

Evaluating the Ability of Commercial Search Engines to Help People Answer Health Questions

by

Kamyar Setayesh Ghajar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2023

© Kamyar Setayesh Ghajar 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The act of seeking information pertaining to medical treatments and self-diagnosis is one of the applications of search engines. However online documents and websites offer convenience and efficiency in accessing information, it is important to acknowledge that they may contain incorrect and also unreliable information, which can potentially lead to adverse consequences such as making harmful medical decisions. This is particularly concerning when search engine users rely solely on the information they encounter through search results, without conducting additional research or seeking guidance from qualified medical professionals. Therefore, it is essential to assess the impact of search engines on users' behavior and decision-making processes, especially when it comes to health-related decisions. Previous research has been conducted to evaluate the extent to which people may be affected by search engine results when they are responding to health-related questions, upon which our study is based (Pogacar et al., 2017; Ghenai et al., 2020). Their findings indicated that individuals tend to make correct decisions when supplied with a series of correct information as search results, and conversely, they tend to make wrong decisions when presented with a group of search results with incorrect information. The prior research studies used a methodology whereby study participants were presented with static search results, without the ability to actively query a search engine. In our study, we designed and conducted a controlled laboratory study which followed a within-subject design that consisted of presenting a group of participants with 12 topics from TREC 2021 Health Misinformation track with each topic comprising a particular health issue and its corresponding suggested medical treatment. These treatments were categorized as either helpful or unhelpful for each health issue, but the participants were not aware of the true effectiveness of each treatment. The participants were then asked to evaluate the effectiveness of the treatments both with and without utilizing the search engine experience provided to them. The search engine environment was established using modern commercial search engine

APIs such as Google and Bing as its underlying infrastructure. This approach, unlike previous studies, allowed participants to directly engage with the search engine and submit their own queries to get their desired search results.

Our research revealed that search engine results have a substantial impact on individuals, both in terms of positive and negative effects. Significantly, the study participants made more incorrect decisions when they were engaged with topics with unhelpful treatments. Furthermore, it was discovered that there existed a positive correlation between the participants' level of prior knowledge of health issues and treatments, and their performance in making decisions. One might hypothesize that the results of [Pogacar et al. \(2017\)](#) were due in part of the use of static search result pages rather than a fully interactive search engine, but in our study we found that, even though the participants used a fully interactive search engine, interaction alone was not sufficient for participants to avoid being negatively influenced by the search engine on some search topics.

Acknowledgements

I wish to express my sincere gratitude to my supervisors, Professors Mark D. Smucker and Charles L. A. Clarke, for their invaluable assistance and direction. Their guidance and expertise were critical assets during my studies at the University of Waterloo. I would also like to thank Professors Gordon V. Cormack and Lukasz Golab for their feedback and comments regarding this work. I would also like to acknowledge the financial support provided by the Graduate Research Studentship Award and the International Student Award provided by the University of Waterloo and the Government of Canada.

This work was supported in part by Google, in part by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2020-04665, RGPAS2020-00080, RGPIN-2022-03065) and in part by the Digital Research Alliance of Canada.

Dedication

This is dedicated to my beloved parents, my precious brother and my wonderful significant other, without whose assistance this would not have been achievable.

Table of Contents

Author's Declaration	ii
Abstract	iii
Acknowledgements	v
Dedication	vi
List of Figures	x
List of Tables	xi
List of Abbreviations	xiv
1 Introduction	1
1.1 Research Motivation	3

1.2	Thesis Overview	3
1.3	Contributions	4
2	Related Work	6
2.1	Health-related Questions and Search Engines	7
2.2	Search Results effect on User Behavior	9
2.3	Web Search Bias	11
3	Methods and Materials	13
3.1	TREC Health Misinformation Track	14
3.2	Data	14
3.2.1	Topics	15
3.3	Study Design	18
3.3.1	Study Application UI	20
3.3.2	Within-Subjects Study Design	24
3.4	Participants	28
3.5	Data Cleaning	29

4	Results and Discussion	31
4.1	Search Engine Effect on Decision Making	31
4.2	Topic-by-topic Analysis	37
4.3	Search Engine Interactions	39
4.3.1	Clicks	39
4.3.2	Queries	43
4.3.3	Decision Time	44
4.4	Confidence and Prior Knowledge	50
4.4.1	Prior Knowledge	50
4.4.2	Confidence	56
5	Conclusion	64
	References	67
	APPENDICES	73
A	Participants' Queries	74

List of Figures

3.1	TREC 2021 Health Misinformation track topic example.	16
3.2	Study Web Application’s SERP Interface.	23
3.3	Study Web Application’s without SERP Interface.	24
3.4	12x12 Graeco-Latin square block created using sixteen 3x3 squares.	27
4.1	Study Web Application’s Pre-Task Questionnaire Example.	52
4.2	Study Web Application’s Post-Task Questionnaire Example.	60

List of Tables

3.1	Topics used in the study. Topics 106 and 123 were used for practice tasks.	19
3.2	Decision Categories.	21
4.1	Confusion matrices with accuracy and harmfulness of participants' decisions of the control and experimental conditions.	33
4.2	Confusion matrices of participants' y/n responses to the task questions in control and experimental conditions.	35
4.3	Final decisions made by participants for <i>Inconclusive</i> judged topics.	36
4.4	Search engine helpfulness analysis compared to the control condition per topic for participants' 3 level decisions (H: helpful, I: inconclusive, U: unhelpful).	40
4.5	Search engine helpfulness analysis compared to the control condition per topic for participants' responses to the task question.	41
4.6	Average number of total and unique clicks per-task based on decision accuracy.	42

4.7	Average number of total and unique clicks per search task on each SERP result rank.	43
4.8	Query type examples among different topics.	45
4.9	Average number of unique and total queries made by each participant per topic.	46
4.10	Average time in seconds it took participants to make their decision with and without responding to the post-task questionnaire.	47
4.11	Average time in seconds it took participants to make their decision per search task with and without responding to the post-task questionnaire.	49
4.12	Participants' prior knowledge of the health issue for search and control tasks.	54
4.13	Participants' average prior health issue knowledge score for control and search engine tasks. The weighted average scores are calculated by assigning following weights to each knowledge level: Nothing=0, Heard=1, General=2, Familiar=3, Extensive=4.	55
4.14	Participants' prior knowledge of the treatment for search and control tasks.	57
4.15	Participants' average prior treatment knowledge score for control and search engine tasks. The weighted average scores are calculated by assigning following weights to each knowledge level: Nothing=0, Heard=1, General=2, Familiar=3, Extensive=4.	58
4.16	Participants' decision confidence for search and control tasks.	62

4.17	Participants' decision confidence score for control and search engine tasks. The weighted average scores are calculated by assigning following weights to each confidence level: Very Uncertain=1, Somewhat Uncertain=2, Neutral=3, Somewhat Certain=4, Very Certain=5.	63
A.1	Queries made by participants per topic sorted by number of users put the query.	74

List of Abbreviations

API Application Programming Interface [20](#), [21](#), [32](#)

GCS Google Custom Search [20](#)

HTTP Hyper Text Transfer Protocol [21](#)

ISR Information Seeking and Retrieval [12](#)

LLM Large Language Model [10](#)

NIST National Institute of Standards and Technology [14](#), [15](#)

SERP Search Engine Results Page [22](#), [38](#), [39](#)

TREC Text REtrieval Conference [13–15](#), [32](#), [64](#)

UI User Interface [13](#), [20](#)

Chapter 1

Introduction

A better search engine is typically thought of as one that aids users in finding the most relevant results more quickly than other search engines. Additionally, results that seem relevant but may include inaccurate information must also be dealt with by modern search engines. If the user receives this inaccurate information, they may utilize the search engine results to make harmful decisions. Consider a patient who receives a cancer diagnosis, turns to a web search engine, and finds documents advertising a variety of treatments for sale. However, many of these treatments will have little impact or even detrimental effects on the patient's health. Incorrect treatment information may prevent people from seeking proven helpful treatments or cause them to choose harmful ones instead. In the past, several researchers have conducted studies to measure the impacts of misinformation on searchers' decision-making ([Bondarenko et al., 2021](#); [Ghenai et al., 2020](#); [Pogacar et al., 2017](#); [Song and Jiang, 2022](#); [White, 2014](#)).

