

Medical fake news in polish on the World Wide Web during the COVID-19 pandemic period: the spread dynamics analysis

Małgorzata Chlabicz, Aleksandra Nabożny, Jolanta Koszelew, Wojciech Łaguna, Paweł Sowa, Wojciech Budny, Katarzyna Guziejko, Magdalena Róg-Makal, Sławomir Pancewicz, Maciej Kondrusik, Piotr Czupryna, Beata Cudowska, Dariusz Lebensztejn, Anna M. Moniuszko-Malinowska, Karol A. Kamiński, Anna Szpakowicz

Submitted to: Journal of Medical Internet Research
on: April 14, 2023

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	28
Figures	29
Figure 1.....	30
Figure 2.....	31
Figure 3.....	32
Multimedia Appendixes	33
Multimedia Appendix 1.....	34
Multimedia Appendix 2.....	34
Related publication(s) - for reviewers eyes onlies	35
Related publication(s) - for reviewers eyes only 0.....	35

Medical fake news in polish on the World Wide Web during the COVID-19 pandemic period: the spread dynamics analysis

Ma?gorzata Chlabicz^{1, 2} MD, PhD; Aleksandra Nabo?ny³ MSc; Jolanta Koszelew⁴ PhD; Wojciech ?aguna⁵ MSc; Pawe? Sowa¹ MD, PhD; Wojciech Budny⁶ MD, PhD; Katarzyna Guziejko⁷ MD, PhD; Magdalena R?g-Makal² MD, PhD; S?awomir Pancewicz⁸ MD, PhD, Prof Dr; Maciej Kondrusik⁸ MD, PhD; Piotr Czupryna⁸ MD, PhD; Beata Cudowska⁹ MD, PhD; Dariusz Lebensztejn⁹ MD, PhD, Prof Dr; Anna M. Moniuszko-Malinowska⁸ MD, PhD, Prof Dr; Karol A. Kami?ski^{1, 10} MD, PhD, Prof Dr; Anna Szpakowicz¹⁰ MD, PhD

¹Department of Population Medicine and Lifestyle Diseases Prevention Medical University of Bia?ystok Bia?ystok PL

²Department of Invasive Cardiology Medical University of Bia?ystok Bia?ystok PL

³Department of Software Engineering Gda?sk University of Technology Gda?sk PL

⁴R&D Department Science4People sp. z o.o. Szczecin PL

⁵Faculty of Computer Science Bialystok University of Technology Bia?ystok PL

⁶Department of Allergology and Internal Medicine Medical University of Bia?ystok Bia?ystok PL

⁷2nd Department of Lung Diseases and Tuberculosis Medical University of Bia?ystok Bia?ystok PL

⁸Department of Infectious Diseases and Neuroinfection Medical University of Bia?ystok Bia?ystok PL

⁹Department of Pediatrics, Gastroenterology, Hepatology, Nutrition, Allergology and Pulmonology Medical University of Bialystok Bia?ystok PL

¹⁰Department of Cardiology Medical University of Bia?ystok Bia?ystok PL

Corresponding Author:

Aleksandra Nabo?ny MSc

Department of Software Engineering

Gda?sk University of Technology

ul. Gabriela Narutowicza 11/12

Gda?sk

PL

Abstract

Background: The phenomenon of the rapid generation and spread of misinformation about the novel coronavirus COVID-19 during the pandemic was evident and attracted many researchers. However, there are more health-related topics polluting the Internet with misinformation that are not as widely studied.

Objective: The aim of our experiment was to assess the reach of the most popular medical content, not limited to the subject of the pandemic, in the World Wide Web (WWW). We evaluated the content for subject matter and credibility in 2021 and 2022. The content was evaluated according to the principles of Evidence Based Medicine (EBM) by experienced clinicians.

Methods: We used 274 keywords to perform web page searches using the BuzzSumo Enterprise Application. The keywords were selected according to medical topics which were obtained via questionnaires given to the medical practitioners. The search was limited to two specific date ranges: (1) Jan. 1, 2021 to Dec. 31, 2021; (2) Jan. 1, 2022 to Dec. 31, 2022. We limited searches to web pages in Polish language and filtered them by the selected dates. The analysis included 161 web pages retrieved from 2021 and 105 retrieved from 2022. Each web page was reviewed by an experienced doctor in terms of credibility. The evaluation was performed in accordance with EBM standards. Additionally, we collected data about social media engagements related to the web pages.

Results: The proportion of non-credible contents about COVID-19 significantly decreased in 2022 as compared to 2021. The percentage of non-credible web pages with the topics COVID-19 and general vaccinations changed from 57% to 24% and 42% to 30%, respectively. However, simultaneously there was a significant rise in non-credible contents and their spread in social media regarding other medical topics. The percentage of non-credible web pages with the topic cholesterol and statins and cardiology increased from 17% to 26% and from 18% to 26%, respectively.

Conclusions: Actions taken during the COVID-19 pandemic to prevent the spread of fake news appear to have been effective.

However, our analysis indicates that such interventions should be applied continuously to existing and emergent medical topics. It seems that topics other than the virus were “filling up the vacuum” created by the decreasing interest in the pandemic.

(JMIR Preprints 14/04/2023:48130)

DOI: <https://doi.org/10.2196/preprints.48130>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

Original Manuscript

Medical fake news in polish on the World Wide Web during the COVID-19 pandemic period: the spread dynamics analysis

Małgorzata Chlabicz^{1,2*}, Aleksandra Nabożny³, Jolanta Koszelew⁴, Wojciech Łaguna⁵, Anna Szpakowicz⁶, Paweł Sowa¹, Wojciech Budny⁷, Katarzyna Guziejko⁸, Magdalena Róg-Makal², Sławomir Pancewicz⁹, Maciej Kondrusik⁹, Piotr Czupryna⁹, Beata Cudowska¹⁰, Dariusz Lebensztejn¹⁰, Anna M. Moniuszko-Malinowska⁹, Karol A. Kamiński^{1,6}

¹Department of Population Medicine and Lifestyle Diseases Prevention, Medical University of Białystok, 15-269 Białystok, Poland.

² Department of Invasive Cardiology, Medical University of Białystok, 15-276 Białystok, Poland.

³Department of Software Engineering, Gdańsk University of Technology, Gdańsk, Poland.

⁴R&D Department, Science4People sp. z o.o., ul. Warszawska 6/32, Poland.

⁵Faculty of Computer Science, Białystok University of Technology, 15-351 Białystok, Poland.

⁶Department of Cardiology, Medical University of Białystok, 15-276 Białystok, Poland.

⁷Department of Allergology and Internal Medicine, Medical University of Białystok, Białystok, Poland.

⁸2nd Department of Lung Diseases and Tuberculosis, Medical University of Białystok, Żurawia 14, 15-540, Białystok, Poland

⁹Department of Infectious Diseases and Neuroinfection, Medical University of Białystok, 15-540 Białystok, Poland.

¹⁰Department of Pediatrics, Gastroenterology, Hepatology, Nutrition, Allergology and Pulmonology, Medical University of Białystok, Poland

*Correspondence: mchlabicz@op.pl; Tel.: +48-50699663

The authors declare no conflict of interest.

Introduction

Background

Although fake news was already known in the early days of the printing press in the previous centuries, it has begun to play an increasingly important role in the era of the Internet. A technology historian Melvin Kranzberg said: "Technology is neither good nor bad; nor is it neutral." [1]. The same could be said of the Internet. Free access to information via the World Wide Web (WWW) has many advantages but also leads to faster propagation of non-credible or fake stories. Disinformation harms society in various aspects of life, but fake news concerning Online Health Information (OHI) can particularly affect people's well-being and have long-term consequences for their health [2]. In this study, we aimed to detect medical-related topics that generate much unreliable content prevalent on the Web (e.g. COVID-19 pandemic, vaccination, statins, diet), and observe their dynamics over two years (2021-2022).

Years 2020, 2021 and 2022 were dominated by the Sars-cov-2 pandemic in the medical-related OHI. It is therefore not surprising that research on the propagation of medical disinformation from this period focused mainly on topics related to COVID-19. Moreover, the research was highly focused on the perspective of the lay person, not medical practitioners who had to face the consequences of disinformation spread. Our study aimed to face two challenges: to perceive OHI from a broader view than just the Pandemic and to explore non-credible OHI from the perspective of medical practitioners.

