misinformation, others argue that engaging false narratives about misinformation with sound arguments will more effectively stem the tide of misinformation (Burki, 2020). Thus, a crucial component of engaging these narratives is gaining a thorough understanding of the misinformation that is spread. By observing rhetorical patterns and themes utilized by those who spread misinformation online, a set of standards might be developed to assess websites and rate their reliability and trustworthiness.

This project investigates content from websites known to pass on medical misinformation related to Covid-19 along with their social media activity on Facebook and Twitter. For years, many medical professionals have hosted websites that provide “alternative” health advice and frequently market health related products. During the pandemic, these websites have used issues related to Covid-19 to attract readers and followers. Often the information they share is not widely accepted by the larger medical community and is thus viewed as medical misinformation. Prior to Covid-19, these same websites pressed an anti-vaccine agenda while also offering more innocuous advice about vitamins and broader health related practices. The websites offer a similar anti Covid-19 vaccine message, which is driving the current problem of vaccine hesitancy in response to the push to achieve herd immunity through vaccination.

According to the authors of the United Nations Children’s Fund Vaccine Misinformation Management Field Guide, misinformation is defined as “false information that’s shared by people who don’t realize it is false and don’t mean any harm” while disinformation is “deliberately engineered and disseminated false information with malicious intent or to serve agendas” (Thompson & Finnegan, 2020). While this definition is helpful for a basic understanding of the dynamics of misinformation and disinformation, it still leaves open the question as to who is responsible for defining false information. For this reason, for the purpose of this study, Covid-19 misinformation will be defined as views or ideas that are generally contradictory to information that is widely accepted by the larger medical and scientific community. Unlike Thompson and Finnegan’s definition, less emphasis is placed on the motivation of the purveyor of the information (hence the dichotomy between misinformation and disinformation) and more on defining the reliability of that information.

2.1 mercola.com

After documenting a variety of websites containing specific Covid-19 misinformation, three websites were selected to perform an in-depth study of medical misinformation in the context of each website's archived articles. The first website, Mercola.com, is known as one of the largest proponents of "alternative" health information (Mercola.Com - #1 Natural Health Website). His website has published articles since 2008. Mercola earns his living by directing his followers to an online store where vitamins and other health related products are sold. According to Stephen Barrett of Quackwatch.org (Barrett, 2020), Mercola had a net worth in 2017 of more than \$100 million. Mercola's website states that he has more than a million subscribers. A typical email subscriber will receive one daily newsletter with links to a few articles. While some articles reference a short description of a supplement or health practice, other articles in the daily newsletter contain information that would typically fall under the category of misinformation including common conspiracy theories, anti-vaccination information, and debunked ideas pertaining to the Covid-19 virus.

Mercola's activities in spreading Covid-19 misinformation have not gone unnoticed by social media platforms or the press. In 2019 a Washington Post article detailed ties held by Mercola to the anti-vaccine community including the National Vaccine Information Center, an organization run by Barbara Loe Fisher, who also shares Covid-19 misinformation regularly on her the NVIC website as well as on various social media platforms including Twitter and Facebook. According to the article, Mercola has donated almost 40 percent of this organization's funding (2.9 million dollars) to support anti-vaccination efforts along with another additional 4 million dollars to support other anti-vaccine advocacy groups (Satija & Sun, 2019). Interestingly, according to its website, NVIC was recently deplatformed from Facebook for spreading misinformation regarding Covid-19 vaccines (National Vaccine Information Center - Your Health. Your Family. Your Choice.). Ironically, during the pandemic, Mercola has also received \$335,000 in federal pandemic relief loans from the Paycheck Protection Program. The program also granted loans to NVIC, Children's Health Defense and two other organizations reputed to share Covid-19 and vaccine misinformation online (Cowley, 2021).

While claiming to have 1.8 million followers on Facebook, in July of 2019, Mercola announced on his website and Facebook page that he was leaving the platform due to the social media company's privacy policies (Forget Facebook – Take Control of Your Privacy by Dr. Mercola). He instead encourages his followers to sign up for his daily email newsletter. He currently maintains a platform on YouTube with 378K subscribers, a Twitter account with 289.4K followers, a Telegram account with 4.4k followers and an Instagram account with 286K followers. Mercola's articles are also shared on various other websites including NVIC and Health Nut News (another website that publishes Covid-19 misinformation) run by Erin Elizabeth, who lives with Mercola (Barrett, 2020). Mercola maintains Facebook pages in ten other languages (Facebook's Algorithm, 2020) including a Spanish language page with more than a million followers. Mercola will release a book titled "The Truth about Covid-19: Exposing the Great Reset, Lock-downs, Vaccine Passports, and the New Normal" in April of this year (Mercola & Cummings, 2021) which is being presold on several sites including Amazon. The foreword is written by Robert F. Kennedy of Children's Health Defense.

NewsGuard, which rates websites based on 9 journalistic standards, gives Mercola's website a score of 35 out of 100 (NewsGuard mercola.com Nutrition Label).

2.2 **childrenshealthdefense.org**

The second website investigated, Children's Health Defense, was created by Robert F. Kennedy, a lawyer known for his work in environmental law and son of the late U.S. Attorney General Robert F. Kennedy. The organization was formerly known as the World Mercury Project (created in 2016) but was changed to CHD in May 2018 (NewsGuard Childrenshealthdefense.Org Nutrition Label). In a study published in January 2020, researchers found that CHD was one of two organizations which were responsible for the largest portion of anti-vaccine advertising on Facebook (Jamison et al., 2020). CHD's news archive, while less extensive than Mercola's, focuses mostly on health issues related to children including vaccinations, the environment, fluoride, and other health-related concerns. One study finds that since late 2019, CHD began posting anti 5G information every week (Mostrous, 2020.) Since the pandemic, CHD's content has shifted significantly to subject matter pertaining to Covid-19 misinformation.

CHD maintains an active social media presence with 151.8k Facebook followers, 57.2k followers on Twitter, 203k followers on Instagram, and 42.8k subscribers on their YouTube channel. Additionally, Robert F. Kennedy maintains a personal social media presence on Twitter with 233.9k followers, 312.7k followers on his personal Facebook page, and a smaller presence on his YouTube channel with 2k subscribers. His personal account was recently banned by Instagram because of persistent sharing of misinformation (Iyengar, 2021). CHD's Instagram account, however, has remained active. In a study of combined Instagram and Facebook posts about medical misinformation, Tortoise Media found that 7% of social media posts are responsible for 80% of measured interactions on the platforms. Of the 156 superspreaders they identified in the combined accounts; Robert F. Kennedy's was found to be the number one superspreader among the 145,000 investigated posts with a combined total followers of 665k who accounted for 486k interactions in the data. Mercola and Erin Elizabeth rated sixth with a combined following of 630k followers and 64k interactions (Super-Spreaders, 2020). NewsGuard assigned CHD's website a score of 12.5 out of 100 (NewsGuard Childrenshealthdefense.Org Nutrition Label).

2.3 **greenmedinfo.com**

The last website investigated, GreenMedinfo, was created by Sayer Ji, who holds an undergraduate degree in philosophy (Jarry, 2019). The website was created in 2008 and offers a variety of articles written by GreenMedinfo staff and articles reposted by other alternative health websites (NewsGuard GreenMedinfo.Com Nutrition Label.) In March of this year Twitter deleted GreenMedinfo's account after the Center for Countering Digital Hate named GreenMedinfo to be one of their top ten anti-vaccine websites (Ji, 2021). Ji still maintains a personal Twitter account with a much smaller following of 9.3k. GreenMedinfo's Facebook page has a total of 536.5k followers, however the last post made on February 26, 2021 claims that only .1% of followers are receiving the posts and encourages followers to

join GreenMedinfo on Telegram and MeWe as well as subscribe to its newsletter on the GreenMedinfo website. Currently the Telegram group has 43k members and the MeWe group has 6.6k members. GreenMedinfo's Facebook page claims that their daily newsletter is received by 450k subscribers (GreenMedinfo.Com | Facebook). GreenMedinfo's Instagram account was deleted "without explanation" in September 2020 according to Ji (Ji, 2020) however in a July GreenMedinfo article, Ji states that Instagram had threatened to delete the account because of misinformation related to the Covid-19 pandemic (Ji, 2020). Interestingly, the same article questions whether their LinkedIn account will also be deleted due to censorship, however this account only shows a total of 560 followers. NewsGuard assigned the GreenMedinfo website a score of 30 out of 100 (NewsGuard GreenMedinfo.Com Nutrition Label).