The internet is home to a significant amount of false information and urban legends, particularly in the area of medicine and health ([Dai et al., 2020](#); [Windfeld and Meier, 2021](#)). Moreover, the proportion of individuals using Internet resources to seek health-related in-

formation has increased to 74.4% (Finney Rutten et al., 2019), and the primary method that people use to access online content on the internet is via modern and widely-used search engines. Individuals, despite the factual truth, often exhibit a pervasive inclination to have an uncontrolled tendency towards identifying treatments as effective (White and Hassan, 2014). Additionally, research has shown that search engines have the potential to display bias by inaccurately promoting the effectiveness of medical treatments that lack empirical evidence (Bondarenko et al., 2021; White, 2014; White and Hassan, 2014). Several previous research studies have been conducted to assess the impact of search engines on individuals' decision-making processes while seeking answers to their inquiries. Furthermore, researchers have conducted user-centric experiments aimed at quantifying the impact of web-based searches on the process of decision making (Kattenbeck and Elswailer, 2019), which led to the indication of individuals' ambiguity in assessing the reliability of online sources. The impact of search engines on people's decision-making processes is significant, since individuals' responses to search results may be greatly impacted by the prominence of certain answers within the search results (Abualsaud and Smucker, 2019; Draws et al., 2021; Ghenai et al., 2020; Song and Jiang, 2022). Therefore, academics, particularly those in the area of information retrieval, have found it essential to investigate user behavior throughout the decision-making process, and apply different enhancement models to the retrieval processes and algorithms, especially while in the context of self-diagnosis efficacy and the assessment of treatment success, search engines play a significant role (Popat et al., 2017, 2018; Pugachev et al., 2023). In cases when search results exhibit bias towards inaccurate information, there is an apparent decline in people's accuracy, possibly resulting in harmful decisions, especially if individuals rely on the reliability of online material (Ghenai, 2017; Pogacar et al., 2017; White, 2013, 2014).

1.1 Research Motivation

This study is a direct extension of the prior research conducted by [Pogacar et al. \(2017\)](#) and [Ghenai et al. \(2020\)](#), in which they provided study participants with a fixed list of made up search results for each task, without giving them the ability to generate their own queries as they would typically do on a search engine. Hence, the primary motivation of this research study was to examine the results of a comparable study design, focusing specifically on the topics from the TREC 2021 Health Misinformation track aiming to provide study participants with the ability to engage with a search engine environment, enabling them to issue their own queries and retrieve results from a modern commercial search engine like Google and Bing. By conducting this study, we aimed to assess the effect of modern commercial search engines in mitigating the adverse impact they may have on individuals' decision-making processes. Additionally, we intended to analyze users' behavioral patterns during their engagement with a search engine while seeking health-related information.

1.2 Thesis Overview

In this thesis, a user experiment was conducted to assess the decision-making proficiency of participants while answering health-related questions while using modern commercial search engines, Google and Bing in this study, in contrast to situations when no search engines are used. The participants were recruited to perform 12 tasks to answer treatment effectiveness questions selected from the topics from TREC 2021 Health Misinformation track. Chapter 2 is focused on the previous related work and studies in information retrieval systems effect on user decision-making process, related health misinformation studies and web search bias. Chapter 3 mainly discusses about the user study design, development and design of the study application, data and topics used for the study and participant

recruitment. In Chapter 4, we report our findings and results from the study such as the effects of search engines on users' decision-making process when responding to health related questions, participants' behavior analysis with and without a search environment and their correlations with individuals' accuracy in their decision about medical treatment effectiveness. Finally the thesis concludes in Chapter 5.

1.3 Contributions

This thesis makes the following contributions:

- Our findings reveal that even interacting with modern commercial search engines does not completely alleviate the problem of search engine's negative impacts with some unhelpful treatments. Therefore, while the accuracy without using a search engine on helpful and unhelpful treatments are 49% and 29%, respectively, using a search engine increases the participants' decision accuracy on helpful topics to 85%, but on unhelpful topics, the performance remains poor with an accuracy of 33%.
- Limited number of user clicks on search results (less than 2 clicks on average) throughout the search process might potentially lead to a higher likelihood of making incorrect decisions.
- Having prior knowledge of a health issue or medical treatment can shield users from potentially inaccurate information found in search results and help them achieve higher accuracy when assessing a treatment's effectiveness.
- Utilizing a search engine has the potential to enhance users' confidence in their decision-making processes, even in instances where their final decisions may prove

to be incorrect. However, it has been observed that individuals who express a high degree of certainty in their decisions often demonstrate a greater accuracy in their final decisions.

Chapter 2

Related Work

Our work directly builds upon the previous work of [Pogacar et al. \(2017\)](#) and [Ghenai et al. \(2020\)](#) in which the researchers requested participants to evaluate the effectiveness of various medical treatments for certain health conditions using a set of search results that had been carefully selected.

[Pogacar et al. \(2017\)](#) defined ten tasks for each participant in which they used a balanced study design with five helpful and five unhelpful medical treatments. Participants were exposed to a control condition, without the search results, for two of the ten tasks. This control condition enabled the researchers to calculate the proportion of accurate and harmful decisions that participants would make if they did not use the search engine results and their web pages as screenshots. Also, for some tasks the search results were intentionally skewed, with some favoring correct information while others favored incorrect information. Their research showed that search engine results have a substantial impact on individuals, both in good and bad ways. In contrast to abstaining from the use of search engine results, individuals' level of accuracy declined from 43% to 23% when they engaged with search results that were biased towards incorrect information. They also found out, when peo-

ple engage with search results that are skewed toward correct information, their accuracy increases to 65%. Additionally, in their research, it has been shown that people exhibit a bias towards wanting treatments to be helpful. This bias, when coupled with incorrect information, has the potential to inflict harm on people.

Ghenai et al. (2020) designed an experiment that combined participants' think-aloud and one-on-one interviews with the previous work of Pogacar et al. (2017). In this study, the researchers asked participants to rate the effectiveness of eight medical treatments by providing them with search engine results and cached web pages. In the experiment, to introduce bias to the given results, they showed eight correct search results and two incorrect ones when biased toward correct information. They displayed eight incorrect search results and two correct ones when biased toward incorrect information. They had asked participants to explicitly indicate what they thought by saying it out loud while performing the tasks. In addition, during the interview, they asked participants about their decisions throughout the task, as well as their usual usage of search engines for answering similar health-related questions. Their study results revealed that, even when participants pay close attention to the task, they are highly impacted by a search result bias. Participants' accuracy reached 67% when search results were skewed toward correct information, and 32% when search results were biased toward incorrect information.

In what follows in this chapter, we cover the previous studies and background related to this research study.

2.1 Health-related Questions and Search Engines

Search engines have become key tools for individuals to get information, serving as one of the most important ways of finding answers to their questions. This is particularly evident

in the context of health-related inquiries when individuals rely on search engines to address issues pertaining to their well-being or medical condition. Also, in light of the growth of Web 2.0 and the emergence of social networks, internet users now have the opportunity to get required information by consulting the thoughts and postings of other individuals, which can even be potentially harmful (Ghenai, 2017). In fact, Individuals are likely to encounter a combination of fact-based information and personal stories about patients' experiences (Blythe et al., 2017). When considering the wider context of the online landscape following the advent of social networks, it was noted that in the United States in the year 2012, 72% of internet users actively sought health-related information online, and 77% of these individuals initiated their search process by utilizing a search engine platform such as Google, Bing, or Yahoo, However, just 50% of them would seek consultation from a healthcare professional subsequent to doing a web search (Fox and Duggan, 2013; Jimmy et al., 2019; Rekabsaz et al., 2021). This percentage of internet users who use online content for their health-related questions has reached 74.4% in 2017 (Finney Rutten et al., 2019). Moreover, when individuals encounter a health issue, they have a tendency to engage in self-diagnosis and actively seek out further information pertaining to their specific health condition (Best et al., 2016), especially when the majority of internet users in the United States exhibit a sense of confidence in their search abilities, firmly believing that they are able to find information that is correct and reliable (Purcell et al., 2012). Online health information encompasses a wide range of topics, including but not limited to symptoms, diagnosis, and treatments of various illnesses, as well as general knowledge about weight reduction, healthy diets, and wellness recommendations (Ghahramani and Wang, 2020). The use of online platforms for accessing health information enables individuals to acquire knowledge pertaining to their health concerns, effectively manage health-related challenges, make informed choices about their well-being, and facilitate positive modifications in their behavior (Ghahramani and Wang, 2020; Lambert and Loiselle, 2007). The internet is widely populated with a significant amount of misinformation and super-

stitutions, particularly in the field of health (Dai et al., 2020; Windfeld and Meier, 2021). The significance of misinformation within the field of medicine and health has gained considerable relevance in relation to the COVID-19 pandemic. During the first stages of the pandemic, a substantial amount of misinformation pertaining to the origins and modes of transmission of the virus was spread. In the context of the COVID-19 pandemic, as well as other health-related conditions, the importance of ensuring the legitimacy and accuracy of online information, as well as tackling misinformation on the internet, cannot be overstated (Dai et al., 2020; Geldsetzer, 2020).