The starting point was the surveys collected among doctors, indicating the topics encountered in their clinical practice. As a result, we have collected information about false content that had a real negative impact on attitude to medical recommendations among patients. In addition, we collected data on disinformation about the covid pandemic and six other medical topics, which provided an image of the overall dynamics of medical online disinformation spread. We systematically searched for the most popular articles and social media engagements related to those articles in 2021 and 2022. Then their credibility was assessed by an experienced team of clinicians, in accordance with the principles of Evidence-Based Medicine (EBM). This allowed us to identify important areas for

improvement in interventions to stop the spread of non-credible OHI.

Literature review

Although there exist several definitions of information credibility [3], as well as non-credible information (such as disinformation or misinformation) [4] there is still no clear definition of false information. Alternatively, the term *medical fake news* is used in the literature [5]. It was defined as "news that is intentionally false and could mislead readers" and "information discrepant with medical knowledge" [6]. The aim is to manipulate and mold public opinion. Many publications report the unfavorable runoff of fake news spreading on the Internet to the community [7,8]. One is the Betsch team's study, which found that accessing vaccine-critical websites for only five to ten minutes increased the perception of vaccine risk and resulted in negative perceptions of the risk of vaccinations and a decreasing intention to vaccinate [9].

Evidence Based Medicine (EBM) provides factual information that is applicable to everyday medical practice, as well as for experts working to debunk medical fake news. The term was first used in 1991 [10]. EBM is "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" [11]. EBM aims to combine the clinician's experience, the patient's values, and the best available scientific information to guide clinical management decisions. The principles of EBM apply not only to individual medical practice but also to institutions and the health care system.

The acute respiratory syndrome coronavirus 2 (SARS-CoV-2) induced the novel coronavirus disease of 2019 (COVID-19). It was not only responsible for the pandemic of COVID-19 but also a trigger for wide-scale disinformation campaigns [12,13,14]. The internet was flooded with millions of posts about COVID-19 [15]. The challenges of establishing proper diagnostics tests, therapeutic protocols, and methods of preventing SARS-CoV-2 infections have made the area of the pandemic an ideal space for manipulation and disinformation, especially in decision-making in the health field [13,14,15]. Fake information about the pandemic spread through mass media, social media, and text messaging applications causing such effects as:

- reducing patients' willingness to vaccinate
- obstructing measures to contain disease outbreaks
- instigating the physical interruption of access to health care, among others [16].

Fake news about COVID-19 has fueled disinformation in related areas, such as vaccination, putting the science to the test [17,18]. Although the anti-vaccination movement

spread conspiracy theories even before, e.g. the idea that large pharmaceutical companies and other interested groups exaggerate the benefits of vaccination and hide the risks or dangers of vaccines [19], the pandemic has increased the phenomenon of questioning immunization. Vaccine hesitancy was already observed in the case of well-known vaccinations like MMR [20], and flu vaccine [21], and studies showed that attitudes and conspiracy beliefs (which could be seen as a particular set of behavior) are related to willingness to vaccinate. Dubé et al.[22] showed that parents who vaccinated their children were most likely to be informed about vaccines by a healthcare professional and were less likely to seek information on the Internet. Moreover, Sommariva et al. revealed that disinformation on social networks could hinder disease prevention efforts [23]. As far as COVID-19 vaccines are concerned, some new forms of conspiracy theories have emerged, like microchip injection, non-checked adverse effects, lack of safety assessment before distributing the vaccine to people, and non-proven effectiveness of the vaccine. It has been observed that exposure to fake news lowered the intention to get the COVID-19 vaccine [24]. Even short contact with false information can modify or change human behavior [22]. Consequently, conspiracy beliefs, trust in information, and news online that people are exposed to for a short time could affect their vaccination choices [24]. The exposure to online disinformation about COVID-19 vaccines declined vaccination intent significantly in both the U.S. (6.4% points) and U.K. (6.2% points) samples, according to Loomba et al. in a randomized controlled trial [25].

The problem of fake news in cardiology, and consecutively with statins, has already been described [26, 27] It was superseded during the pandemic of COVID-19 and is now coming back. Waszak et al. [28] analyzed the topics that attracted public attention between 2012 and 2017. The main topics concerned cancer, vaccination, heart attack, stroke, AIDS/HIV, hypertension and diabetes. The topic most tainted by fake news was vaccination (90%), followed by hypertension and HIV/AIDS (both at 70%).

Medical information involving nutrition-related issues and qualified as false or misleading has been analyzed in the literature [7, 29]. While the impact of dis- and misinformation on the level of knowledge and nutritional decision-making is being discussed regularly, the extent of such content on the Internet is relatively rarely analyzed [30]. An attempt to quantify the phenomenon of nutrition disinformation during the pandemic period has been made by several centers - a Turkish study estimated that about 20% of analyzed YouTube videos on nutrition topics contained fake news [31], a Korean study as much as

37% [32]. From a qualitative point of view, the nutrition misinformation in COVID-19 pandemic era included issues concerning food immunity boosters, which were supposedly designed to minimize the risk of infection or severity of infection. Studies described the effect of supplementation with zinc, vitamin C, vitamin A, herbs, or nutraceuticals on minimizing the risk mentioned above [33, 34]. An Indian study found that immunity-boosting food was reflected in dietary decisions in over 70% of respondents [8]. The pandemic also caused an increase in the popularity of fake news related to the protective effects of alcohol in the context of Sars-Cov2 infection - these included the hypothesis that strong alcohol kills the virus. There is a continuous flow of articles stating that light alcohol (beer, wine) stimulates the immune system to fight the coronavirus [7, [35, 36]. This misinformation proved particularly egregious for the 60 people who lost sight and the nearly 6,000 hospitalized after 'preventive' methanol drinking [7, 36]. Although there is a study from Hong Kong indicating that approximately 19% of respondents experienced content describing a beneficial association between drinking alcohol and the prevention of sars-cov2 infection [37], few scientific reports estimate the extent of fake news about this topic. Please note that for the topic of dietary interventions we use the term "misinformation" rather than "disinformation". It is because some of the "off the shelf" recommendations may have been in the form of inept self-help. Therefore, such information does not fall into the category of intentionally misleading (disinformation).

In conclusion, we have many reports on the pernicious impact of medical dis- and misinformation on social security, but reports on the dynamics of the spread of unreliable OHI are still lacking.

Methods

Preliminary questionnaires

The first step towards data collection was to select the areas within which the topics most likely to generate unreliable content would be identified. Initially, the myths and false beliefs that are most often propagated by patients in doctors' offices and hospital wards were identified. We consider those beliefs to realistically hinder doctors from achieving their goal of successfully treating patients. For the purpose of identification, a questionnaire was developed and applied [Appendix 1]. The questionnaires were collected in the Medical University of Bialystok (MUB) Clinical Hospital, Poland, as well as through an online survey. The online survey was conducted thanks to the institutions with which the MUB has signed letters of intent (College of Family Physicians, Polish Society of Cardiology). The

questionnaires were collected from November 29, 2021, to December 26, 2021 (Supplement. Table 1). The questionnaire included questions about the physician's age, gender, specialization, length of service in the profession, place of work, and included an open question: "What false medical information do you encounter in your professional practice when talking to patients?". Responses were obtained from 121 physicians. The mean age of the doctors was 41.1 ± 11.4 years. Women accounted for 64.5% (n=78). Duration of employment in the profession was 15.2 ± 11.4 years. The specialties the doctors had were internal medicine (n=41), cardiology (n=34), pediatrics (n=21), family medicine (n=18), surgery (n=8), hematology (n=6), allergology (n=6), ophthalmology (n=4), rheumatology (n=1), pulmonology (n=1), emergency medicine (n=1), and doctors in training (n=3). Some physicians had several medical specialties. Most of the doctors worked only in hospitals (n=70), some in hospitals, public and private clinics (n=8), in hospitals and public clinics (n=8), or in hospitals and private clinics (n=19). Moreover, some respondents worked only in a public clinic (n=11) or in a private clinic (n=5). The physicians described all the topics they had encountered in their professional practice and/or quoted specific questions that patients had asked them. Based on the questionnaires, we identified common problems and then divided them into nine groups (topics) that were consistent in scope and were related to false medical information.

Web page and social media engagements retrieval using BuzzSumo

Next, queries were formulated based on topics (problem groups). Then, topics were assigned to each member of the other team of physicians, who then queried the DuckDuckGo search engine. Queries were single words or several words pouching a particular meaningful whole, i.e., phrase. Physicians took into account only queries that yielded unreliable websites. 274 queries were collected using this method.