After studying the data associated with the above three websites, several observations were made. First, while the articles on the website include a variety of more general health related material, they also promote some very specific anti-vaccination content and content that embraces several common conspiracy theories related to Covid-19. This information is often grouped around a small subset of topics related to Covid-19 including mitigation measures, treatments, origin stories, vaccine efforts, and conspiracy theories related to technology, purported world domination by technocrats and Covid-19. Second, the strategy embraced by the operators of these websites appears to utilize social media platforms to attract subscribers their sites to share daily information through email and SMS messaging. Additionally, various social media platforms appear to have had some level of success in limiting the spread of misinformation although in many cases misinformation from these websites continues to flow in some capacity. Twitter appears to be less successful in stemming the propagation of misinformation by those who utilize the platform. Although GreenMedinfo's account was recently deleted, Children's Health Defense and Mercola still continue to share a wide variety of Covid-19 misinformation on Twitter and use the platform to direct followers to their websites for more detailed misinformation and stories that feature common conspiracy theories.

3 Problem Definition/Hypotheses

Websites that pass on medical misinformation tend to follow a specific pattern of communication. Most websites have some innocuous articles that pertain to health issues including advice about supplements, healthy eating, exercise, and healthy lifestyles. These articles may be attractive to those who are mildly interested in health topics but not typically drawn to conspiracy theories that stretch the imagination more than the simple idea of trying a vitamin or other supplement. These types of innocuous articles may be a kind of lure or "gateway drug" that draws the user in. Most websites press the viewer to share an email to gain access to articles and once that email is captured, most websites will send a daily news digest that highlights additional reading material. Thus, over time, the user is exposed to a variety of topics including the more nefarious conspiracy theories that would be categorized as medical misinformation. One might expect then to observe in the text data, that during times when uncertainty is more prevalent due to a health crisis, articles may skew more towards current events and during times when uncertainty in public health is

less of a concern that topics will skew more towards articles about healthy lifestyles. One may also hypothesize that social media sites, for this type of misinformation, are more utilized to form connections to readers and to draw traffic to the website rather than to necessarily gain converts to the social media site itself. The real value in a tweet or a Facebook post in this case is not necessarily to generate discussion or retweets but to gain clicks that will lead users to the website for the purpose of capturing their email and following up with additional newsletters and ultimately product advertisements.

4 Methods

4.1 Website Selection

After a careful review of the literature regarding online misinformation and its impact on the Covid-19 pandemic, a survey of popular alternative health websites was undertaken. While there are many websites that contain a mixture of political and health related materials, for this study, only websites that focused nearly exclusively on health topics were considered. Lists found on various websites that describe alternative health websites were consulted. Some of these lists included Quackwatch (Barrett, 2020), and various pages posted on NewsGuard on the Coronavirus Misinformation Tracking Center page (detailed lists and data were formally publicly shared although they are no longer available except by request). Reports published by the Center for Countering Digital Hate were also consulted (counterhate.com). Just this month CCDH released an additional study which named Joseph Mercola, Robert F. Kennedy, and Sayer Ji as among a list that they identify as the “disinformation dozen”, a group that they find is responsible for most of the current vaccine misinformation being spread on social media (The Disinformation Dozen | Center for Countering Digital Hate).

Finally, using Massmine, a collection of tweets was compiled using common hashtags associated with Covid-19 misinformation (#vaccineskill, #nomaskonme, #scamdemic, #covidhoax, #vaccineinjury, #nomaskmandates, #Vaxxed2, #nomorelockdowns, #masksdontwork, #vaxxed, #nochip, #learntherisk, #antimask, #BillGates, #plandemic, #stopbillgates, #iwillnotcomply, #Lockdownskill, #nogates, #NoCovidVax, #burnyourmask, #CovidisOver, #nomasks, #endlockdownsnow, #covidots, #nomask, #HerdImmunity). By following links listed on Twitter posts and investigating the number of responses and followers, frequently mentioned websites were identified. Ultimately the resulting data was used to compose a database of 38 websites which included additional information about the individuals who were named as authors of the sites along with key players who were associated with the websites. Twitter and Facebook usernames along with the number of followers and likes for each account were also researched and recorded.

Each website in the database was investigated to gauge the number of articles on the site that were available and the type of content available. Some websites (like NVIC) were eliminated simply because there were few articles that existed on the website while other websites (like HealthNut News by Erin Elizabeth) were eliminated because a larger portion of the articles on the website were reprints from Mercola’s website. Finally, other websites

were eliminated because their content consisted of mostly videos or “courses” that individuals could enroll in to learn more about the information shared on the website. After applying these criteria, three websites were chosen for scraping and data gathering: mercola.com, childrenshealthdefense.org, and greenmedinfo.com.

4.2 Data Collection

Given the aim of our study was to track changes in web page and social media content during the Covid-19 pandemic, the full corpus of archived articles published by Mercola, CHD, and GreenMedinfo in 2020 were scraped from the websites. Using both RSelenium and rvest packages in R, URLs were identified for potential scraping. Some of the URLs identified had broken links and were eliminated. Additionally, URLs that were not associated with health articles or which contained very little text (either recipes, exercise illustrations or videos) were also omitted. Once these URLs were eliminated, the text body of the article, article date, and article title were scraped and imported into CSV files that contained the url, date, title and text. Because many articles had inconsistencies in the way in which author information was coded and many do not post an author name at all, author data was not collected.

CrowdTangle was used to gather all available posts made on the associated Facebook group pages during 2020 and a history of 2020 tweets from each organization were gathered using MassMine. Traffic analysis data was acquired from Semrush.com in the form of csv files of traffic sources and top page traffic. The data was investigated to note patterns and trends in the texts of the website articles and how those compared to the social media content and web page data. Traffic source csv files were compiled and visualized to show monthly website traffic trends along with the top sources for website traffic each month. Top page files were extracted for each month and December’s top page file was cleaned and trimmed to visualize articles with the most frequent unique visitors for all three websites during the month of December.

4.3 Data Descriptions

While the entirety of articles from 2008 to 2020 (a total of 8676 articles) were scraped from Mercola’s archive of health-related articles, the articles investigated for this paper include those written and posted in 2020. After eliminating non-text website content, this data set included 793 articles with title, date, url, and the full body text of each page.

The collection of Mercola Facebook archived posts from CrowdTangle include a csv file of 25,027 observations. Variables included number of page likes, number of comments, post likes, other post responses, shared links, videos, number of times videos are viewed, the post text, and a score for overperformance. Most posts included range from mid-2009 to August 2019. As of August 20, 2019, Mercola stopped posting on his Facebook page although posts by Mercola are still widely circulated on Facebook. He announced that he was leaving Facebook and includes a link to his website for more information. On his website, he encourages members to subscribe to his email, SMS messages, and podcasts for daily information. Just recently, in March 2021, Mercola appeared to resume posting on the

Facebook page group. Mercola's Twitter data set is a collection of 1943 tweets ranging from January 1, 2020 to December 31, 2020.