Furthermore, in contrast to the findings of studies which suggest that a substantial number of search results shown on the first page of search engines provide reliable and credible information to individuals in search of diverse health-related subjects (Kitchens et al., 2014), some other studies have provided evidence that search engines tend to prioritize websites that have captions with potentially alarming content, resulting in a higher likelihood of increased click-through rates (White and Horvitz, 2013). Even after controlling for rank position and relevance by means of click-through data down-sampling and click inversions, individuals searching for symptoms exhibit a notable inclination towards considering and clicking on captions that contain health-related content that may potentially be alarming (White and Horvitz, 2013). Click inversions refer to situations where the click-through rate (CTR) of a search result is higher than the result immediately above it, thus overcoming the biases related to position and visual evaluation of the results (Joachims et al., 2007).

2.2 Search Results effect on User Behavior

Studies have shown that search engines may exhibit bias by erroneously suggesting the effectiveness of medical treatments that lack empirical support (White, 2014). Further-

more, such inaccuracies may be further magnified due to individuals' bias towards favoring positive information (Kayhan, 2013; White, 2013). Research has shown that search engines may exhibit a bias towards favorably representing the helpfulness of medical treatments, regardless of their correctness which in some cases has even led to confirming misbeliefs (Bondarenko et al., 2021; White and Hassan, 2014). The influence of search engines on individuals' decision-making processes is considerable, as the manner in which individuals react to search results might be highly influenced by the prominence of particular answers within the search results (Abualsaud and Smucker, 2019; Draws et al., 2021; Ghenai et al., 2020; Song and Jiang, 2022). Specifically, in instances when search results are biased towards incorrect information, there is a notable decline in individuals' accuracy, which may potentially lead to harmful decisions, particularly if individuals place trust in the credibility of online content (Pogacar et al., 2017). As it is established that users who are exposed to a greater quantity of correct information, compared to users who are exposed to a bigger quantity of incorrect information, tend to make a higher proportion of correct and non-harmful decisions (Abualsaud and Smucker, 2019), search engine researchers have used current advancements in knowledge-based ranking methods (Vakili Tahami, 2022) and Large Language Models (LLMs) to extract the linguistic features of documents in order to determine their credibility prior to aggregating them for generating a final prediction about claims found on the web. (Popat et al., 2017, 2018; Pugachev et al., 2023; Zhang et al., 2022).

There also has been studies working on the efficacy of online web search as self-diagnosing tool. Some work have utilized an extensive collection of online search data to predict the adverse reactions associated with pharmaceutical substances. In these studies, the authors have constructed a predictive model by using individuals' search queries for post-illness side effects. This model aims to anticipate potential adverse effects of various medications that may need a longer duration to become apparent (Odgers et al., 2014; White et al., 2016; Yom-Tov et al., 2013).

Moreover, researchers have designed user-centered studies to measure the effects of web search on health issue diagnosis or finding a treatment. For example, conducting a think-aloud user research with the aim of elucidating the cognitive processes used by individuals in evaluating the credibility of search result pages ([Kattenbeck and Elsweler, 2019](#)). The results of their study indicate that individuals exhibit uncertainty when evaluating the credibility of internet sources.

2.3 Web Search Bias

Search engines use information retrieval techniques to retrieve the most relevant documents, resulting in search result pages that exhibit various algorithmic biases. There has been also previous research pertaining to bias in search results and also addressing bias in search behavior. In some studies, they have examined the impact of results' rank on people's level of accuracy in responding to health-related questions when using internet search platforms by implementing distinct online search interfaces with different characteristics of ranking bias ([Allam et al., 2014](#); [Lau and Coiera, 2007, 2009](#)).

There also has been research conducted to study the influence of cognitive biases and heuristics on the processes of information retrieval and decision-making. A cognitive bias refers to a consistent and predictable tendency in human thinking that may result in inaccuracies in judgments and decision-making processes ([Tversky and Kahneman, 1974, 1992](#)). Individuals may encounter cognitive biases while engaging in the process of knowledge retrieval. Several elements, including a person's previous belief (known as the anchoring effect), the order in which documents were accessed (referred to as the order effect), and the amount of time individuals spent on documents (referred to as the exposure effect), have been identified as influential factors in shaping the way questions were answered after doing a search ([Lau and Coiera, 2007](#)). Cognitive biases may have a significant influence

on the behaviors and results of [Information Seeking and Retrieval \(ISR\)](#). The area is particularly concerned with the potential amplification of cognitive biases as a result of the immediate access to vast amounts of information and the potential exploitation of these biases by search engines and content authors, whether intentional or unintentional ([Baeza-Yates, 2018](#); [Carr, 2008](#); [Cho and Roy, 2004](#)). When considering these biases originating from both the system and the user, it is evident that they may interact and potentially intensify the impact, whether it be positive or negative ([Kumpulainen et al., 2015](#)). It has become evident that the current state of knowledge within the field is just preliminary in comprehending the intricate nature of cognitive biases in [ISR](#) operations ([Azzopardi, 2021](#)).

Chapter 3

Methods and Materials

In this study, we brought participants to measure how well they can assess the effectiveness of various treatments for the associated health issues. To ask the participants how effective a treatment is for a particular health issue, we exposed them to our developed controlled search engine [User Interface \(UI\)](#) and questionnaires about each treatment to ask them about their prior and subsequent knowledge. Additionally, using the same web app we had created, we collected information about participants' general demographics via a questionnaire at the start of the research. All participant interactions with the program, including clicks, used search queries, answers, and how long it takes them to find the answer, were logged throughout the study. In this chapter, we first talk about the topics we used as the set of health issues and their corresponding treatment based on the 2021 [Text REtrieval Conference \(TREC\)](#) Health Misinformation track's collection; then, we mention the development of the dedicated web app as the study environment and explain the study design. We finally provide our study participants' insights.

3.1 TREC Health Misinformation Track

The [National Institute of Standards and Technology \(NIST\)](#) is an agency of the United States Department of Commerce that annually hosts the [TREC](#) workshop to enhance information retrieval research. Since TREC’s inception in 1992, it has doubled the effectiveness of retrieval, and by creating several large test data collections, it has also made the large-scale evaluation of the text retrieval process easier and more available ([Voorhees et al., 2005](#)). [TREC](#) offers participants a variety of tracks to choose from each year. The Health Misinformation track, formerly known as the Decision track, aimed to elevate accurate and reliable health-related information on the web. In our study, we have used the topics and run collections from the [TREC 2021 Health Misinformation track](#)¹. The run collection refers to the all 71 runs submitted and evaluated in the ad-hoc retrieval task from the [TREC 2021 Health Misinformation track](#) ([Clarke et al., 2021](#)).

3.2 Data

For this research study, we used the topics from the [TREC 2021 Health Misinformation track](#). Each topic focuses on a particular health issue and a treatment specific to the corresponding health issue, which reflects a user seeking information to assist them in deciding whether that treatment is effective in addressing the health issue or not. We also used the run collection presented by the track’s participants that year to select a subset of topics from all of them to conduct the research based on those selected topics. The [TREC 2021 Health Misinformation track](#) had 50 topics. Each topic has the following fields:

- **Number:** a unique number per topic from 101 to 150

¹<https://trec-health-misinfo.github.io/2021.html>

- **Query:** some keywords related to the topic
- **Description:** the topic paraphrased as a natural language question
- **Narrative:** more information about the topic
- **Disclaimer:** lets the reader know that none of the material is intended to be medical advice that is the same text for all topics
- **Stance:** answers the topic question with one of the following options:
 1. *Helpful:* track organizers deemed the topic’s suggested treatment helpful for the specified health issue based on a credible source.
 2. *Unhelpful:* track organizers deemed the topic’s suggested treatment unhelpful for the specified health issue based on a credible source.
- **Evidence:** a link to a credible source that supports the track organizers labeled stance

Figure 3.1 shows an example of a topic.