Then, the BuzzSumo Enterprise Application (BuzzSumo Limited, Brighton, United Kingdom) available through the website was used [38]. BuzzSumo is a social media analytics tool that searches the Internet for web documents based on queries and generates engagement reports on their comments, likes and shares on social media. BuzzSumo gathers data across several platforms, namely Facebook, Pinterest, Reddit, and Twitter, to generate a list of article web pages with the highest online engagement. *Engagement* is defined as the total number of interactions that users have with a particular article link, including actions such as "liking," "sharing," and "commenting". After entering a

query, the application returns statistics for the linked pages. The subsequent steps of the web search are shown in Figure 1.

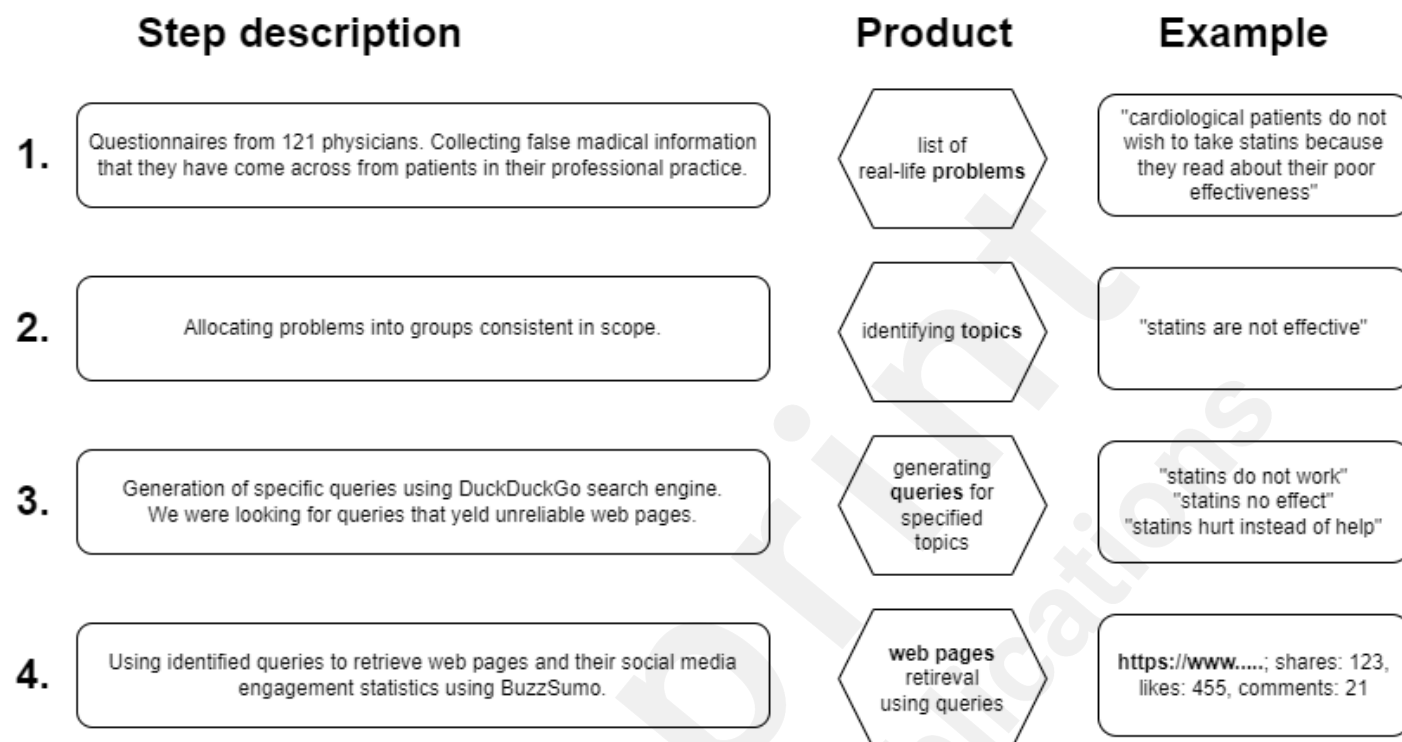


Figure 1. Subsequent stages of web search research.

The search was limited to two specific dates: (1) Jan. 1, 2021, to Dec. 31, 2021; (2) Jan. 1, 2022, to Dec. 31, 2022. The searches were filtered by time and Polish language. Our analysis took place in two consecutive phases. The first phase covered the whole year 2021 and took place in January 2022. The second phase took place in January 2023 and covered the entire 2022. The analysis carried out in this way allowed us to assess the current websites, because the online environment is constantly changing and web pages are often deactivated or stop generating traffic. The latter statement is also confirmed by our work. Despite using exactly the same queries to search websites in both periods, no website from 2021 was selected again in 2022.

Only pages that generated at least ten social media engagement and only pages with text were included in the analysis; videos were not evaluated. Then, ten most popular pages per query were analyzed. "Popular" in this context means those with the highest number of total engagements. Each web page was reviewed by an experienced doctor

(with at least five years of professional work) who is also a scientist at the MUB. In order to clarify the web pages examined, categories were introduced according to their content: medical, non-medical, and difficult to evaluate. Out of 274 queries, only 58 of them resulted in medical content retrieval in 2021, and 32 in 2022. Our workflow is shown in Figure 2.

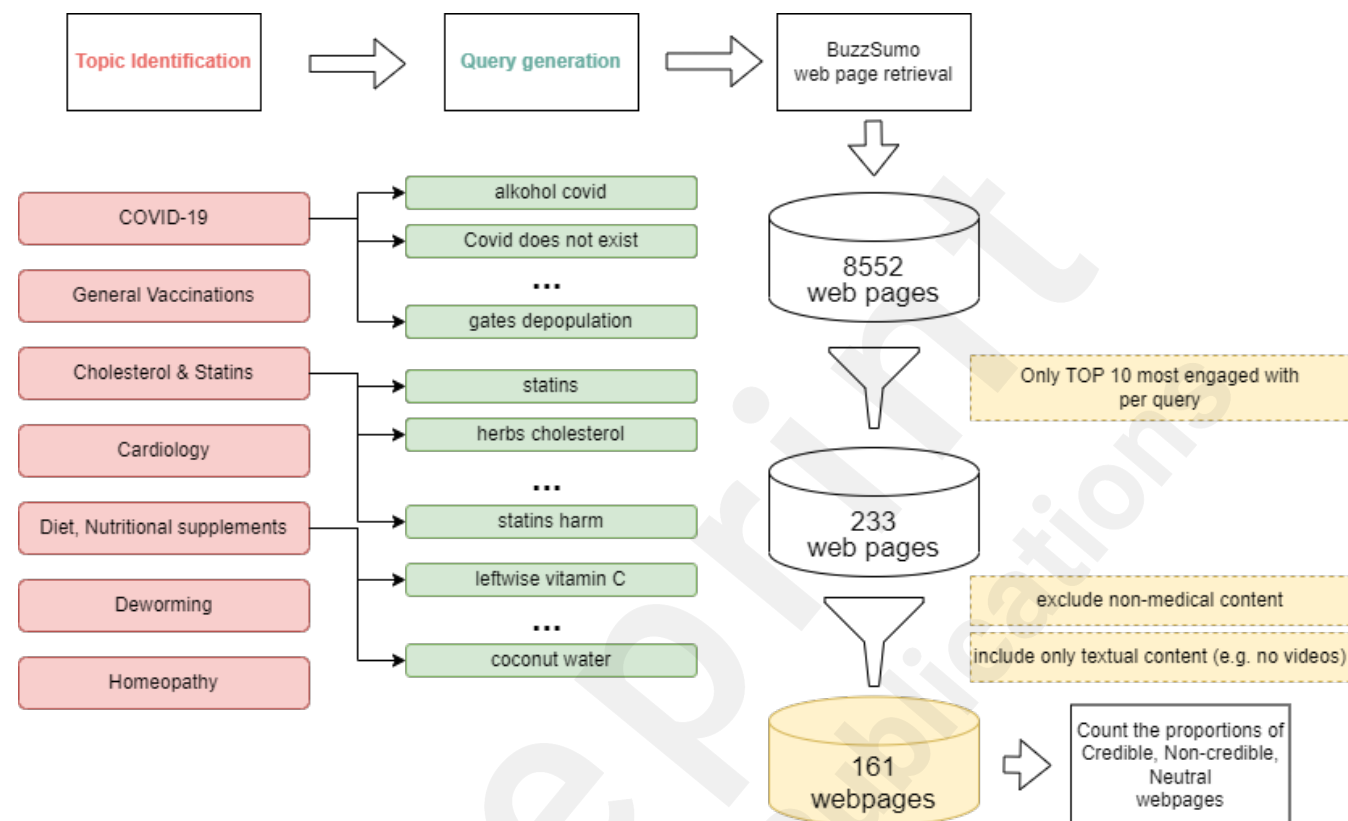


Figure 2. Subsequent stages of the research. Example for the year 2021.