A total of 921 articles were scraped from Children's Health Defense website ranging from 2005 to 2020, however, only 25 articles were dated before 2017. Once again, for this study only the articles written and posted in 2020 were studied (a total of 568 articles). The CrowdTangle Facebook dataset used for this study contained 3959 observations including the same variables as described above in the Mercola dataset. Post dates begin in September of 2016 and continue through 2021. 1630 posts from 2020 were analyzed for this study. 758 tweets were collected and studied using MassMine ranging from January 3, 2020 to December 30, 2020.

Website articles were collected from GreenMedinfo for 2020 and the data resulted in a total of 412 observations. Facebook data collected from CrowdTangle included 23,416 observations with dates ranging from July 2009 through 2020 of which 1246 posts were from 2020 and used for this study. 3174 tweets were collected using MassMine ranging from August 24, 2017 to January 11, 2021. A total of 519 tweets were analyzed from 2020.

4.4 Data Cleaning

A variety of r packages were utilized to clean the resulting data scraped from the three websites as well as Facebook and Twitter text data. Date columns in website, Twitter and Facebook files were parsed using the r package lubridate and a separate column was generated with extracted months for sorting. The text columns were processed to remove numbers, UTF characters, and symbols. The "SMART" stop word list was utilized to remove commonly occurring words in the text column. Additionally, carriage returns, new lines and punctuation were removed, all letters were changed to lowercase, and groups of white space was collapsed to one single space. Finally, a few additional words and some short statements that were included in every article were removed from each data set. Snowballc was then used to stem the text of each dataset and various synonyms used for Covid-19 (Covid-19, SARS-CoV-2, coronavirus etc.) were combined into the term covid.

Twitter data in json files were streamed in using the ndjson package in r and converted to csv files. Columns from the Twitter data included the following information about each post: date (filtered for 2020 tweets), text, user description, total favorites, total followers, total friends, user location, username, screen name, total statuses, tweet mentions, retweet mentions, url, retweet url, and total retweets.

4.5 Data Analysis Methods

Word frequencies and word correlations were created using tidyr to tokenize and count the text material for articles. The scales package was used to find correlations between the website collections. Sentiment analysis was performed utilizing the syuzhet package. For analysis of Twitter co-occurrences, the readtext and quanteda packages were used. Wordcloud was used to visualize the top pages ranked by unique visits in the three websites.

5 Observations of Website Data

5.1 Top Words in Website Articles

In 2020, the words “vaccine”, “health”, and “covid” were the most frequently occurring words in all three collections of websites while CHD had significantly more mentions of vaccines. The words “risk” and “death” also are found more frequently in all three datasets. GreenMedinfo and Mercola both appear to have more terms related to general health conditions while Children’s Health Defense words appear to focus more on civic issues.

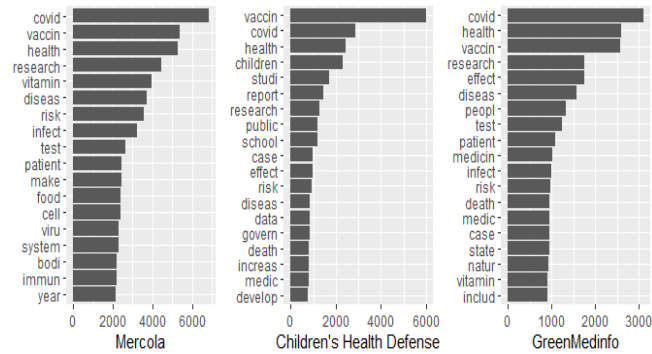


Figure 5.1: Top Website Article Words in 2020

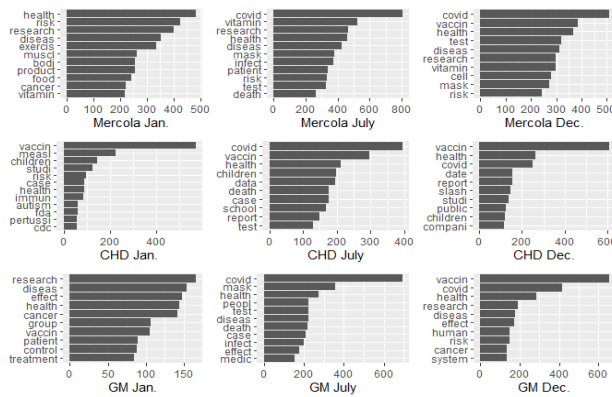


Figure 5.2: Top Words for Website Articles in Jan., July, and Dec. 2020

Throughout 2020 the change in the most frequent words used in the articles posted on the three websites studied shifted notably. While vaccine related words were used predominately previous to the pandemic in January in both CHD and GM’s collection the use of the word vaccine increased significantly in Mercola and GM’s collection with a slight increase in use in the CHD website. By July the word covid was found to be the top used word and by December it was still at least among one of the top three words used by each website. All three websites both have the word mask rated highly in their collections in July or

December. The word health is found consistently among the top words throughout the websites over the three months visualized.

5.2 Word Correlations

Top correlated words give more insight into word usage in the articles found in the three websites. In January, before the pandemic, top words correlated to the words health and vaccine found in the three collections appear to be more neutral in their discussions with more general words frequently mentioned as opposed to specific health conditions or diseases. The once exception to this pattern is found in the Children’s Health Defense website which has focused more heavily on vaccine safety throughout its existence. In this collection the word measles is found and specific references to the CDC and FDA are among top correlated words.

In July there is a notable shift in the correlation between words and their focus in website articles. The vaccine discussion includes references to the pandemic, covid, and even a mention of Bill Gates (gate) and the CDC are included in the discussion. The word risk is now used more commonly found in particular in reference to health in the GreenMedinfo and Mercola articles, and while it previously figured highly in Children's Health Defense articles it now is slightly less correlated. The words death, hospital, virus, and covid are found to be correlated with vaccines in GreenMedInfo and Mercola articles in July while the Children's Health Defense website shows the word death to be correlated to vaccines in January but not in July, although it is found to be correlated with health in July in CHD's articles.

By December, the correlations between pandemic related words become even more specific. Covid is now the most highly correlated word in December in all three websites (with vaccine in Mercola and GM and health with Children's Health Defense). Pfizer is the most correlated word with vaccine in Children's Health Defense December articles followed by the FDA and Moderna. Mask, pandemic and death are all highly correlated with vaccine in the Mercola articles in December while pandemic, companies, and technologies are among the top correlated words in the GreenMedinfo articles suggesting a discussion regarding the involvement of technology companies in the pandemic. Words correlated with the word health now appear to be more specific to the pandemic and words oriented more towards risk and danger.