3.2.1 Topics

To conduct this study, we needed to select a subset of [TREC 2021 Health Misinformation](#) track topics with an equal number of helpful and unhelpful stances to be able to have a statistically balanced analysis of the outcomes of the study. Moreover, we were looking to choose the most informative subset of topics among those 50 topics that had the most correlation with the whole set. We used the [TREC 2021 Health Misinformation](#) track’s ad-hoc retrieval task submitted runs to perform this topic subset selection. The [NIST](#)

Topic 114

Number: 114

Query: vitamin e cream for skin scars

Description: Can applying vitamin E cream remove skin scars?

Narrative: Skin scars form when a wound in the skin heals and leave marks on the skin that look and/or feel different from its normal appearance. Vitamin E is an antioxidant. Vitamin E cream is a cream with Vitamin E as an ingredient that could be applied to skin. A very useful document would discuss the efficacy of applying Vitamin E creams for improving the cosmetic appearance of scars. A useful document may include information about improving the cosmetic appearance of skin scars, the use of topical Vitamin E creams for skin, or both.

Disclaimer: We do not claim to be providing medical advice, and medical decisions should never be made based on the stance we have chosen. Consult a medical doctor for professional advice.

Stance: unhelpful

Evidence: <https://pubmed.ncbi.nlm.nih.gov/10417589/>

Figure 3.1: TREC 2021 Health Misinformation track topic example.

assessors were only able to provide assessed documents for 35 out of the total 50 topics, and also among those 35 topics, three topics had no assessed harmful document. 14 topics of these 32 topics were labeled *helpful* and the rest 18 topics were labeled *unhelpful* by the track organizers.

In total, 71 runs had been submitted for the ad-hoc retrieval task (Clarke et al., 2021) for the aforementioned 32 topics, and we used all of these run submissions to select a subset of topics from the 32 topics. Each run submission was a file submitted by one of the participant groups in the track, which had a ranked list of documents per topic, ranked with the corresponding team’s ranker algorithm at the time based on the *compatibility* metric (Clarke et al., 2020). Regardless of the track’s document set and qrels² and the retrieval method used by each team to generate those submitted run files, we wanted to select a subset with an equal number of *helpful* and *unhelpful* topics that had the most correlation with the whole topics set ranking, in terms of similarity of test collections rankings between the selected topics and the whole set of topics. To measure this similarity, we used Kendall’s Tau(τ) rank correlation coefficient (Kendall, 1938).

We had planned to present all selected topics to all participants so we could gather as much data as possible. Thus, as we had three different experimental conditions (interfaces) to be studied in our research, to have a balanced study design, it was required for the number of selected topics to be divisible by three so we could expose every participant to each interface an equal number of times. We calculated the Kendall’s Tau(τ) rank correlation coefficient between each subset of the 32 assessed topics with an equal number of *helpful* and *unhelpful* topics, and the whole 32 topics, based on all track submitted run files. Then, we picked the 12 topics with the highest τ as the topic set of our study. We also selected topics two other topics to use for the practice tasks we presented to the participants. Topic 106 was used for a practice task with no search engine (control condition), and topic 123 was used for a practice task with the search engine. Table 3.1 shows all the 12 topics we

²query relevance set

used in the study plus the practice task topics.

3.3 Study Design

Before starting the main study, we went through a screening phase to ask the eligible volunteers, based on our screening criteria, to participate in our study. The screening criteria comprised: 1)the volunteer’s ability to read, comprehend, and communicate fluently in English as all study materials were in English, 2)having access to a desktop/laptop computer with a standard up-to-date browser and a reliable internet connection as the study web application was served online for the participants to log in and participate in the research study, and 3)being a resident of Canada as we were aiming to remunerate the participants in Amazon.ca e-gift cards in Canadian Dollars (CAD). After the screening, we invited the eligible volunteers to participate in the study and fill out and sign a consent form. Once participants are requested to complete and sign the consent form using an online Google form, they will be provided with an anonymous participant ID and the study-designated web application URL for logging in. Upon commencement of the research, participants were required to access the web application and thereafter complete a brief demographics questionnaire pertaining to their background and prior experience with web search. Subsequently, the participants were provided with on-screen instructions pertaining to the research, accompanied by a quiz and two practice tasks. The quiz’s purpose at the end of the instructions and the practice tasks was to facilitate the participants’ acquisition of the necessary skills to execute the actual study tasks successfully. After that, participants were redirected to the actual study’s set of 12 tasks, wherein they will be asked about the efficacy of a suggested treatment for a particular health issue, with or without using our designated search engine platform. Each task included the completion of a pre-task questionnaire in which we asked the participant’s level of prior knowledge of the task’s health

Table 3.1: Topics used in the study. Topics 106 and 123 were used for practice tasks.

#	Question	Stance	Issue	Treatment
101	Will wearing an ankle brace help heal Achilles tendonitis?	unhelpful	Achilles tendonitis	ankle brace
103	Can folic acid help improve cognition and treat dementia?	unhelpful	dementia	folic acid
107	Does yoga improve the management of asthma?	helpful	asthma	yoga
109	Does selenium help prevent cancer?	unhelpful	cancer prevention	selenium
112	Can evening primrose oil help treat eczema?	unhelpful	eczema	evening primrose oil
115	Can magnesium help prevent migraines?	helpful	migraine prevention	magnesium
118	Can dupixent treat eczema?	helpful	eczema	dupilumab
128	Does steam from a shower help croup?	unhelpful	croup	breathing hot shower steam
134	Can I remove a tick by covering it with Vaseline?	unhelpful	remove tick from skin	Vaseline
136	Can eating dates help manage iron deficiency anemia?	helpful	iron deficiency anemia	dates
140	Can fungal creams treat athlete's foot?	helpful	athlete's foot	fungal cream
149	Will at-home exercises manage hip osteoarthritis pain?	helpful	osteoarthritis	at-home exercise
106	Can vitamin B12 and sun exposure together help treat vitiligo?	helpful	vitiligo	vitamin B12 and sun exposure
123	Can I get rid of a pimple overnight by applying toothpaste?	unhelpful	pimples	toothpaste

issue and suggested treatment and also a post-task questionnaire in which we asked about the participant’s confidence in their decision about the effectiveness of the treatment and their explicit Yes/No response to the task’s topic question. Upon completing the 12 tasks, we expressed our gratitude to the participants, and then they logged out of the online study web application. At the end of the study, the participants were remunerated and required to complete and sign an acknowledgment form as a means of verifying that they had received remuneration for their participation.

3.3.1 Study Application UI

We conducted the research study by developing an online web application for participants to log in and perform their assigned study tasks. Each participant was assigned 12 tasks in which we asked them to judge a suggested medical treatment’s effectiveness for a certain health issue. In order to assess the efficacy of each medical treatment, the participants were presented with three alternatives for evaluation: 1)*helpful*, 2)*unhelpful*, and 3)*inconclusive*. Table 3.2 shows these decision categories with an example. For 4 out of 12 tasks, the participants were assigned to the control condition, whereby they did not obtain our provided search engine environment, so they had to judge the medical treatment effectiveness with their own prior knowledge. For the other 8 of the 12 tasks, every participant had to interact with our provided search engine environment to help them determine the task’s sought answer. In 4 of these search tasks we used the [Google Custom Search \(GCS\) Application Programming Interface \(API\)](https://developers.google.com/custom-search) ³ and for the other 4 tasks, without any differences in the study web app UI compared to the other 4 search tasks, we used the Bing Web Search [API](https://www.microsoft.com/bing/apis/bing-web-search-api) ⁴ as the back-end of the search engine environment. To make a unified user interface design for the search engine results page, we used the mentioned [API](#) documentations to

³<https://developers.google.com/custom-search>

⁴<https://www.microsoft.com/bing/apis/bing-web-search-api>

Table 3.2: Decision Categories.

Category	Explanation	Example
Helpful	The treatment is successful and has a direct and beneficial impact on the health problem that was described before.	Does Advil help relieve pain? Yes, this treatment helps .
Inconclusive	If medical professionals are still uncertain as to whether the treatment will have a good, negative, or no impact at all on the health condition, they may choose not to proceed with it.	Does reminiscence therapy help dementia? Medical professionals had inconclusive findings about the effectiveness of this treatment.
Unhelpful	Either the treatment does not have any effect at all, or it has an immediate and unfavorable impact on the particular health condition.	Does drinking dairy products help the stomach flu? No, this treatment does not help .

match with our desired data which we wanted to show to the participants as the search results. Thus, we implemented a mapping of the [Hyper Text Transfer Protocol \(HTTP\) API](#) response fields to our required fields which were the pages title, snippet, url and rank. Also, we implemented the pagination functionality based on the provided [APIs](#) to make the search engine experience as similar to the participants prior everyday experience as possible.