For web pages labeled as medical, the content was assessed in terms of credibility: credible, non-credible, or neutral. Specialists in the relevant medical field evaluated the websites corresponding to the field, e.g. topics on statins cardiologists, topics on COVID-19 infectious disease specialists. The credibility of the health information in the news was assessed by comparing it with current textbooks, guidelines, and scientific articles in the relevant field. The evaluation was performed following EBM standards. The evaluation criteria are shown in Table 1.

Table 1. The description of three content categories that were used to tag web pages. The expert medical practitioners tagged web pages as credible, non-credible or neutral.

Content categories	
Credible	Contains verifiable and up-to-date medical information, does not
medical	contain information intended to mislead the viewer/ reader.

websites	
Non-credible medical websites	1. Contains false medical information. 2. Contains unverified medical information. 3. Contains outdated medical information that is inconsistent with current standards. 4. Contains medical information consistent with EBM, but the implications of the fact presented are falsely exaggerated. 6. Contains partially not credible medical information: miscalculations or erroneous conclusions from true analyses; a mixture of credible and not credible information. 7. Contains non-credible medical information/inconsistent with EMB, but part of the text softens the implication of the sentence.
Neutral medical websites	It contains information related to medicine, but describes a story related to a disease, reports on a medical encounter, and presents regulations in the health care sphere.
Non-medical website	It does not contain information related to medicine.
EBM, Evidence Based Medicine	

The data was collected and analyzed in Microsoft Excel (Microsoft, Redmond, Washington, USA). In this scientific study, descriptive statistics were computed for the categorical variables, and the frequencies of the variables under investigation were presented. The chi-square test was used to assess the distribution of medical web pages categorized as credible, non-credible, or neutral between the years 2021 and 2022. This test was employed to determine if there was a statistically significant difference in the distribution of both the number and popularity of these web pages across the mentioned categories. The intention was to examine if there were shifts in the credibility classification of web pages and their corresponding popularity over the two years. The Fisher exact test was performed in cases where expected values were less than five. However, the least represented data, the neutral medical web pages' popularity for 2021 and 2022, must not be included since the Fisher exact test consumed significant computer power.

Results

Preliminary questionnaires

The most common topics indicated by doctors were: COVID-19 infection, COVID-19

vaccination, vaccination against other infectious diseases, hypercholesterolemia and statins, hypertension, nutritional supplements, and antibiotics. A list of topics is shown in Table 2.

Table 2. A list of topics (problem groups) indicated by doctors in questionnaires, the number of times the topic was mentioned, and the resulting number of queries generated from this topic for data collection purposes.

No	Topics	Number of times a topic was indicated by doctors	Queries
1	Coronavirus disease 2019 (COVID-19)	74	85
2	General vaccinations	64	18
3	Cholesterol and statins	52	45
4	Cardiology	27	68
5	Diet, Nutritional supplements, Alcohol	26	32
6	Antibiotics	12	12
7	Deworming	4	9
8	Lyme disease	2	1
9	Homeopathy	1	4

Web page and social media engagements retrieval using BuzzSumo

Overall, 8552 web pages were obtained in 2021 and 6404 in 2022. COVID-19 topics were the most popular in 2021, while cholesterol and statins were the most popular in 2022. Detailed data by topic are presented in Table 3.

Table 3. The number of all web pages based on the queries, retrieved by BuzzSumo (2021-2022).

Topics	The number of all web pages retrieved by BuzzSumo in 2021	The number of all web pages retrieved by BuzzSumo in 2022
Coronavirus disease	5448	1495

2019 (COVID-19)		
General vaccinations	468	59
Cholesterol and statins	1152	3282
Cardiology	1326	1456
Diet, Nutritional supplements, Alcohol	150	23
Deworming	3	8
Homeopathy	8	80
Total	8552	6404

After the steps described in Figure 1, ultimately, 161 web pages from 2021 were analyzed, and 105 from 2022. The filtering process is depicted in Table 4.

Table 4. The number of web pages retrieved from 2021 and 2022, respectively, during each stage of web search and filtration. In the first stage web pages were retrieved by BuzzSumo using queries specified according to the questionnaires conducted among doctors. In the second stage only web pages that provoked at least 10 interactions were selected and at a last stage only web pages with medical messages were considered.

Stages	2021	2022
The number of total web pages retrieved by BuzzSumo using selected queries	8552 web pages	6404 web pages
The number of web pages that provoked engagements at least ten times (videos were not evaluated)	233 web pages	137 web pages
Number of web pages considered by experts as containing medical content	161 web pages	105 web pages

In 2021, the most popular (in terms of the number of engagements) web pages were

related to the COVID-19 topic, with a number of 221,301, followed by the general vaccinations topic, with a score of 22,414. Similarly, in 2022 the most popular web pages were related to the COVID-19 topic (generated 18,264 engagements) followed by cholesterol and statins (generated 16,844 engagements). The topic most polluted with fake or non-credible content in 2021 was COVID-19 (57%), followed by general immunization (42%), and in 2022 diet, nutritional supplements, and alcohol (60%), followed by general vaccinations (30%). The percentage of fake or non-credible web pages related to COVID-19 and general vaccinations decreased from 57% to 24% and 42% to 30%, respectively. In contrast, the percentage of fake or non-credible web pages related to cholesterol/statins and cardiology increased from 17% to 26% and 18% to 26%, respectively (Figure 3).

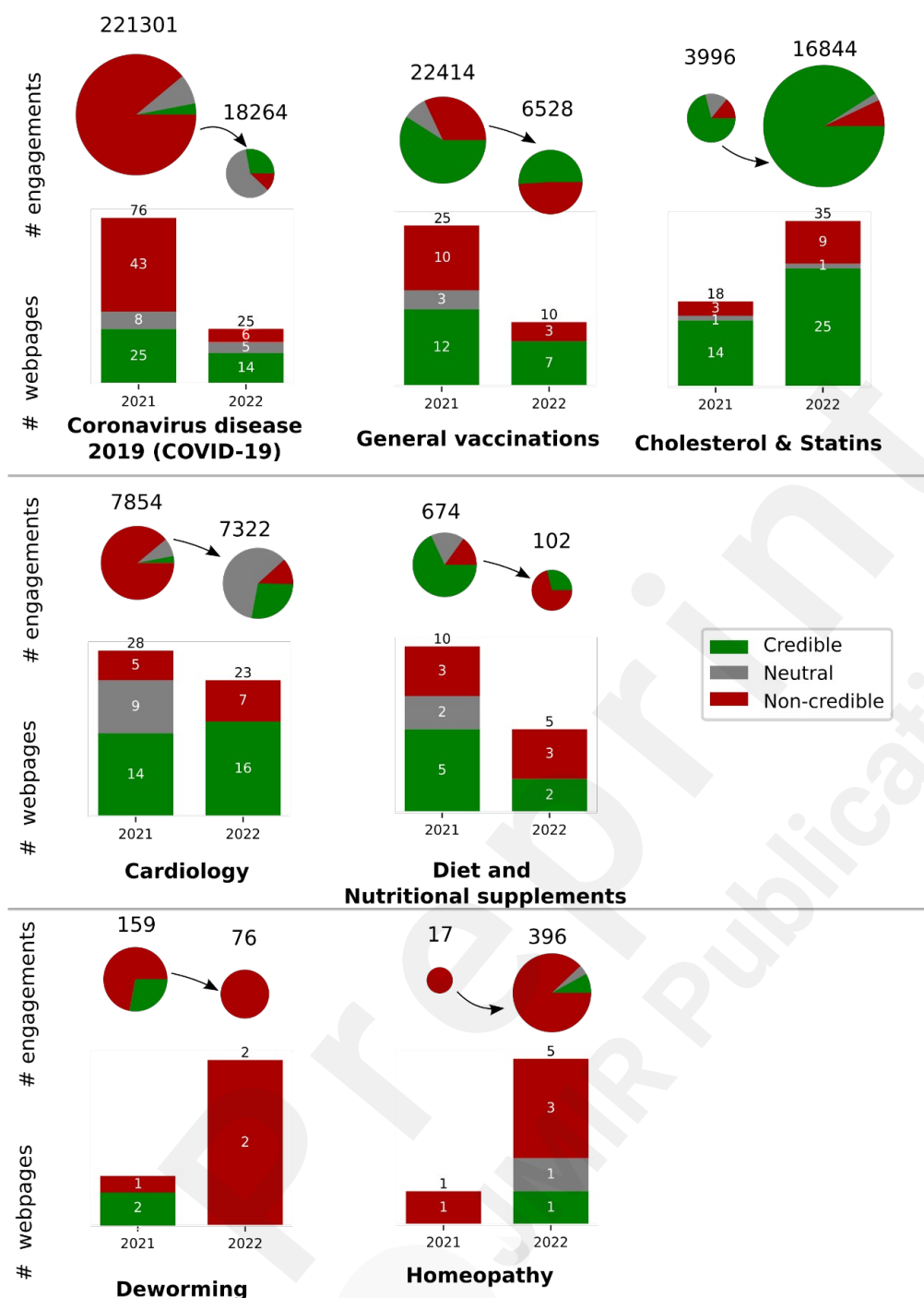


Figure 3. The distribution of the number of web pages and related social media engagements. Distributions are divided by topics identified during the years 2021 and 2022