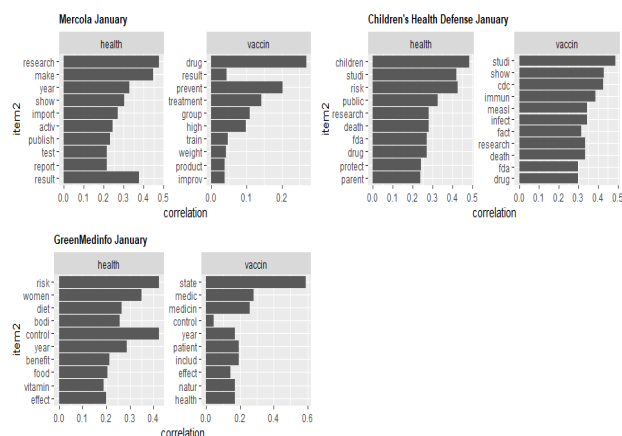


Figure 5.3: Top Correlated Words for January Website Articles

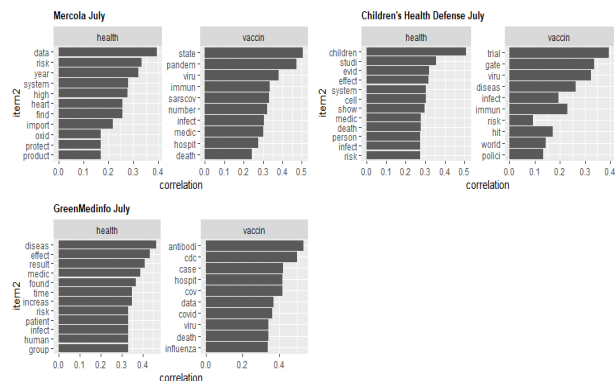


Figure 5.4: Top Correlated Words for July Website Articles

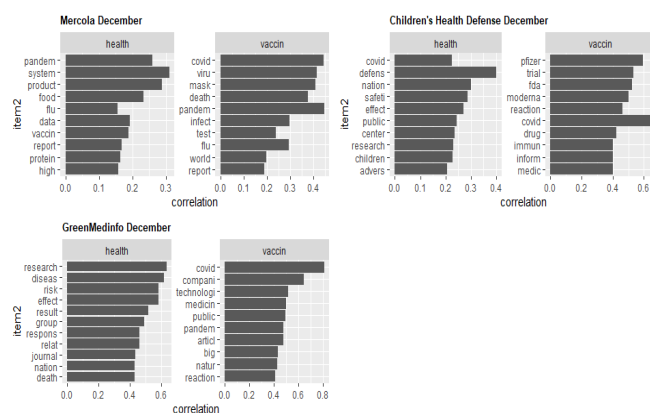


Figure 5.5: Top Correlated Words for December Website Articles

5.3 Comparing the Top 5000 Words Across the Collections for 2020

Collections of words for the three websites tend to be most correlated around covid, vitamin, and research while words like people, risk, test, data and evidence were also among the more correlated words between documents. Words around the center line tend to revolve around mostly health related words like nutrients, and words related to physiology and disease. A Pearson's product-moment calculation shows all three texts to be highly correlated in their words counts, with the lowest correlation between Mercola and GreenMedinfo being 0.7632147 and the highest correlation between Children's Health Defense and GreenMedinfo articles being 0.8956503. The Pearson correlation between Children's Health Defense and GreenMedinfo was 0.8032828.

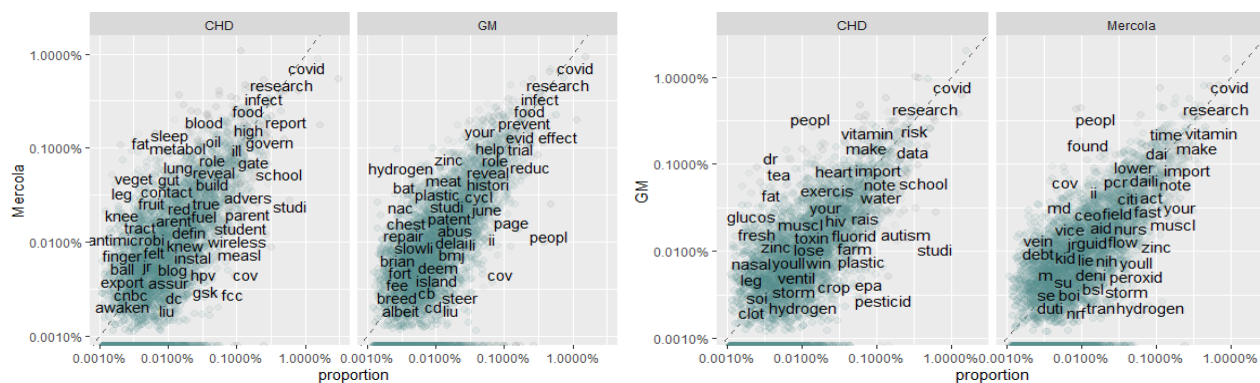


Figure 5.6: Comparing Website Word Correlations

5.4 Sentiment Analysis

Sentiment word clouds help to capture the approach to Covid-19 discussions within the articles collected. The word cloud for Mercola's 4th quarter articles capture a sentiment of disgust for words like vaccine, vaccination, covid and Facebook. The words great and reset in the anger grouping refer to a discussion of a world program that in health website circles is thought of as a conspiracy by technocrats to gain world domination.

In the word cloud for Children's Health Defense, in contrast, the word vaccine is not present. However cov (covid) is present with the disgust category as are mentions of the FDA. The word peg in the disgust category references an ingredient that is used in Covid-19 vaccines that some claim will cause many allergic reactions with the vaccine roll-out. The word pcr refers to the type of lab test used to detect Covid-19 which some misinformation websites believe has been misused to create a "plandemic."

Of interest in the GreenMedinfo word cloud are the reference to the CDC in the fear category as well as the reference to the FDA in the surprise category. While the word lock-down is found in the anticipation category, words related to vaccines are not found.

6.2 Facebook Posts, Likes, Comments and Shares for 2020

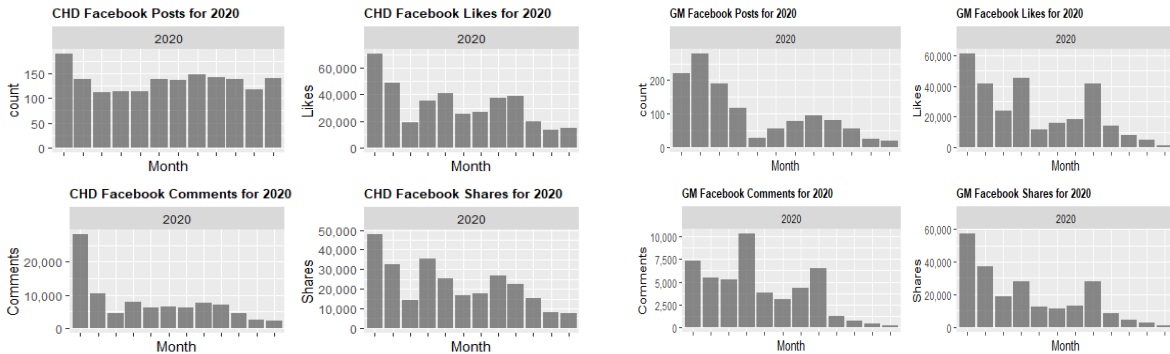


Figure 6.2: Children's Health Defense and GreenMedInfo's Facebook Posts, Likes, Comments, and Shares for 2020

Posts on the Children's Health Defense Facebook page remain consistent throughout the year. There is a notable drop in comments for 2020 after January along with a steady drop in likes and shares throughout 2020. GreenMedinfo's posts drop significantly after the first quarter for 2020 followed by a slight increase over the summer with another drop in November and December. A similar pattern is also noted for posts, likes, comments, and shares with very little activity noted in November and December.

6.3 Visualization of CHD FB Posts: Top Words for 2020

Because Mercola did not have posts in 2020, only posts for Children's Health Defense and GreenMedinfo are investigated. Top words in posts for Children's Health Defense show a lot of discussion around the topic of covid and vaccines with health and children still among the most common words in posts. Interestingly in GreenMedinfo's posts, the words newsletter, join, gift, free, sign, telegram, and subscribe all appear to point to a push on Facebook to use posts to encourage followers to either sign up for GreenMedinfo's newsletter or to join GreenMedinfo on Telegram to avoid Facebook's censorship of GreenMedInfo's posts. A sample post from the GreenMedinfo page in figure 12 illustrates this practice.