For each task, the medical treatment question was prominently shown at the top of the

page throughout the entire task from the pre-task questionnaire up until the end of the post-task questionnaire, accompanied by a concise boxed paragraph with definitions pertaining to the health issue and its corresponding treatment. Before each task, in a pre-task questionnaire, we asked about the participant’s level of familiarity with both the health issue and the suggested treatment. Unlike the previous work of [White \(2014\)](#) and [White and Horvitz \(2015\)](#) but similar to studies of [Pogacar et al. \(2017\)](#) and [Ghenai et al. \(2020\)](#), in the pre-task questionnaire, asking for participants’ prior beliefs before the search task was omitted due to an expectation of biased conduct. It was anticipated that participants may exhibit resistance towards changing their already mentioned beliefs, as they may be reluctant to acknowledge the potential error in their own decision. So, after the participant judged the effectiveness of the treatment for the health issue in the actual task, we then redirected them to a post-task questionnaire in which we asked about their confidence in their decision, and we also asked their opinion as a Yes/No question to the task’s main question that is mentioned in the task information box on top of each page.

For the tasks that we had provided the search engine environment, unlike the aforementioned previous work, participants were able to put in their own desired queries and then get the search results they had sought. They also could access the search results by clicking on them to be redirected to the web page they might have been looking for to judge the treatment’s effectiveness and come back to the [Search Engine Results Page \(SERP\)](#) to check other results or go to the next pages for more results. A reminder of the definitions pertaining to the several kinds of medical treatment effectiveness, namely “*helpful*,” “*inconclusive*,” and “*unhelpful*,” buttons in green, grey, and red colors respectively had been shown on the right side of the search results page. [Figure 3.2](#) shows the [SERP](#) interface design of the study’s developed web application for a task with a provided search engine.

In 4 of the 12 tasks that we asked each participant to perform in the study, we did not provide them with a search engine. In these tasks, we were looking for participants’ prior knowledge responses so we could compare the effects of the availability of a search engine

Task Question: Does selenium help prevent cancer?

Treatment: Selenium is a mineral essential for human health.

Health Issue: Cancer is a disease where the cells in the body start to grow and spread uncontrollably.

Selenium for preventing cancer - PMC

Overall, there is no evidence to suggest that increasing **selenium** intake through diet or supplementation prevents **cancer** in humans. However, more research is ...
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6491296/>

Selenium for preventing cancer

Jan 29, 2018 ... A hypothesis generated by the Nutritional **Prevention of Cancer** Trial - that individuals with low blood **selenium** levels could reduce their risk ...
<https://pubmed.ncbi.nlm.nih.gov/29376219/>

Potential Role of Selenium in the Treatment of Cancer an...

Feb 17, 2022 ... Several studies have shown that **selenium** has a tumor-promoting effect. The NPC trial, for example, found that **selenium** supplementation (as ...
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8879146/>

Selenium in cancer prevention: a review of the evidence ...

Interventions with Se have shown benefit in reducing the risk of **cancer** incidence and mortality in all **cancers** combined, and specifically in liver, prostate, ...
<https://pubmed.ncbi.nlm.nih.gov/16313696/>

Selenium | Memorial Sloan Kettering Cancer Center

Dec 15, 2021 ... For **Patients & Caregivers** - To **prevent** or treat **cancer**. Clear evidence is lacking. To reduce **cancer** treatment side effects. Preliminary studies ...
<https://www.mskcc.org/cancer-care/integrative-medicine/herbs/selenium>

Selenium - Health Professional Fact Sheet

Mar 26, 2021 ... Randomized controlled trials of **selenium** supplementation for **cancer prevention** have yielded conflicting results. The authors of a Cochrane ...
<https://ods.od.nih.gov/factsheets/Selenium-HealthProfessional/>

Selenium and Vitamin E Cancer Prevention Trial (SELECT)....

Jul 7, 2015 ... There are no clinical trials that show a benefit from taking vitamin E or **selenium** to reduce the risk of prostate **cancer** or any other **cancer** or ...
<https://www.cancer.gov/types/prostate/research/select-trial-results-qa>

Selenium Exposure and Cancer Risk: an Updated Meta-analy...

Jan 20, 2016 ... The results indicated that high **selenium** exposure had a protective effect on **cancer** risk (pooled OR = 0.78; 95%CI: 0.73–0.83). The results of ...
<https://www.nature.com/articles/srep19213>

Instructions: Use the search engine to help you determine the effectiveness of the specified treatment for the corresponding health issue.

Category	Explanation
Helpful	The treatment is effective and has a direct positive influence on the specified health issue.
Inconclusive	If medical professionals are still unsure if the treatment will have a positive, negative, or no influence on the health issue.
Unhelpful	The treatment either has no effect or has a direct negative influence on the specified health issue.

When you have decided on the effectiveness of the treatment, please click on the button corresponding to your answer below:

Figure 3.2: Study Web Application’s SERP Interface.

Task Question: Can vitamin B12 and sun exposure together help treat vitiligo?

Treatment: Vitamin B12 is a vitamin essential for human health. Sun exposure means to expose the skin to sunlight.

Health Issue: Vitiligo is a condition where patches of skin lose their pigment.

Instructions: For this task, you are to judge the effectiveness of the treatment without the aid of a search engine.

Category	Explanation
Helpful	The treatment is effective and has a direct positive influence on the specified health issue.
Inconclusive	If medical professionals are still unsure if the treatment will have a positive, negative, or no influence on the health issue.
Unhelpful	The treatment either has no effect or has a direct negative influence on the specified health issue.

When you have decided on the effectiveness of the treatment, please click on the button corresponding to your answer below:

Figure 3.3: Study Web Application’s without SERP Interface.

when people are seeking answers to their health-related questions. So, in these tasks, we just showed the task question, the treatment description, and the health issue description to the participants, and they had to judge the effectiveness of the treatment towards the mentioned health issue without the aid of a search engine. Figure 3.3 shows the study’s developed web application’s interface for a task that was presented to a participant without the search engine aid.

3.3.2 Within-Subjects Study Design

In order to assess the impact of different search engine results on participants’ ability to ascertain the efficacy of a proposed medical treatment according to a specific health issue, we conducted a within-subjects, controlled laboratory study. We used a 12x12 Graeco-Latin square to randomize both study variables: 1) the experiment conditions (interfaces), which were the different search engines plus the control conditions, and 2) the topics to have a fully balanced design. A Graeco-Latin square is a composite structure formed by combining two or more distinct Latin squares. So, each task is defined as a topic accompanied by an experiment condition, which can be either the control condition or one

of the search engines. As we wanted to balance the number of *helpful* and *unhelpful* tasks for each participant and throughout the whole study, we needed a 12x12 Graeco-Latin square which each of its rows was to be assigned to a particular participant.

In this design, each row has 12 different tasks in which there are six different *helpful* and six different *unhelpful* topics. Also, the three different experimental conditions are assigned to the 12 tasks equally, which means each experimental condition is assigned to exactly four tasks per row. The reason why we used a Graeco-Latin square was to have orthogonal Latin squares, one for assigning the topics and another for assigning the experiment conditions to each task and combining them over each other to have a study design without the. By using a randomization technique to arrange the rows and columns of a 12x12 Graeco-Latin square and thereafter allocating each row to a participant in the study, we have successfully achieved a fully balanced study design. There are two primary issues with this design if we did not use a Graeco-Latin square. The first is related to the order of the topics, while the second concerns the order of the experiment conditions (search engines and control) among the sequence of tasks assigned to a specific participant during the study. A Latin square might have been used as a method to mitigate the influence of one of these factors, but it was incapable of simultaneously controlling for both. A Graeco-Latin square is required in order to effectively manage both variables.

Moreover, in particular, participants' *learning* and *fatigue* might provide study results that can be ascribed to the experimental design itself rather than the specific experimental condition being studied. As participants go through each topic, their understanding of the task scenario and the experimental conditions deepens, which could impact the study results in the next tasks; this is called the *learning* effect. Also, at the start of the study, participants may exhibit higher levels of motivation and attentiveness compared to the end of the study. Thus, when the participants become *fatigued* throughout the study, as they go on, the results may be impacted by the participants' haste in performing the study tasks. When both the orders of topics and experiment conditions are rotated using

a Graeco-Latin square, it will effectively eliminate the potential biases arising from both *learning* and *fatigue* factors (Kelly et al., 2009).