Table 5 shows the number of medical web pages rated by medical experts as credible, non-credible, or neutral. Web pages classified as medical neutral did not contain any medical information in the context of diagnosis or treatment and mainly referred to health stories or described changes in medical regulations or medical conferences. In 2021, the number of medical web pages classified as credible was 72, non-credible was 66, and neutral was 23, while in 2022, they were 66, 32, and 7, respectively. In 2022, the percentage of non-credible content from all web pages was much lower (30%) than in 2021, which was 41%.

Table 5. Number of medical web pages, divided into credible, non-credible and neutral. The proportions are compared between the years 2021 and 2022 using chi-squared test or the Fisher's exact test ($p < 0.05$, *chi2 test, **Fisher's exact test).

Topics	2021			2022			p
	Credible medical web pages	Non-credible medical web pages	Neutral medical web pages	Credible medical web pages	Non-credible medical web pages	Neutral medical web pages	
Coronavirus disease 2019 (COVID-19)	25	43	8	14	6	5	.018*
General vaccinations	12	10	3	7	3	0	.467**
Cholesterol and statins	14	3	1	25	9	1	.760**
Cardiology	14	5	9	17	6	0	.007**
Diet, Nutritional supplements, Alcohol	5	3	2	2	3	0	.620**
Deworming	2	1	0	0	2	0	.400**
Homeopathy	0	1	0	1	3	1	1.000**
Total	72	66	23	66	32	7	.010*

In 2021, credible web pages generated 24,398 engagements, while in 2022, they generated 30,301 engagements. In 2021 and 2022, web pages containing fake news generated 210,018 and 7,942 engagements, respectively. Detailed data are shown in Table 6.

Table 6. The number of aggregated engagements in social media of medical web pages, divided into credible, non-credible, and neutral. The proportions are compared between the years 2021 and 2022 using chi-squared test or the Fisher's exact test ($p < 0.05$, *chi2 test, **Fisher's exact test).

Topics	2021			2022			p
	Credible medical web pages popularity	Non-credible medical web pages popularity	Neutral medical web pages popularity	Credible medical web pages popularity	Non-credible medical web pages popularity	Neutral medical web pages popularity	
Coronavirus disease 2019 (COVID-19)	6739	197100	17462	5138	2145	10981	<.001*
General vaccinations	13234	7226	1954	3325	3203	0	<.001**
Cholesterol and statins	2831	579	586	15324	1228	292	<.001*
Cardiology	1090	4883	1881	6454	868	0	<.001**
Diet, Nutritional supplements, Alcohol	459	99	116	30	72	0	<.001**
Deworming	45	114	0	0	76	0	<.001**
Homeopathy	0	17	0	30	350	16	.329**
Total	24398	210018	21999	30301	7942	11289	<.001*

Discussion

Principal results

Many researchers tried to create a model of algorithms to identify dis- and misinformation on the Web. Machine learning protocols and artificial intelligence can detect SARS-CoV-2 infection-related fake news from social media [*39, 13, 14]. Iwendi et al. proposed new features to recognize fake news about COVID-19 in online information with an accuracy of over 86% [14]. Machine Learning-supported actions taken during the COVID-19 pandemic to prevent the spread of fake news appear to have been effective. However, such interventions should be applied continuously to new medical fields susceptible to creating unreliable content due to its enormous impact on the community.

The current study was focused on the perspective of the lay person, as it is

difficult for a lay person to understand medical information and find reliable sources. Consequently, lay people are at risk of encountering non-credible information [9,25]. Whereas, health care providers use known reliable sources of medical knowledge based on EBM [10].

COVID-19 was doctors' most commonly reported "fake news generating" topic in their daily clinical practice in the last two years. Our study showed that although there are guidelines based on EBM for the treatment of COVID-19 [40], in Poland, the most popular web pages were still those related to the use of alternative methods of treating SARS-COV-2 infection (e.g., amantadine). In our study, the proportion of non-credible web pages in the selected sample regarding COVID-19 was 0,57 and 0,24 in 2021 and 2022, respectively. It is worth mentioning that credible web pages were less prevalent on the web, with a much lower number of engagements than non-credible ones in both years (6739 vs. 221301 and 5138 vs. 18264). It indicates that false information about COVID-19 was still prominent on the Web in 2021. On the other hand, the significant decrease in total engagements related to non-credible content may indicate that interventions to detect, remove, and correct disinformation about COVID-19, withheld by big tech companies like Google and Facebook, had a positive impact on diminishing the spread of fake news. On-platform interventions included highlighting, surfacing and prioritizing content from authoritative sources cooperation with fact-checkers and health authorities to flag and remove disinformation [41]. Another possible phenomenon for the reduced spread of Covid-related fake news could have been declining interest in the pandemic. Among possible reasons are that people got used to it, the reduction of the restrictions, and another global crisis that emerged in 2022 (the war in Ukraine). However, our study has found that while interest in non-credible information about COVID-19 has decreased from 2021 to 2022, the popularity of other topics (especially cholesterol/statins and cardiology) has increased. It seems that new topics were "filling up the vacuum" created by the decreasing interest in the virus.

In the case of the topic „cholesterol and statins“, social media has seen a huge increase in interest in this field between 2021 and 2022. Number of aggregated total engagements grew more than fourfold: from around 4,000 to around 16,800. This happened in parallel with an increase of interest in other cardiology topics and a decrease in interest in COVID-19. Most web page contents were assessed as credible, fake news content accounted for 17% in 2021 and raised to 26% in 2022.

During the period of increasing interest in the topic of “cholesterol and statins”, the percentage of content classified as non-credible also increased. This phenomenon was confirmed in the other studied categories. Hypercholesterolemia is a condition that does not directly cause clinical symptoms. For a long time it does not affect the comfort of life, until complications arise such as stroke or myocardial infarction. Patients come to doctor's offices for other reasons, instead they receive a diagnosis of hypercholesterolemia. Afterwards, doctors evaluate indications for pharmacological treatment. As a result of this specific compilation, doctor's competence in this field is often questioned. Due to low trust in the profession in Poland in general, fears of side effects and distant prospects of clinical benefits are high [26]. Patients are looking for alternative information on the Internet and in social media, and the increased interest in the subject is a breeding ground for the creation of fake news content.

According to the obtained results the topic of deworming generated the least engagement. Nevertheless, this problem was indicated as prevalent among patients in our entering questionnaire. The number of engagements for non-credible web pages regarding this topic increased from 30% in 2021 to as many as 100% in 2022. The absolute numbers of web pages are small (3 and 2 for 2021 and 2022, respectively) but the lack of any credible source of information about this topic in 2022 should be alarming to the medical community. The qualitative insight revealed the key role of social media influencers in the deworming narratives. Those facts indicate that deworming topics may not generate a lot of clicks, but provide a strong, convincing message to the lay person. There exists a study that shows that the analysis of internet searches predicts seasonal peaks in hairworm numbers with high accuracy and identifies host characteristics that exactly match the scientific data [42]. Considering this study, we can conclude that lay Internet users often look for information about parasites, but rarely reach reliable sources.