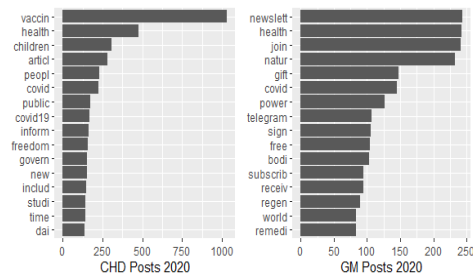


Figure 6.3: Top Words for Facebook Posts in 2020

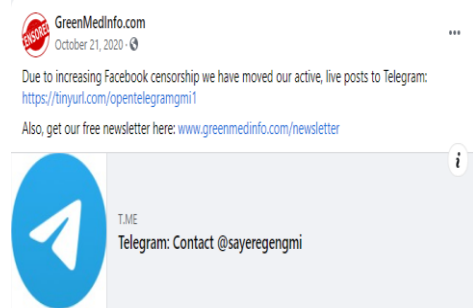


Figure 6.4: GreenMedinfo Facebook Post Example



Figure 6.7: Sample Children's Health Defense Post



Figure 6.8: CHD Facebook Post Encouraging Followers to Move to MeWe

GreenMedinfo's web page demonstrates both the wordiness of their Facebook posts and the regular presence of an invitation to subscribe to their newsletter in order to receive a free gift (a 500-page document). The word telegram is found in the trust category.



Figure 6.9: Sentiment Word Cloud for GreenMedinfo Facebook Posts

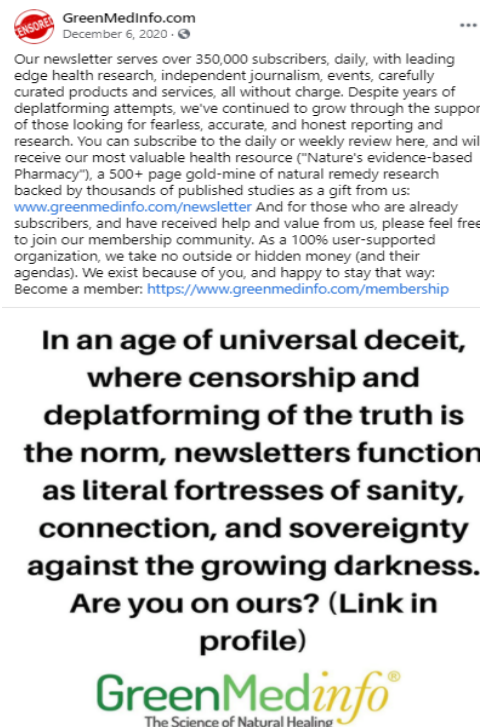


Figure 6.10: Sample GreenMedinfo Facebook Post - Newsletter Invitation

7 Social Media Analysis - Twitter Data

7.1 Tweets by Month for 2020

Mercola's tweets peaked in the beginning of 2020 with roughly 240 tweets or more for the first three months of 2020. In April, his tweets dropped significantly and stayed low during May. In June, tweets peaked once again followed by a drop in tweets for July and August.

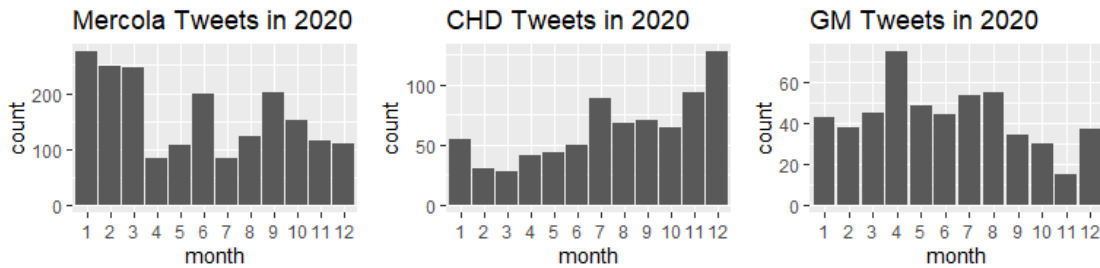


Figure 7.1: Tweet Count by Month for 2020

September tweets neared the level of Mercola's tweets in the first quarter but again dropped the following two months. Tweets in November and December were also relatively low. Mercola's tweets decrease in 2020, particularly after the pandemic begins. However, he is still averaging more than a few tweets a day. Children's Health Defense tweets more than double in 2020 over the year. GreenMedinfo's tweets are the fewest of the three websites with some decreases in tweets towards the end of 2020, particularly in November.

7.2 Top Words for Tweets in 2020

For the top words found in 2020 Mercola tweets, covid is the highest ranked with significantly more mentions than any other word. The phrase "stopcovidcold" is the third most frequent "word." This phrase is a reference to another Mercola website by the same name. This website contains information that exclusively addresses purported ways to avoid

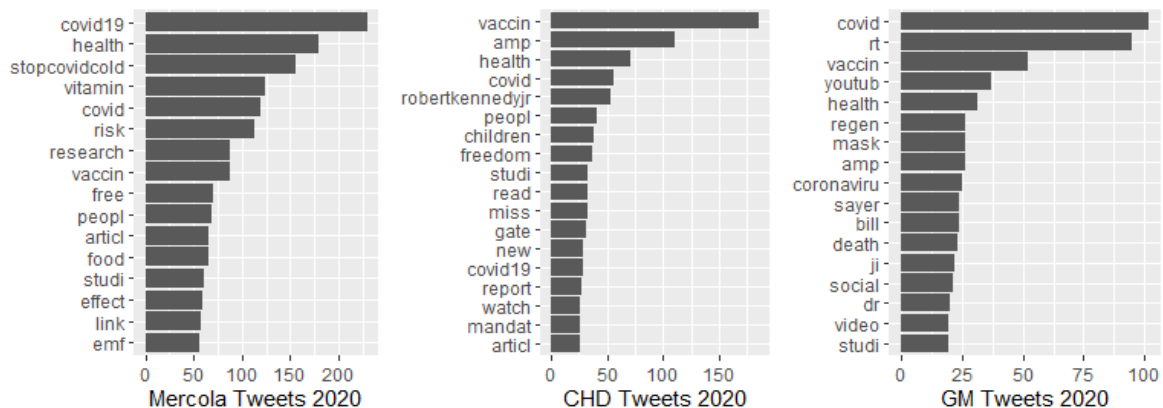


Figure 7.2: Top Words for 2020 Tweets

7.3 Sentiment Analysis for 2020 Tweets

Mercola Tweet Sentiment - 2020

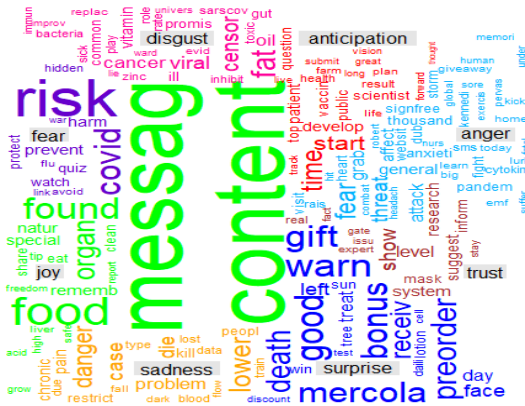


Figure 7.3: Sentiment Analysis for 2020 Mercola Tweets

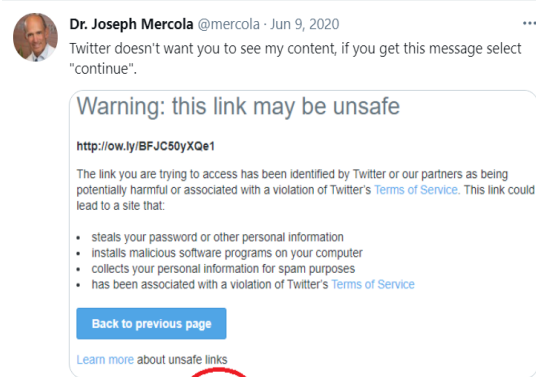


Figure 7.4: Sample Mercola Tweet

Children's Health Defense tweets refer to their followers as "Health Freedom Advocates" and these three words found in the wordcloud refer to frequent mentions in tweets that encourage their followers to take action against the numerous issues that CHD tweets about. Interestingly, Fauci is found in the sadness category. The CDC, vaccines, shot, and Trump are among words mentioned in the surprise category.