Initially, as a rational first attempt, we intended to build the 12x12 Graeco-Latin square by creating four 6x6 Graeco-Latin squares and combining them to build a 12x12 Graeco-Latin square. As a consequence of the fact that there are no two 6x6 Latin squares that are orthogonal and a Graeco-Latin square should be built using two orthogonal Latin squares on top of each other, there is no 6x6 Graeco-Latin square (Cochran and Cox, 1948). Thus, we could not proceed with the mentioned design, and we had to come up with another idea.

To create a 12x12 Graeco-Latin square for this balanced study design, we divide both sets of *helpful* and *unhelpful* topics into two equal subsets of 3 topics. In 3.1, if we consider H_1 and H_2 as the *helpful* topics subsets and U_1 and U_2 as the *unhelpful* topics subsets that are mentioned above, H and U will be the whole sets of *helpful* and *unhelpful* topics we had selected for the study, respectively. Also, in 3.1, h_i is a *helpful* and u_i is an *unhelpful* topic.

$$\begin{aligned}
 H_1 &= \{h_1, h_2, h_3\} & U_1 &= \{u_1, u_2, u_3\} \\
 H_2 &= \{h_4, h_5, h_6\} & U_2 &= \{u_4, u_5, u_6\} \\
 H &= H_1 \cup H_2 & U &= U_1 \cup U_2
 \end{aligned}
 \tag{3.1}$$

We built the 12x12 Graeco-Latin square by combining sixteen 3x3 Graeco-Latin squares shown in Figure 3.4. In each 3x3 square, we used only three topics to create the Graeco-Latin square with the three different experimental conditions (interfaces).

For example, the 3x3 square with the H_1 label in Figure 3.4 is showing a 3x3 Graeco-Latin square created with the topics in the subset $H_1 = \{h_1, h_2, h_3\}$ that was mentioned in 3.1 with all three experimental conditions: 1) Bing, 2) Google and 3) Control, in which no

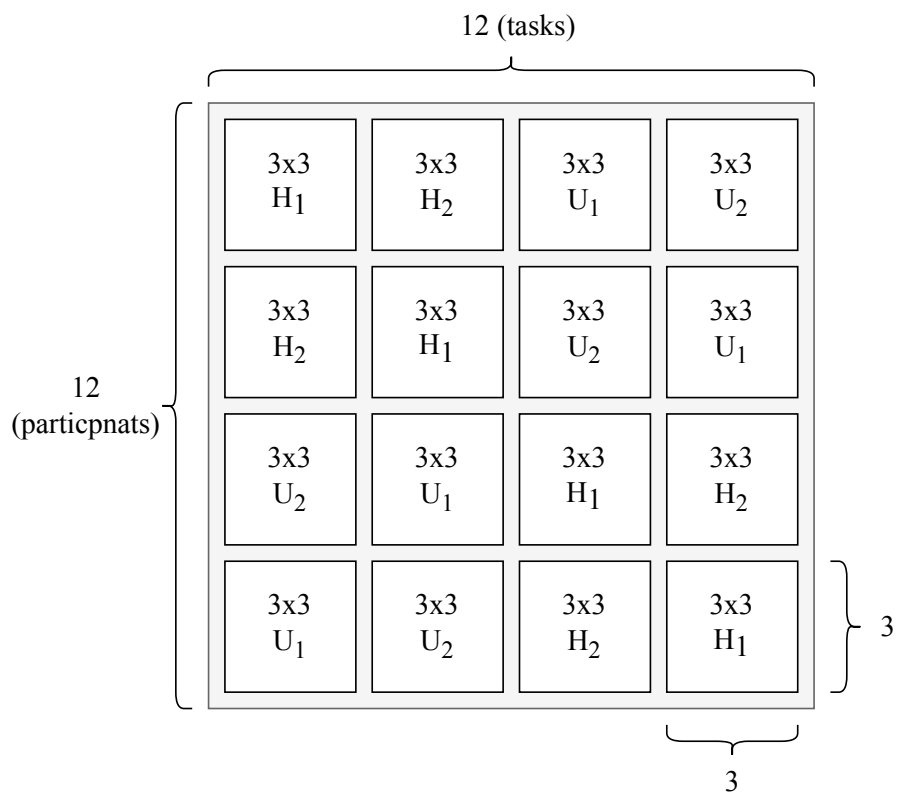


Figure 3.4: 12x12 Graeco-Latin square block created using sixteen 3x3 squares.

search engine is provided to the participant for the task. This implies, in this square each one of the three experimental conditions will appear in only one row and one column. Also, each member of the H_1 set appears only once in each row and once in each column of this 3x3 square. This design will guarantee the equal appearance of each experiment condition in each row and each column in the whole 12x12 square. Additionally, as each topic is a member of exactly one of the H_1 , H_2 , U_1 or U_2 sets, in the 12x12 final square, each topic of the whole 12 topics appears exactly once per row and once per column. Finally, to remove any aforementioned bias in this within-subjects study design, we randomize all rows and columns and assign each row to each participant of the study. For example 3.2 shows a possible task sequence assignment to participant # i (P_i) as one of the rows of the final 12x12 Graeco-Latin block, in which each pair of (t_i, I_j) is a study task presented with topic t_i that $t_i \in H \cup U$ (defined in 3.1) and providing the experiment interface I_j that $I_j \in \{Bing, Google, Control\}$.

$$\begin{aligned}
 P_i : &<(h_1, I_1), (h_2, I_2), (h_3, I_3), \\
 &(h_4, I_1), (h_5, I_2), (h_6, I_3), \\
 &(u_1, I_1), (u_2, I_2), (u_3, I_3), \\
 &(u_4, I_1), (u_5, I_2), (u_6, I_3) >
 \end{aligned}
 \tag{3.2}$$

3.4 Participants

The pilot study for the testing purposes phase included the involvement of my supervisor and two other graduate students as participants. After receiving clearance from the University of Waterloo’s ethics committee, we recruited volunteers by posting flyers around campus and sending email notifications to several graduate student email groups. Out of a total of 176 people who filled out our screening registration Google Form, 162 people

were eligible to participate in the study based on the screening criteria that are mentioned earlier in section 3.3. After that, we invited the eligible volunteers in batches of 20 people at a time, so we did not exceed the intended 60 participants we required. We aimed to recruit 60 participants so we could fill out five full blocks of 12x12 Graeco-Latin squares mentioned in subsection 3.3.2. In the invitation, we asked the participants to consent to participate in another Google Form, which was given by all of the participants after being fully informed about the study in the same form description. Then, each participant was scheduled for an online participation at a particular 1.5-hour time-slot. Each participant was to be remunerated with an Amazon.ca e-gift card with a value equal to C\$30 or a prorated amount of C\$10 per 30 minutes of study participation to a maximum of C\$30 if they did not finish the study. We recruited 60 participants, of which 27 were men (45%), 32 were women (53.3%), and one was non-binary (1.6%). The participants were between the ages of 18 and 61 (10% 20 and under 20, 41.6% 21-25, 31.6% 26-30, 10% 31-40, 1.3% over 40), and 88% of the participants were university students, which 33.3% of them were either Mathematics or Computer Science, 38.3% Engineering, 8% Arts, and 5% were Science students.

3.5 Data Cleaning

After a few rounds of participation, we conducted a preliminary analysis of the data collected from the participants. It was observed that a subset of the participants did not generate any queries during the search tasks and instead just clicked on one of the treatment effectiveness decisions. In addition, a few other participants opted to forgo formulating their own queries and instead turned to just copying the task question and pasting it into the search bar as their search query. Given that both activities were deemed not ideal for our subsequent study, we opted to implement a few modifications to the web user in-

terface. Additionally, we included one-on-one online interviews held via Zoom to enhance the new participants' engagement and their performance on the tasks and substituted new participants with the participants we detected those problems in their logged data. In this course of refining study data, a total of 38 participants' data had to be excluded, necessitating their replacement with new participants. Therefore, we came up with a criterion to evaluate the participant's performance on the two unwanted behaviors. Based on the criterion, we replaced all participants which had fewer than 6 out of 8 search tasks in which they had issued at least one query that was not an exact copy of the task question. As a result, we were able to get a complete dataset consisting of 60 participants without the issues explained above. This dataset was used to populate five blocks of 12x12 squares, as previously described in [3.3](#).

In order to enhance the online user interface (UI) to make it more resilient to the issues mentioned above, we removed the decision-making buttons prior to the participant issuing at least one query in a search task. This way, participants could not go to the next stage of the study without issuing at least one query. Additionally, we disabled the text copy functionality on the parts of the web app in which we had the task defined with the task question and health issue and treatment definitions at the top of the page ([Figure 3.2](#)), which helped mitigate issuing the same task question as the search query by participants.