During the pandemic, we all faced the phenomenon called an “infodemic” - flow of a vast amount of information that includes false and fake news during the outbreak of a particular disease [15]. While COVID-19 is a living example of how harmful and contagious medical disinformation can be, there are still other health topics where online disinformation is prevalent. Those topics already generated disinformation before the Covid pandemic, and still pose a serious threat to society. Our study's strength is that the analysis of most of the unreliable topics appeared in

clinical practice. Moreover, the evaluation of the websites were made by an experienced team of clinicians.

Limitations

The initial questionnaire includes only 121 respondents. Unfortunately, at the time, doctors were overworked during the COVID-19 pandemic. However, they were doctors of different specialties, working in distinct workplaces, both inpatient and outpatient, and of different ages. Moreover, the doctors described all problems with disinformation they had encountered in their professional work. The questionnaire did not suggest which topics to address. Doctors independently, according to their clinical experience, described these phenomena. Thus, despite the relatively small number of respondents, the topics were explored comprehensively.

Analyzing only the first 10 websites selected with BuzzSumo using the queries may not be sufficient to discover the complete picture of the medical fake news spread throughout the Web. However, with our limited resources (time of the experts) we tried to capture the dynamics of the most popular contents as accurately as possible. Another limitation is the selection of websites in both time intervals using exactly the same queries. New medical topics in the second time slot may not have been selected. To overcome this limitation, we intend to reintroduce the questionnaire and check the dynamics for possible new, emergent topics generating much medical non-credible content on the Internet.

Conclusions

Our analysis highlights several important areas for improvement in interventions to stop the spread of non-credible OHI, which can be summarized as follows:

1. Interventions introduced by Big Tech companies have been effective against COVID-19 disinformation. However, the increase in dissemination of non-credible content between 2021 and 2022 in other topics indicates that such interventions should also cover non-COVID-related topics.
2. Topics that generate the most unreliable contents, besides COVID-19, is “cholesterol & statins” and “cardiology”. Perhaps it is worth starting interventions with such topics.
3. The case of the “deworming” topic shows that online activity does not always reflect the actual dissemination of disinformation. The topic, which was indicated frequently in questionnaires, generated few web pages, with only

few total engagements. It is worth mentioning that all web pages found in 2022 from this topic were non-credible. This highlights the importance of repeating questionnaires among physicians.

4. The case of the “deworming” also reveals that there are medical topics poorly covered by OHI from credible sources. Special emphasis should be put on such topics.
5. The dynamics of disinformation dissemination of frequently shared, commented or liked topics should be constantly monitored to react in advance to future infodemics.

Ethical approval

Ethical approval was not required as this study did not involve any individuals. No personal data was accessed and it did not put anyone at risk.

Funding

The research was carried out as part of the research work number INFOSTRATEG-I/0010/2021-00 at the Medical University of Bialystok (Poland), Polish-Japanese Academy of Information Technology in Warsaw (Poland) and Science4People sp. z o.o. in Szczecin (Poland) financed from a subsidy provided by the National Centre for Research and Development, Poland.

Data

The data was collected as part of the research work number INFOSTRATEG-I/0010/2021-00 at the Medical University of Bialystok (Poland), Polish-Japanese Academy of Information Technology in Warsaw (Poland) and Science4People sp. z o.o. in Szczecin (Poland) financed from a subsidy provided by the National Centre for Research and Development, Poland. Due to the involvement of a commercial entity in the research, access to the data is restricted. However, it may be made available upon a reasonable request directed to the corresponding author.

Availability

Conflict of interest

The authors declare no conflict of interest. The funder had no role in the design of the study, in the collection, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- [1] KRANZBERG, M., HISTORY OF TECHNOLOGY. SCIENCE, 1960. 131(3408): P. 1246-7.
- [2] SWIRE-THOMPSON, B. AND D. LAZER, PUBLIC HEALTH AND ONLINE MISINFORMATION:

CHALLENGES AND RECOMMENDATIONS. ANNU REV PUBLIC HEALTH, 2020. 41: P. 433-451. PMID: 31874069

[3] WIERZBICKI, A. (2018). WEB CONTENT CREDIBILITY. CHAM: SPRINGER INTERNATIONAL PUBLISHING.

[4] BAINES, D., & ELLIOTT, R. J. (2020). DEFINING MISINFORMATION, DISINFORMATION AND MALINFORMATION: AN URGENT NEED FOR CLARITY DURING THE COVID-19 INFODEMIC. DISCUSSION PAPERS, 20(06), 20-06.

[5] PACKER, M., ARE META-ANALYSES A FORM OF MEDICAL FAKE NEWS? THOUGHTS ABOUT HOW THEY SHOULD CONTRIBUTE TO MEDICAL SCIENCE AND PRACTICE. CIRCULATION, 2017. 136(22): P. 2097-2099. PMID: 29180491

[6] MAMAK, K., DO WE NEED THE CRIMINALIZATION OF MEDICAL FAKE NEWS? MED HEALTH CARE PHILOS, 2021. 24(2): P. 235-245. PMID: 33398487

[7] ISLAM, M.S., ET AL., COVID-19-RELATED INFODEMIC AND ITS IMPACT ON PUBLIC HEALTH: A GLOBAL SOCIAL MEDIA ANALYSIS. AM J TROP MED HYG, 2020. 103(4): P. 1621-1629. PMID: 32783794

[8] GAVARAVARAPU, S.M., ET AL., IMPACT OF 'INFODEMIC IN PANDEMIC' ON FOOD AND NUTRITION RELATED PERCEPTIONS AND PRACTICES OF INDIAN INTERNET USERS. PLOS ONE, 2022. 17(4): P. E0266705. PMID: 35446865

[9] BETSCH, C., ET AL., THE INFLUENCE OF VACCINE-CRITICAL WEBSITES ON PERCEIVING VACCINATION RISKS. J HEALTH PSYCHOL, 2010. 15(3): P. 446-55. PMID: 20348365

[10] GUYATT, G., D. COOK, AND B. HAYNES, EVIDENCE BASED MEDICINE HAS COME A LONG WAY. BMJ, 2004. 329 (7473): P. 990-1. PMID: 15514320

[11] SACKETT, DAVID L., ET AL. EVIDENCE BASED MEDICINE: WHAT IT IS AND WHAT IT ISN'T. BMJ, 1996. 312 (7023): P. 71-72. PMID: 8555924

[12] WIERSINGA, W.J., ET AL., PATHOPHYSIOLOGY, TRANSMISSION, DIAGNOSIS, AND TREATMENT OF CORONAVIRUS DISEASE 2019 (COVID-19): A REVIEW. JAMA, 2020. 324(8): P. 782-793. PMID: 32648899

[13] PENNYCOOK, G., ET AL., FIGHTING COVID-19 MISINFORMATION ON SOCIAL MEDIA: EXPERIMENTAL EVIDENCE FOR A SCALABLE ACCURACY-NUDGE INTERVENTION. PSYCHOL SCI, 2020. 31(7): P. 770-780. PMID: 32603243

[14] IWENDI, C., ET AL., COVID-19 FAKE NEWS SENTIMENT ANALYSIS. COMPUT ELECTR ENG, 2022. 101: P. 107967. PMID: 35474674

[15] INFODEMIC. , WHO, EDITOR.: AVAILABLE ONLINE: [HTTPS://WWW.WHO.INT/HEALTHTOPICS/INFODEMIC#TAB=TAB_1](https://www.who.int/healthtopics/infodemic#tab=tab_1).