GreenMedinfo's tweets appear to have more content revolving around covid with negative mentions of covid-related words. Mask and face are mentioned in fear suggesting the presence of anti-mask messages. Vaccine falls in the surprise category. Covid, censor, test and other virus-related words are found in the disgust category.

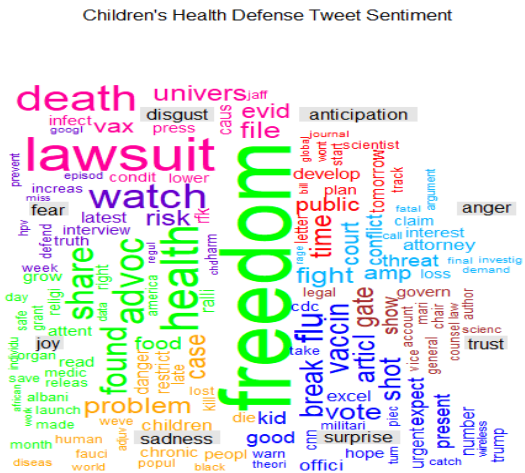


Figure 7.5: Sentiment Word Cloud for Children's Health Defense 2020 Tweets

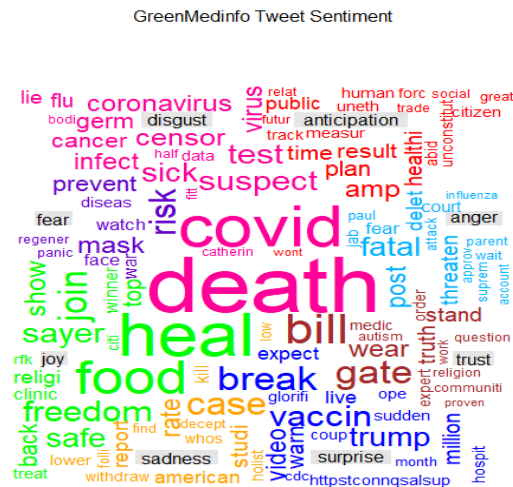


Figure 7.6: Sentiment Word Cloud for GreenMedinfo 2020 Tweets

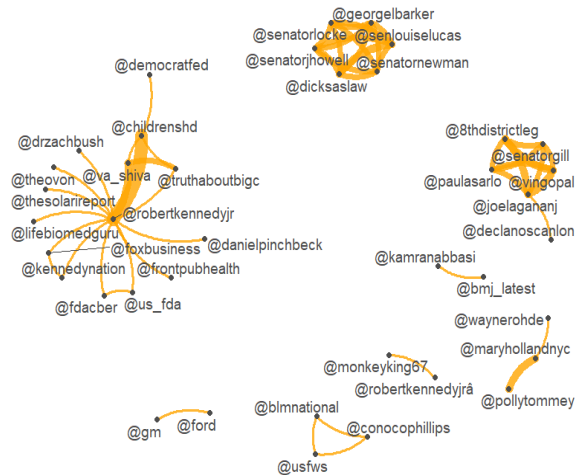


Figure 7.10: User Co-occurrence Graphs for 2020 Children's Health Defense Tweets

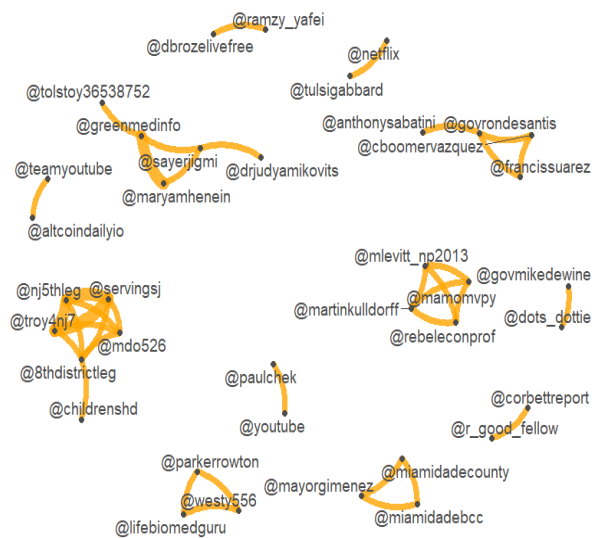


Figure 7.12: User Co-occurrence Graphs for 2020 GreenMedinfo Tweets

GreenMedinfo appears to tweet at a few political figures as well as Children's Health Defense. Dr. Judy Mikovits is a controversial figure in the anti-vaccine movement who helped to produce the Plandemic video. She is frequently referred to in articles by all three websites.

8 Observations of Website Traffic Data

8.1 Traffic Sources

Traffic to Mercola's website appears to have slowed over 2020. While traffic from direct sources appears to be relatively consistent, traffic from searches has decreased steadily. Traffic from referrals and social media appears to remain the same throughout the year.

The Children's Health Defense website experienced significant growth through both direct traffic sources and to a lesser degree sources from social media. Search traffic appears to fluctuate somewhat but does not appear to have changed much overall. Referral traffic has also shown some growth.

Direct traffic to the GreenMedinfo website has grown throughout 2020 with the most growth in April and May. Traffic from social media has decreased somewhat and search traffic is much lower than in the beginning of 2020.

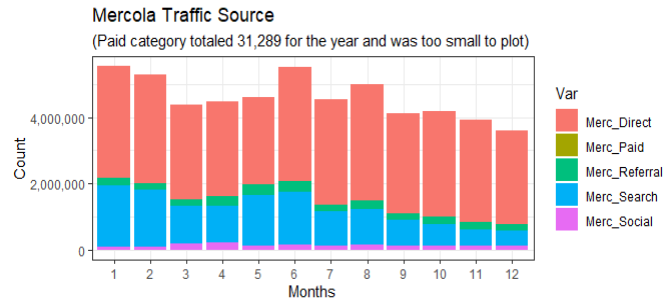


Figure 8.1: Website Traffic Sources for Mercola by Month in 2020

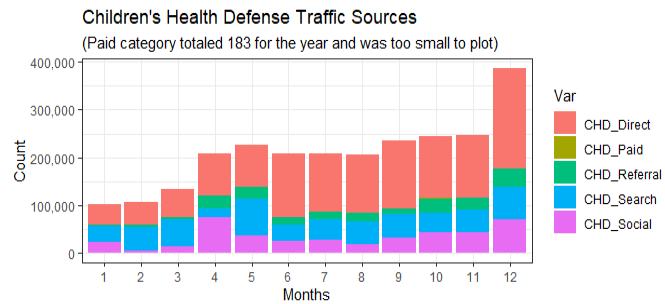


Figure 8.2: Website Traffic Sources for Children's Health Defense by Month in 2020

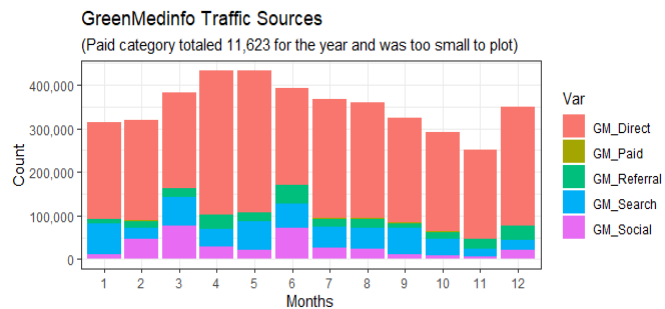


Figure 8.3: Website Traffic Sources for GreenMedinfo by Month in 2020

8.2 Top Pages

The following three visualizations are phrase clouds generated with the titles of the top 25 most visited pages for the three websites in December 2020. These titles reflect the articles that had the most unique visits during the month. For all three articles, there is a mix of both health-related articles along with mostly articles that represent common discussions that include Covid-19 misinformation. This list of top visited pages illustrates exactly where websites visitors who either subscribe to website newsletters or click through a social media post arrive and what they are reading about.