Chapter 4

Results and Discussion

In this chapter we present the findings of the study mentioned in Chapter 3, which aim to illustrate the various measures examined in the study participants' interactions. These measures are used to assess the effectiveness of different treatments, both with and without the utilization of a modern search engine. Subsequently, further investigation is conducted to explore the participants' prior knowledge of health issues and treatments, as well as their level of confidence in their final decisions.

4.1 Search Engine Effect on Decision Making

The primary findings of our study focuses on the impact of the modern search engines on the participants' ability to accurately assess the effectiveness of the 12 medical treatments mentioned in Section 3.3.2. The participants are presented with two conditions: one where they engage with a search engine (referred to as the experimental condition) to aid in their response to the question, and another where they are instructed to provide an answer to

the question without having access to a search engine (referred to as the control condition). In the experimental settings, a contemporary user interface for a search engine has been constructed, using the APIs offered by the two prominent and widely used search engine services, namely Google and Bing. Therefore, half of the topics (4 of them) assigned to the experimental condition in the balanced design described in Section 3.3.2 had the Google and the other half had the Bing search engine back-end API with the same user interface design of the search engine page presented to the participants.

In order to assess the effect of experimental conditions compared to the control condition, first, our research focuses on comparing the study outcomes based on participants' interactions with or without the use of a search engine to assess the effectiveness of treatments. As mentioned in Section 3.2, the topics we have used for the study derived from the TREC 2021 health misinformation track, have a stance field that indicates the truth about the effectiveness of the suggested treatment for the corresponding health issue. Based on the stance field of each topic, two key metrics that provide valuable insights in this regard are the average accuracy and harmfulness of participants' decisions. Table 4.1 shows the confusion matrices of study participants' decisions over the actual truth and also average accuracy and harmfulness metrics for control and experimental conditions respectively. In a broad sense, among the three levels of options every participant had during the study to decide about a treatment's effectiveness: *helpful*, *inconclusive* and *unhelpful*, a decision is deemed accurate when the participant's determination of the efficacy of the treatment aligns with the stance field established on the topic that has one of the helpful and unhelpful values. Therefore, the determination of an *inconclusive* decision has been considered as an inaccurate decision in our calculation of each study setting's 3-level accuracy. Moreover, a decision is considered to be harmful, when the participant has exactly chosen the opposite decision option to say when a participant judged a treatment as *unhelpful* while the truth stance label of a topic was helpful and vice versa. This way, a decision indicating a treatment's effectiveness to be *inconclusive* is not a harmful decision. So, the average

Table 4.1: Confusion matrices with accuracy and harmfulness of participants’ decisions of the control and experimental conditions.

Control Condition (no search engine)						
Truth	Participant Decision			Total	Accuracy	Harmfulness
	Helpful	Inconclusive	Unhelpful			
Helpful	59	37	24	120	0.49	0.20
Unhelpful	34	51	35	120	0.29	0.28
Total	93	88	59	240	0.39	0.24

Experimental Conditions (interaction with a search engine)						
Truth	Participant Decision			Total	Accuracy	Harmfulness
	Helpful	Inconclusive	Unhelpful			
Helpful	203	35	2	240	0.85	0.01
Unhelpful	62	98	80	240	0.33	0.26
Total	265	133	82	480	0.59	0.13

3-level accuracy and harmfulness mentioned in the Table 4.1 and later on in this chapter are calculated based on the metric definitions mentioned above.

In Table 4.1, we observe that, the total average accuracy increases from 39% to 59% as people use a search engine to determine the treatments effectiveness compared to when they do not. Moreover, the mean overall harmfulness of the experimental conditions is lower compared to the control condition. In general, the utilization of contemporary search engines to address health-related questions enables individuals to enhance the accuracy of their assessments and make less harmful decisions. Additionally, it is seen that when the topic is one of the helpful ones, the use of a search engine leads to a significant reduction in harmfulness, decreasing from 0.20 to 0.01. Furthermore, the total accuracy of the search

engine improves significantly, increasing from 49% to 85%. On the contrary, it is evident that the use of search engines proved to be ineffective in situations when the topic was one of the unhelpful ones. This suggests that search engine algorithms have a proclivity to direct users towards results in favor of the effectiveness of treatments, which is consistent with prior research in the field (Kayhan, 2013; White, 2013; Bondarenko et al., 2021; White and Hassan, 2014).

Another notable result is that participants had a higher tendency to choose the *inconclusive* choice when the treatment’s stance was unhelpful in both the control and experimental conditions. It is also observed that in the experimental conditions, compared to the control condition, although the total number of helpful topics is doubled, the number of *inconclusive* choices made by participants remained nearly the same. This indicates that the inclination to deem a helpful topic’s treatment as *inconclusive* is reduced by half when users utilize a search engine compared to when they do not. Added to this information, when considering the accuracy metric results in both conditions it can be inferred that search engines not only helped users in determining the correct answer but also enhanced their confidence in their decision-making process, thereby reducing confusion and minimizing the selection of the *inconclusive* option when they were asked to judge helpful topics. On the other hand, while using a search engine, participants were almost three times more likely to display a lack of confidence and choose *inconclusive* when judging an unhelpful topic. In Section 4.4, we will provide further elaboration on the participants’ levels of confidence across various circumstances.

In order to align the participants’ decisions and also make them more comparable with the topics’ binary helpful and unhelpful stances, we included a Yes/No question into the post-task questionnaire. This allowed us to examine what the participants’ final decisions would have been if they had not been presented with the *inconclusive* alternative. To fulfill this purpose, for each topic, we asked the topic question mentioned in Table 3.1, only with “Yes” and “No” options, to understand that if the participant had to choose between these

Table 4.2: Confusion matrices of participants’ y/n responses to the task questions in control and experimental conditions.

Control Condition (no search engine)					
Truth	Response		Total	Accuracy	Harmfulness
	Yes	No			
Helpful	66	54	120	0.55	0.45
Unhelpful	39	81	120	0.68	0.33
Total	105	135	240	0.61	0.39

Experimental Conditions (interaction with a search engine)					
Truth	Response		Total	Accuracy	Harmfulness
	Yes	No			
Helpful	219	21	240	0.91	0.09
Unhelpful	91	149	240	0.62	0.38
Total	310	170	480	0.77	0.23

options, what their final decision would have been. The confusion matrices for “Yes” and “No” responses on helpful and unhelpful topics in both experimental and control conditions are shown in Table 4.2. The table presents data on binary accuracy, which is the proportion of instances in which participants accurately determined the effectiveness of the treatment based on their answers to the Yes/No question. Moreover, the harmfulness metric in this scenario pertains to the instances when the study participants were unable to accurately ascertain the helpfulness of the treatment and responded in opposite to the ground truth.

Upon examination of Table 4.2, it is evident that the results for the helpful topics align with

Table 4.3: Final decisions made by participants for *Inconclusive* judged topics.

Condition	Topic Stance	Topic Question Response	
		Yes	No
Control	Helpful	7 (19%)	30 (81%)
Control	Unhelpful	4 (8%)	47 (92%)
Engine	Helpful	19 (54%)	16 (46%)
Engine	Unhelpful	32 (33%)	66 (67%)
All	Helpful	26 (36%)	46 (64%)
All	Unhelpful	36 (24%)	113 (76%)

the findings shown in Table 4.1. This suggests that search engines play a significant role in assisting users in finding the necessary search results to substantiate the effectiveness of helpful treatments. On the contrary, it is observed that when it comes to unhelpful topics, the use of search engines has a tendency to attract users, leading them to believe that the treatment is beneficial for the associated health issue. We also traced the *inconclusive* judged topics by study participants and saw that among control tasks 88%, search tasks 62% and in total 72% of them, finally turned to a “No” answer to the task question in the post-task questionnaire. Thus, it can be inferred that most of the *inconclusive* decisions which are made in confusion, turn out to be a negative response to the treatments’ effectiveness in participant’s opinion, but when people use search engines, this effect gets less predominant compared to when they do not. By looking at Table 4.3, we observe that confused *inconclusive* decisions in the control condition are turned into negative topic question responses regardless of actual stance of the topic. On the other hand though, for the search engine tasks, *inconclusive* choices have comparably turned into more accurate responses by the participants. This can imply that although most of the confusions end up to be considered as a “No” decision by the users, search engines drive people to select

“Yes”, thus the overall accuracy of making the correct decision about the helpful topics, after selecting *inconclusive*, increases. This also may show that users look to be more confident about their decisions when using a search engine, so they make less conservative choices. We will include more details about the participants prior knowledge and decision confidence in Section 4.4.