[16] AN OVERVIEW OF SYSTEMATIC REVIEWS OF THE CURRENT STATE OF THE ART OF INFODEMICS AND HEALTH MISINFORMATION AND ITS REPERCUSSIONS IN PUBLIC HEALTH: RECOMMENDATIONS, CHALLENGES, AND AVAILABLE RESEARCH OPPORTUNITIES, MAY 2022, BULLETIN OF THE WORLD HEALTH ORGANISATION

[17] MONTAGNI, I., ET AL., ACCEPTANCE OF A COVID-19 VACCINE IS ASSOCIATED WITH ABILITY TO DETECT FAKE NEWS AND HEALTH LITERACY. J PUBLIC HEALTH (OXF), 2021. 43(4): P. 695-702. PMID: 33693905

[18] SOWA, P., ET AL., COVID-19 VACCINE HESITANCY IN POLAND-MULTIFACTORIAL IMPACT TRAJECTORIES. VACCINES (BASEL), 2021. 9(8). PMID: 34452001

[19] JOLLEY, D. AND K.M. DOUGLAS, THE EFFECTS OF ANTI-VACCINE CONSPIRACY THEORIES ON VACCINATION INTENTIONS. PLOS ONE, 2014. 9(2): P. E89177. PMID: 24586574

[20] TOLL, M. AND A. LI, VACCINE SENTIMENTS AND UNDER-VACCINATION: ATTITUDES AND BEHAVIOUR AROUND MEASLES, MUMPS, AND RUBELLA VACCINE (MMR) IN AN AUSTRALIAN COHORT. VACCINE, 2021. 39(4): P. 751-759. PMID: 33218781

[21] BORAH, P. AND J. HWANG, TRUST IN DOCTORS, POSITIVE ATTITUDES, AND

VACCINATION BEHAVIOR: THE ROLE OF DOCTOR-PATIENT COMMUNICATION IN H1N1 VACCINATION. *HEALTH COMMUN*, 2022. 37(11): P. 1423-1431. PMID: 33685304

[22] DUBÉ E, L.C., GUAY M, BRAMADAT P, ROY R, BETTINGER JA, VACCINE HESITANCY. . *HUMAN VACCINES AND IMMUNOTHERAPEUTICS*, 2013. 9(8): P. 1763-1773. PMID: 23584253

[23] SOMMARIVA, S., C. VAMOS, A. MANTZARLIS, L.U.-L. ĐÀO, AND D. MARTINEZ TYSON., SPREADING THE (FAKE) NEWS: EXPLORING HEALTH MESSAGES ON SOCIAL MEDIA AND THE IMPLICATIONS FOR HEALTH PROFESSIONALS USING A CASE STUDY. *AMERICAN JOURNAL OF HEALTH EDUCATION*, 2018. 49: P. 246-255.

[24] PAKALNISKIENE, V., ET AL., COULD BELIEF IN FAKE NEWS PREDICT VACCINATION BEHAVIOR IN THE ELDERLY? *INT J ENVIRON RES PUBLIC HEALTH*, 2022. 19(22). PMID: 36429616

[25] LOOMBA, S., ET AL., MEASURING THE IMPACT OF COVID-19 VACCINE MISINFORMATION ON VACCINATION INTENT IN THE UK AND USA. *NAT HUM BEHAV*, 2021. 5(3): P. 337-348. PMID: 33547453

[26] CATAPANO, A.L., O. WIKLUND, AND S. EUROPEAN ATHEROSCLEROSIS, THINK AGAIN ABOUT CHOLESTEROL SURVEY. *ATHEROSCLER SUPPL*, 2015. 20: P. 1-5. PMID: 26671304

[27] HILL, J.A., ET AL., MEDICAL MISINFORMATION. *CIRC CARDIOVASC INTERV*, 2019. 12(2): P. E007796. PMID: 30695791

[28] WASZAK, P.M., W. KASPRZYCKA-WASZAK, AND K. KUBANEK, THE SPREAD OF MEDICAL FAKE NEWS IN SOCIAL MEDIA – THE PILOT QUANTITATIVE STUDY. *HEALTH POLICY AND TECHNOLOGY*, 2018. 7: P. 115-118.

[29] REYNEKE, G., J. HUGHES, AND S. GRAFENAUER, CONSUMER UNDERSTANDING OF THE AUSTRALIAN DIETARY GUIDELINES: RECOMMENDATIONS FOR LEGUMES AND WHOLE GRAINS. *NUTRIENTS*, 2022. 14(9). PMID: 35565721

[30] RUANI, M.A. AND M.J. REISS, SUSCEPTIBILITY TO COVID-19 NUTRITION MISINFORMATION AND EATING BEHAVIOR CHANGE DURING LOCKDOWNS: AN INTERNATIONAL WEB-BASED SURVEY. *NUTRIENTS*, 2023. 15(2). PMID: 36678321

[31] INAN-EROGLU, E. AND Z. BUYUKTUNCER, YOUTUBE: IS IT A RELIABLE SOURCE OF NUTRITION INFORMATION ON COVID-19 PANDEMIC? *HEALTHCARE (BASEL)*, 2022. 10(10). PMID: 36292358

[32] MOON, H. AND G.H. LEE, EVALUATION OF KOREAN-LANGUAGE COVID-19-RELATED MEDICAL INFORMATION ON YOUTUBE: CROSS-SECTIONAL INFODEMIOLOGY STUDY. *J MED INTERNET RES*, 2020. 22(8): P. E20775. PMID: 32730221

[33] AHMED, I., ET AL., BEHAVIORAL PREVENTIVE MEASURES AND THE USE OF MEDICINES AND HERBAL PRODUCTS AMONG THE PUBLIC IN RESPONSE TO COVID-19 IN BANGLADESH: A CROSS-SECTIONAL STUDY. *PLOS ONE*, 2020. 15(12): P. E0243706. PMID: 33306725

[34] HAMULKA, J., ET AL., DIETARY SUPPLEMENTS DURING COVID-19 OUTBREAK. RESULTS OF GOOGLE TRENDS ANALYSIS SUPPORTED BY PLIFECOVID-19 ONLINE STUDIES. *NUTRIENTS*, 2020. 13(1). PMID: 33375422

[35] LEBIN, J.A., A. MUDAN, AND A.H.B. WU, CHRONIC ALCOHOL USE DOES NOT PROTECT AGAINST COVID-19 INFECTION. *AM J EMERG MED*, 2021. 45: P. 640-641. PMID: 33221114

[36] AGHABABAEIAN, H., L. HAMDANIEH, AND A. OSTADTAGHIZADEH, ALCOHOL INTAKE IN AN ATTEMPT TO FIGHT COVID-19: A MEDICAL MYTH IN IRAN. *ALCOHOL*, 2020. 88: P. 29-32. PMID: 32693023

[37] LUK TT, ZHAO S, WENG X, ET AL., EXPOSURE TO HEALTH MISINFORMATION ABOUT COVID-19 AND INCREASED TOBACCO AND ALCOHOL USE: A POPULATION-BASED SURVEY IN HONG KONG. *TOBACCO CONTROL*, 2021. 30: P. 696-699. PMID: 32855353

- [38] BUZZSUMO. AVAILABLE ONLINE: WWW.BUZZSUMO.COM.
- [39] KHAN, S., ET AL., DETECTING COVID-19-RELATED FAKE NEWS USING FEATURE EXTRACTION. FRONT PUBLIC HEALTH, 2021. 9: P. 788074. PMID: 35059379
- [40] FLISIAK, R., ZALECENIA DOTYCZĄCE POSTĘPOWANIA W ZAKAŻENIACH SARS-CoV-2 POLSKIEGO TOWARZYSTWA EPIDEMIOLOGÓW I LEKARZY CHOROÓB ZAKAŻNYCH Z 23 LUTEGO 2022 ROKU. MED. PRAKT., 2022(3): P. 49–77.
- [41] OECD. COMBATTING COVID-19 DISINFORMATION ON ONLINE PLATFORMS, OECD, 2020, PARIS.
- [42] DOHERTY, J.F., ET AL., THE PEOPLE VS SCIENCE: CAN PASSIVELY CROWDSOURCED INTERNET DATA SHED LIGHT ON HOST-PARASITE INTERACTIONS? PARASITOLOGY, 2021. 148(11): P. 1313-1319. PMID: 34103103