is coronavirus everywhere
pandemic wealth inequality
coronavirus vaccine side effects
covid conspiracy
vitamin d prevent respiratory tract infections
technocracy and the great reset
do asymptomatic people spread coronavirus
anti vaccine propaganda censorship
vitamin d to prevent tumor death
asymptomatic covid testing
covid christmas rules
coronavirus hoax
pufas iron intake and dpn
pro and anti vax
coronavirus lawsuit pcr test reliability
how to improve immune system
thomas lewis eye brain connection
efficacy of surgical masks
covid vaccine cicip
face mask on flight
high dose vitamin c therapy major diseases
cultured meat from human stem cells
aborted fetal cells in coronavirus vaccines
linoleic acid health effects

Figure 8.4: Mercola Top Pages for December 2020

The Mercola top pages list includes articles that appear to imply that the pandemic is a conspiracy and a hoax manufactured with inaccurate testing processes which are being used by “technocrats” to achieve what many conspiracy theorists refer to as “The Great Reset”. Many in this group argue that pandemic prescriptions and regulations such as lockdowns and masking along with vaccines are being used to achieve some sort of compliance from the general public which will help with future population control efforts. The “covid Christmas hoax” phrase illustrates how interest in December turned to encouraging readers to not comply with requests on the part of health officials to refrain from travel and large gatherings in order to mitigate the spread of Covid-19 during the holidays. The following month saw the most Covid-19 cases and deaths recorded throughout the entire pandemic.

retract study flu vaccines protect against covid
cdc flawed science discredit pentagon flu study
save whales dolphins seabirds ditch plastic
pfizer covid vaccine reaction fda peg
pfizer covid vaccine allergic reactions
welcome to the defender
pandemic beneficiaries technocrats
mosquito control pesticides cancer causing pfas chemicals
pfizer covid vaccine trial pathogenic priming
covid 19 vaccine news
johns hopkins plan vaccinate ethnic minorities mentally challenged
tribute brandy vaughn
fda cancer cells in vaccines
unvaccinated children healthier than vaccinated children
fauci fda pfizer moderna covid vaccines
yes bill gates said that
vaccine mandates an erosion of civil rights chd e book
us governments plan track covid vaccine recipients
fight the real deep state
membership after sign up
fda petition halt pfizer vaccine efficacy
police launch investigation death brandy vaughn
mercola childrens health targeted national security risk
toxic vaccine ingredients the devils in the details

Figure 8.5: Children's Health Defense Top Pages for December 2020

The Children’s Health Defense top page list focuses more on articles that attempt to expose purported problems with the Covid-19 vaccines. Vaccine manufacturers are specifically mentioned as are the so-called villains of common Covid-19 conspiracy theories, Fauci and Gates. Children’s Health Defense also points to ways in which the government will use vaccines to gain control over the population, starting with minorities and the mentally challenged. Additionally, Children’s Health Defense most visited articles tend to be ones that point out any possible problem with vaccines including articles about vaccine ingredients and possible adverse health consequences associated with the vaccines. Note the top page that encourages readers to sign up for the regularly emailed newsletter.

GreenMedinfo's list appears to be a mix of Mercola's and Children's Health Defense's most popular articles. Like the Mercola articles, GreenMedinfo articles point to questions about the legitimacy of the pandemic and the activities of technology leaders in using the pandemic for "The Great Reset". Sprinkled throughout are articles about common food related health remedies, along with articles that question the efficacy and safety of the Covid-19 vaccines as well as highlighting the vaccine's "massive side effects".

plan unfolding how vaccines will be monitored
 can vaccines alter child s health trajectory
 covid homeland security biosecurity
 emergency covid vaccines may cause massive side effects
 three spice combo improves chronic knee osteoarthritis
 more reasons black seed remedy everything death
 researchers discover why gum disease causes heart disease
 black seed remedy everything death
 scam has been confirmed pcr does not detect sars cov
 yes bill gates said heres proof
 how clean your arteries one simple fruit
 technocracy and great reset
 when cancer spontaneously disappears
 former pfizer science officer reveals great covid scam
 foods out your risk deadly stroke
 korean superfood may reduce body fat

Figure 8.6: GreenMedinfo Top Pages for December 2020

9 Conclusions

The material explored in this study is in many ways simply the tip of the iceberg when it comes to health misinformation. In fact, the depth and variety of communications found related to the three studied websites were considerably daunting. One mitigating factor, however, was the observation of how connected the major participants in this community are and how common threads are easily elicited from the material. It was interesting to observe that Facebook's efforts to "shadow-ban" content were found to be somewhat effective, however, there is still a great deal of active posting on the social media platform regarding common Covid-19 misinformation, particularly surrounding vaccines, to this day. Additionally, despite some decrease in Facebook activity on the part of some misinformers, traffic simply was directed to websites in other ways.

Several social media companies have had a varied level of success in labeling, discouraging, and sometimes censoring posts created by those responsible for Covid-19 misinformation. Nonetheless, as evidenced by the data collected in this study, even when a social media user finds themselves having lost access to a platform, there are ample means of driving potential followers and readers to their website. Once an email capture is made, that recipient will be "fire-hosed" with health misinformation on a regular basis. Given the prevalence of vaccine hesitancy, adherence to Covid-19 myths, and a significant unwillingness to comply with public health initiatives pertaining to Covid-19 by a large portion of the public, it is likely that these sources of misinformation are meeting their goals of reaching the public, even when they are banished from social media.

Several approaches must be taken to address the problem of medical misinformation in the future. First, vigilance and knowledge of key players in various misinformation communities and a thorough knowledge of commonly held conspiracy theories at any given time are essential for preemptively and responsively sharing information in a way that contradicts and refutes harmful ideas that are shared in a variety of venues. While it may be tempting to fix responsibility for misinformation on its most visible expressions (like social media), when it comes to addressing this problem, it will be important in the future to be aware of the

means by which those who misinform adapt and change their approach in order to evade detection.

It is also notable that not all Covid-19 misinformation and conspiracy theories are equal. Some misinformation appears to get very little traction in the website articles and social media posts, for example, the idea that 5G radiation caused Covid-19 or other misinformation about the origins of Covid-19. Likewise, discredited treatments like HCQ commonly touted by websites that feature misinformation received much press coverage throughout the pandemic and yet did not appear to be a significant contributor to articles that were consumed by readers on these websites. Rather, Covid-19 misinformation topics such as compliance with mitigation efforts by social-distancing, wearing masks, or consenting to take the vaccine appeared to dominate website discussions. Other popular topics such as “The Great Reset” or “The Plandemic” appeared to serve as building blocks which laid a foundation to successfully achieve an acceptance of the misinformation that has led to public non-compliance with mitigation efforts and vaccine hesitancy.

Coupled with the knowledge of how misinformation reaches the general public and which information appears to be the greatest threat to public health efforts, leaders in the medical and scientific community along with those who keep communities informed about public health crises, should be able to discern which specific misinformation creates the greatest danger to the overall public health of communities. Likewise, such information should allow officials to create messages that are fine-tuned to the concerns raised by websites that spread misinformation. These efforts should be able to prevent readers who encounter misinformation on social media or in their inbox to avoid being drawn into false narratives which in turn impact their health choices to the detriment of the health of larger community in which they reside. In this way, website and social analytics can serve as a powerful tool in the effort to create and implement effective public health responses in the midst of major health crises.

10 References

- 132 websites (and counting) are pushing coronavirus conspiracies, says NewsGuard Misinformation Monitor. (2020, March 27). Newsweek. <https://www.newsweek.com/132-websites-are-pushing-coronavirus-conspiracy-theories-says-newsguard-misinformation-monitor-1494557>
- Alexandra S. Levine. (2020, December 21). *Misinformation about the vaccine could be worse than disinformation about the elections*. POLITICO. <https://www.politico.com/news/2020/12/21/social-media-vaccine-misinformation-449770>
- Ball, P., & Maxmen, A. (2020). The epic battle against coronavirus misinformation and conspiracy theories. *Nature*, 581(7809), 371–374. <https://doi.org/10.1038/d41586-020-01452-z>
- Barrett, S. (2020a, July 18). COVID-19 Schemes, Scams, and Misinformation | Quackwatch. <https://quackwatch.org/consumer-protection/covid-19-consumer-protection/>
- Barrett, S. (2020b, July 27). Dr. Joseph Mercola Ordered to Stop Illegal Claims | Quackwatch. <https://quackwatch.org/11ind/mercola/>

Burki, T. (2020). The online anti-vaccine movement in the age of COVID-19. *The Lancet Digital Health*, 2(10), e504–e505. [https://doi.org/10.1016/S2589-7500\(20\)30227-2](https://doi.org/10.1016/S2589-7500(20)30227-2)

Center for Countering Digital Hate. (n.d.). CCDH. Retrieved March 23, 2021, from <https://www.counterhate.com>

Coronavirus Misinformation Tracking Center. (n.d.). *NewsGuard*. Retrieved March 23, 2021, from <https://www.newsguardtech.com/coronavirus-misinformation-tracking-center/>

Cowley, S. (2021, January 18). Vaccine Critics Received More Than \$1 Million in Pandemic Relief Loans. *The New York Times*. <https://www.nytimes.com/2021/01/18/business/anti-vaccine-ppp-loans.html>

Dr. Joseph Mercola—Posts / Facebook. (n.d.). Retrieved March 22, 2021, from <https://www.facebook.com/doctor.health/posts/10157083814730590>

Facebook’s Algorithm: A Major Threat to Public Health. (2020, August 19). Avaaz. https://secure.avaaz.org/campaign/en/facebook_threat_health/

Forget Facebook – Take Control of Your Privacy by Dr. Mercola. (n.d.). Mercola.Com. Retrieved March 22, 2021, from <http://www.mercola.com/forget-facebook.htm>

Hollowood, E., Sorrato, R., & Newell, C. (n.d.). Who are the ‘super-spreaders’ of medical misinformation, and how do their hoaxes spread so quickly on social media? A graphic guide. Retrieved March 22, 2021, from <https://www.theglobeandmail.com/world/article-medical-misinformation-superspreader-investigation/>

Iyengar, R. (2021, February 11). Robert F. Kennedy Jr. Has been banned from Instagram—CNN. CNN. <https://www.cnn.com/2021/02/10/tech/robert-kennedy-jr-instagram-ban/index.html>

Jamison, A. M., Broniatowski, D. A., Dredze, M., Wood-Doughty, Z., Khan, D., & Quinn, S. C. (2020). Vaccine-related advertising in the Facebook Ad Archive. *Vaccine*, 38(3), 512–520. <http://dx.doi.org/10.1016/j.vaccine.2019.10.066>

Jarry, J. (2019, July 11). *Popular Health Guru Sayer Ji Curates the Scientific Literature with His Bachelor’s Degree in Philosophy*. Office for Science and Society. <https://www.mcgill.ca/oss/article/pseudoscience/popular-health-guru-sayer-ji-curates-scientific-literature-his-bachelors-degree-philosophy>

Ji, S. (2020a, July 8). *Instagram Threatens to Delete Greenmedinfo’s Account for Posting THESE Covid-19 Related Posts*. <https://www.greenmedinfo.com/blog/instagram-threatens-delete-greenmedinfos-account-posting-covid-19-related-image2>

Ji, S. (2020b, September 2). *Greenmedinfo Deleted from Instagram and LinkedIn (Facebook Next?) Has The Great Social Media Purge of 2020 Begun?* <https://www.greenmedinfo.com/blog/greenmedinfo-deleted-instagram-and-linkedin-facebook-next-has-great-social-media->

Ji, S. (2021, March 7). *Twitter Deletes GreenMedInfo’s 13-Year-Old Account After Calling Out CCDH’s Digital Hit List*. <https://www.greenmedinfo.com/blog/twitter-deletes-greenmedinfos-13-year-old-account-after-calling-out-ccdhs-digital2>

Mercola, J., & Cummings, R. (2021). *The Truth About Covid-19: Exposing the Great Reset, Lockdowns, Vaccine Passports, and the New Normal*. Chelsea Green Publishing.

Mercola.com—#1 Natural Health Website. (n.d.). Mercola.Com. Retrieved March 22, 2021, from <https://www.mercola.com/>

Mokhtari, H., & Mirzaei, A. (2020). The tsunami of misinformation on COVID-19 challenged the health information literacy of the general public and the readability of educational material: A commentary. *Public Health*, 187, 109–110. <https://doi.org/10.1016/j.puhe.2020.08.011>

Mostrous, A. (2020, November 11). *How a Kennedy became a 'superspreader' of hoaxes on COVID-19, vaccines, 5G and more*. <https://www.theglobeandmail.com/world/article-robert-f-kennedy-jr-medical-misinformation/>

National Vaccine Information Center—Your Health. Your Family. Your Choice. (n.d.). National Vaccine Information Center (NVIC). Retrieved March 22, 2021, from <https://www.nvic.org/>

NewsGuard childrenshealthdefense.org Nutrition Label. (n.d.). NewsGuard. Retrieved March 22, 2021, from <https://api.newsguardtech.com/label/childrenshealthdefense.org?cid=1a184c47-f65b-403a-af0b-824b5b5b9bbe>

NewsGuard greenmedinfo.com Nutrition Label. (n.d.). Retrieved March 23, 2021, from <https://api.newsguardtech.com/label/greenmedinfo.com?cid=1a184c47-f65b-403a-af0b-824b5b5b9bbe>

NewsGuard mercola.com Nutrition Label. (n.d.). NewsGuard. Retrieved March 22, 2021, from <https://api.newsguardtech.com/label/mercola.com?cid=1a184c47-f65b-403a-af0b-824b5b5b9bbe>

Nvic-fb-deplatforming-statement-2021.pdf. (n.d.). Retrieved March 22, 2021, from <https://www.nvic.org/cmstemplates/nvic/Pdf/pressrelease/nvic-fb-deplatforming-statement-2021.pdf>

Satija, N., & Sun, L. H. (n.d.). A major funder of the anti-vaccine movement has made millions selling natural health products. *Washington Post*. Retrieved April 10, 2021, from https://www.washingtonpost.com/investigations/2019/10/15/fdc01078-c29c-11e9-b5e4-54aa56d5b7ce_story.html

Sun, L. H. (2019, November 15). *Majority of anti-vaccine ads on Facebook were funded by two groups*. The Washington Post (Online); WP Company LLC d/b/a The Washington Post. <http://search.proquest.com/docview/2314673862/citation/DCAC66FC89044F93PQ/1>

Super-spreaders. (2020, August 4). Tortoise. <https://www.tortoisemedia.com/2020/08/04/rise-of-the-super-spreaders-infodemic/>

The Disinformation Dozen | Center for Countering Digital Hate. (n.d.). CCDH. Retrieved March 30, 2021, from <https://www.counterhate.com/disinformationdozen>

The Truth About COVID-19 by Joseph Mercola | Chelsea Green Publishing. (n.d.). Retrieved March 22, 2021, from <https://www.chelseagreen.com/product/the-truth-about-covid-19/>

Thompson, A., & Finnegan, G. (2020). United Nations Children's Fund Vaccine Misinformation Management Field Guide. <https://vaccinemisinformation.guide>

Tran, L. (2019, November 7). *Firehosing: The systemic strategy that anti-vaxxers are using to spread misinformation* / Lucky Tran. The Guardian.
<http://www.theguardian.com/commentisfree/2019/nov/07/firehosing-the-systemic-strategy-that-anti-vaxxers-are-using-to-spread-misinformation>