Abbreviations

OHI – Online Health Information

EMB – Evidence Based Medicine

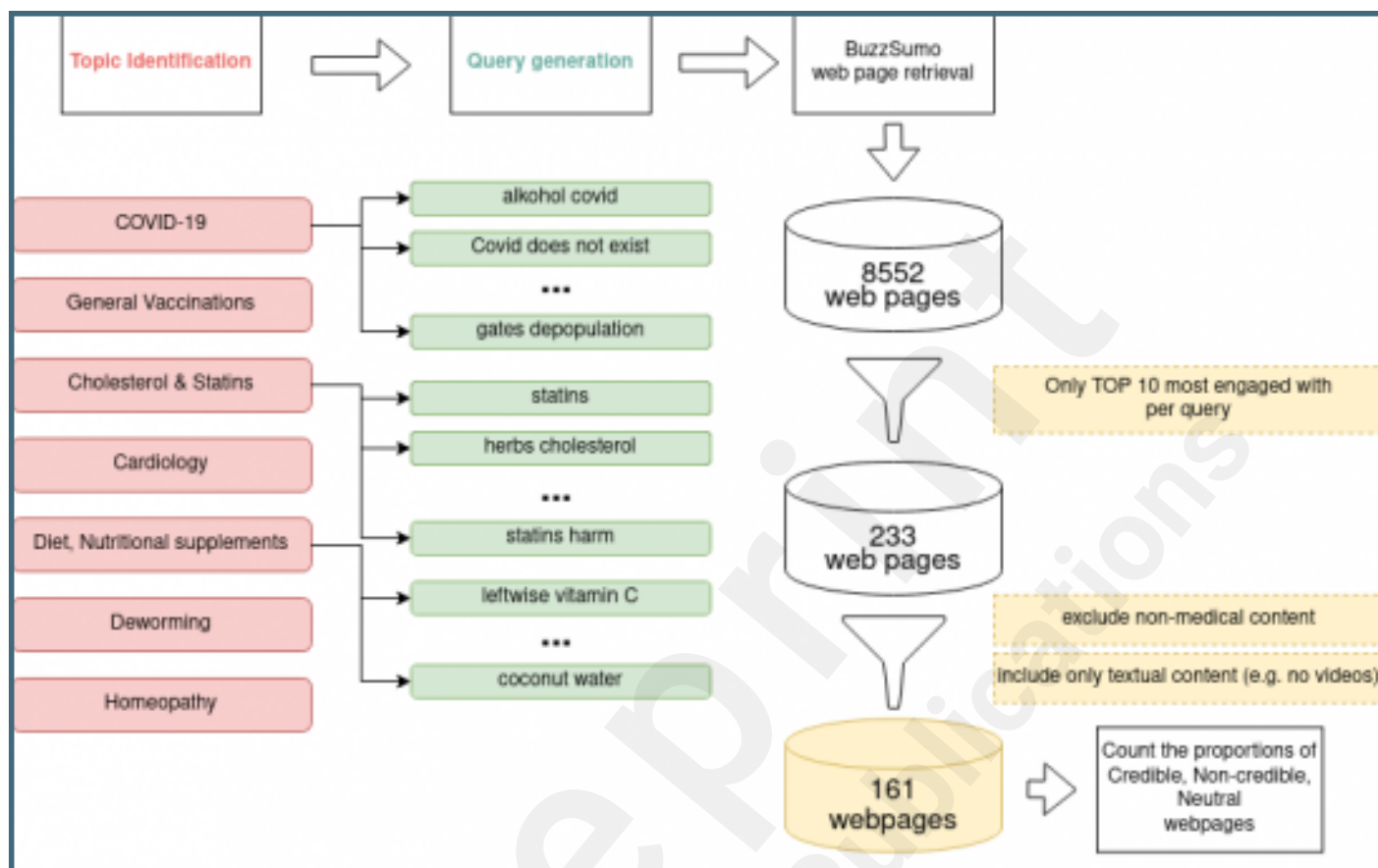
Supplementary Files

Figures

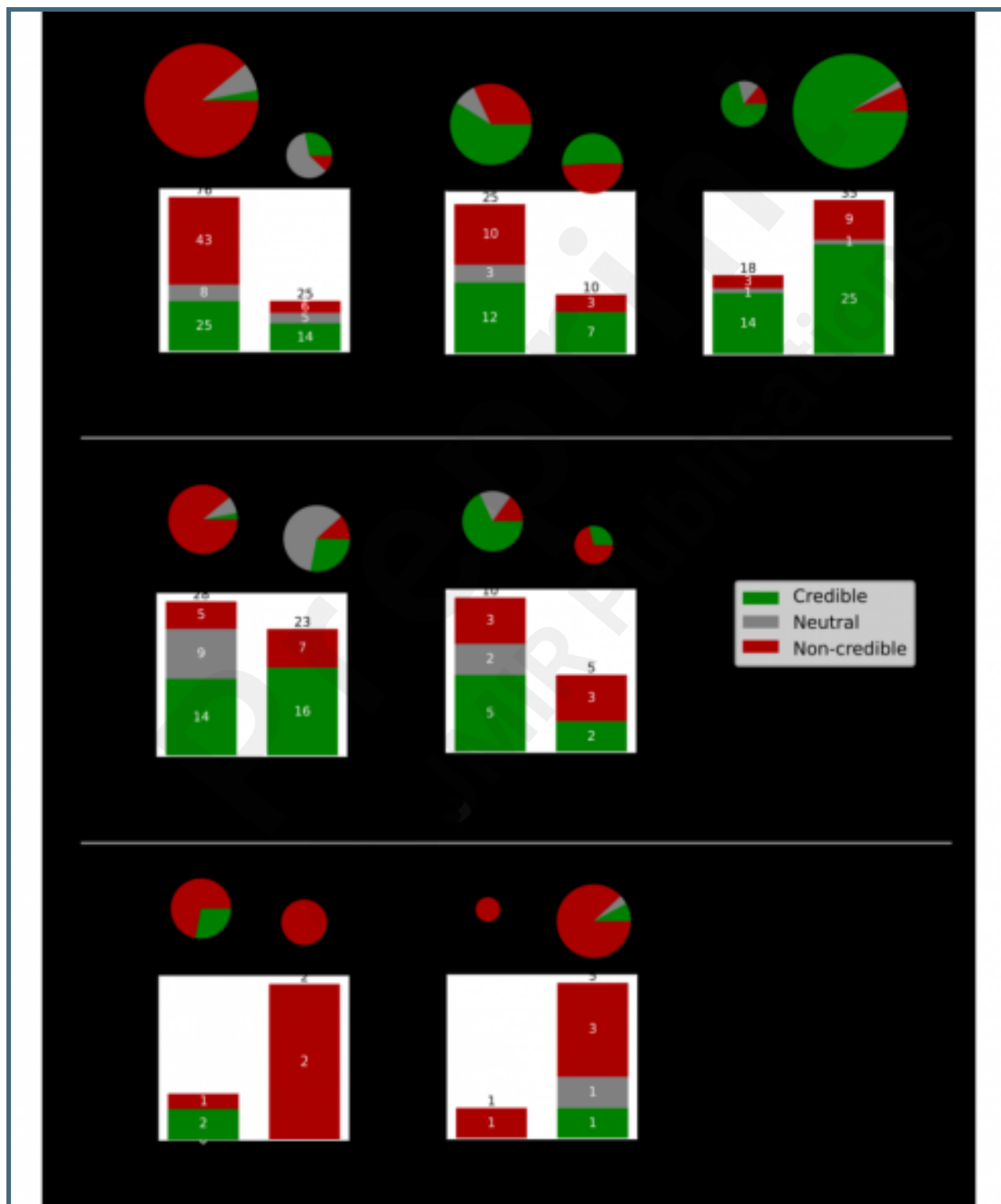
Four subsequent stages of web search research. The figure includes a step description, the product of the step, and an example.

	Step description	Product	Example
1.	Questionnaires from 121 physicians. Collecting false medical information that they have come across from patients in their professional practice.	list of real-life problems	"cardiological patients do not wish to take statins because they read about their poor effectiveness"
2.	Allocating problems into groups consistent in scope.	identifying topics	"statins are not effective"
3.	Generation of specific queries using DuckDuckGo search engine. We were looking for queries that yield unreliable web pages.	generating queries for specified topics	"statins do not work" "statins no effect" "statins hurt instead of help"
4.	Using identified queries to retrieve web pages and their social media engagement statistics using BuzzSumo.	web pages retrieval using queries	https://www..... ; shares: 123, likes: 455, comments: 21

The figure contains a detailed description of the subsequent stages of the web search research, with the year 2021 as an example. Topic identification, query generation, web page retrieval, and the subsequent web page filtration are depicted.



The figure shows the differences in the prevalence and interactions on social media of non-credible, credible, and neutral web pages in the two periods: (1) January 1, 2021 to December 31, 2021; (2) January 1, 2022 - December 31, 2022. The bars indicate the number of websites that exceeded the social media engagement threshold (top 10 web pages mostly interacted with per search query) in a given period related to a given medical topic. Pie charts show the number of interactions (shares, likes, comments) with non-credible, credible, and neutral web pages over a given period, on a given medical topic. The aggregated social media engagement data was retrieved using the BuzzSumo Enterprise Application.



Multimedia Appendixes

The preliminary questionnaire body.

URL: <http://asset.jmir.pub/assets/2fa498eef726d0098fbb861c9522d957.png>

A list of the most popular non-credible websites on a given topic in 2021 and 2022.

URL: <http://asset.jmir.pub/assets/92866ce8677e071ee6f2c7c32e0dad2f.doc>



Related publication(s) - for reviewers eyes onlies

Publication with similar ethical considerations.

URL: <http://asset.jmir.pub/assets/bfafc64b1dca5cfaa2c8fea6a950316c.pdf>