4.2 Topic-by-topic Analysis

Given that the participants’ performance in the experimental conditions was comparatively poorer for unhelpful topics as opposed to helpful ones, we deemed it necessary to conduct a deeper analysis of the data on a topic-by-topic basis. Table 4.4 shows the 3-level accuracy and harmfulness metrics per topic for both control and search engine conditions. The table demonstrates that, along with the findings discussed earlier in this chapter, the use of a search engine has yielded positive outcomes in terms of enhancing accuracy and reducing harmfulness metrics for the helpful topics. This improvement is seen when comparing the results to those obtained from the same topics in control condition. Additionally, it is evident in the table that, search engine usage is not necessarily detrimental for all unhelpful topics. In order to get a more precise evaluation of the participants’ performance in each topic, we have collected the replies to the topic questions from the post-task questionnaire. These responses have been organized in a confusion matrix, which includes the accuracy and harmfulness metric findings for each topic. The specific details of this matrix can be found in Table 4.5. This table also reveals that people exhibited reduced accuracy in their decisions of the effectiveness of treatments for similar unhelpful topics while using a search engine, as compared to the people responded to the same question in control condition. In the following discussion, using the data provided in Table 4.4 and Table 4.5, for each

unhelpful topic, we will comprehensively analyze the impact of search engines on the study participants' decision-making processes.

For topic 101, after we reviewed some of the logs of the search results that were shown to the participants for the topic during the study, it was evident that the search engines had brought a single result to the top of the [SERP](#) list that was an article that unequivocally asserts that braces are ineffective in the treatment of Achilles tendonitis. That particular article likely played a crucial role in enabling the participants to arrive at an accurate decision. In both topics 103 and 128, after comparing the control and engine conditions data, it is evident that the search engines had effectively persuaded people to overcome their inherent skepticism and get convinced that the treatments are beneficial. The participants had initial doubt about the effectiveness of these treatments in addressing the respective health concerns, however, it seems that their perspectives were subsequently influenced upon seeing the search engine results. In relation to topic 109, it is believed that due to the frequent mention of cancer here and there, the participants who were engaged in the topic 109's task under the control condition expressed that if selenium has potential as a remedy or intervention for cancer, they would have encountered information on it on at least one occasion prior to the study. Therefore, the majority of responses in the control condition were negative, as shown by the examination of [Table 4.5](#), which represents the correct answer. Upon analyzing the data presented in [Table 4.4](#) for topic 112, it can be inferred that the using search engines did not yield any significant benefits. However, a closer examination of the final "Yes" and "No" questions in the post-task questionnaire, as illustrated in [Table 4.5](#), reveals that search engines has indeed proven to be advantageous for this particular unhelpful labeled topic. In this instance, upon reviewing the search result logs of the participants, it was observed that there were scholarly articles suggesting the lack of efficacy of the treatment. However, there were also a few documents among the top results that seemed likely to cause confusion among the participants, initially leading to an *inconclusive* decision but ultimately resulting in a "No" response to the topic

question. In topic 134 also, it was observed that while the participants' performance was satisfactory in the control condition, the utilization of search engines played a significant role in enabling users to determine the correct answer. This was due to the fact that search results prominently featured articles that contradicted the effectiveness of Vaseline for tick removal.

4.3 Search Engine Interactions

In our research, alongside recording participants' decisions and responses to the questionnaires, we also observed their interactions comprehensively. This included monitoring their decision times, queries entered into the search bar, clicks on search results, and engagement with [SERP](#) pages. This approach aimed to enhance our understanding of user behavior and interactions with or without a search engine when seeking information to address health-related questions.

4.3.1 Clicks

The quantification of user interactions with a search engine is a crucial aspect of analysis, with the number of clicks on search results contributing as a significant metric. In order to track the frequency of these click events, we tracked the individual clicks made by each user on any search result for each task. The average number of clicks per task by each participant compared in different circumstances is presented in [Table 4.6](#). Occasionally, users may click on a certain search result several times. In order to accurately determine the number of separate search results that were clicked on, we have recorded these repeated clicks as a

Table 4.4: Search engine helpfulness analysis compared to the control condition per topic for participants' 3 level decisions (H: helpful, I: inconclusive, U: unhelpful).

Topic #	Subject	Stance	Type	Decision			Acc	Harm	Engine Effect
				H	I	U			
101	ankle brace achilles tendonitis	unhelpful	control engines	15	3	2	0.10	0.75	helps
103	folic acid dementia	unhelpful	control engines	10	11	19	0.48	0.25	
109	selenium cancer	unhelpful	control engines	2	9	9	0.45	0.10	harms
112	evening primrose oil eczema	unhelpful	control engines	18	21	1	0.03	0.45	
128	steam shower croup	unhelpful	control engines	1	14	5	0.25	0.05	no effect
134	remove tick with vaseline	unhelpful	control engines	6	24	10	0.25	0.15	
107	yoga asthma	helpful	control engines	5	13	2	0.10	0.25	no/little effect
115	magnesium migraine prevention	helpful	control engines	5	28	7	0.18	0.13	
118	dupixent eczema	helpful	control engines	7	6	7	0.35	0.35	harms
136	dates iron deficiency anemia	helpful	control engines	21	11	8	0.20	0.53	
140	hip osteoarthritis at-home exercises	helpful	control engines	4	6	10	0.50	0.20	helps
149	fungus on feet	helpful	control engines	2	3	35	0.88	0.05	
			control engines	7	5	8	0.35	0.40	helps
			control engines	24	16	0	0.60	0.00	helps
			control engines	7	10	3	0.35	0.15	helps
			control engines	31	9	0	0.78	0.00	helps
			control engines	5	10	5	0.25	0.25	helps
			control engines	38	2	0	0.95	0.00	helps
			control engines	8	6	6	0.40	0.30	helps
			control engines	34	5	1	0.85	0.03	helps
			control engines	17	3	0	0.85	0.00	helps
			control engines	40	0	0	1.00	0.00	helps
			control engines	15	3	2	0.75	0.10	helps
			control engines	36	3	1	0.90	0.03	helps

Table 4.5: Search engine helpfulness analysis compared to the control condition per topic for participants' responses to the task question.

Topic #	Subject	Stance	Type	Response		Acc	Harm	Engine Effect
				No	Yes			
101	ankle brace achilles tendonitis	unhelpful	engines control	26	14	0.65	0.35	helps
103	folic acid dementia	unhelpful	engines control	14	26	0.35	0.65	harms
109	selenium cancer	unhelpful	engines control	29	11	0.73	0.28	harms
112	evening primrose oil eczema	unhelpful	engines control	19	1	0.95	0.05	harms
128	steam shower croup	unhelpful	engines control	30	10	0.75	0.25	helps
134	remove tick with vaseline	unhelpful	engines control	12	8	0.60	0.40	helps
107	yoga asthma	helpful	engines control	14	26	0.35	0.65	harms
115	magnesium migraine prevention	helpful	engines control	13	7	0.65	0.35	harms
118	dupixent eczema	helpful	engines control	36	4	0.90	0.10	helps
136	dates iron deficiency anemia	helpful	engines control	14	6	0.70	0.30	helps
140	hip osteoarthritis at-home exercises	helpful	engines control	8	32	0.80	0.20	helps
149	fungus on toenails	helpful	engines control	10	10	0.50	0.50	helps
			engines control	2	38	0.95	0.05	helps
			engines control	13	7	0.35	0.65	helps
			engines control	2	38	0.95	0.05	helps
			engines control	13	7	0.35	0.65	helps
			engines control	5	35	0.88	0.12	helps
			engines control	11	9	0.45	0.55	helps
			engines control	0	40	1.00	0.00	helps
			engines control	2	18	0.90	0.10	helps
			engines control	4	36	0.90	0.10	helps
			engines control	5	15	0.75	0.25	helps

Table 4.6: Average number of total and unique clicks per-task based on decision accuracy.

Dependent Variables	Clicks	Unique Clicks
Correct Decisions (3-level)	2.21	2.13
Inconclusive Decisions	2.69	2.55
Harmed Decisions	1.89	1.86
Unharmed Decisions	2.37	2.27

single click in the unique clicks column of the table. The table displays the dependent variable for each row, indicating the filter applied to pairings of (task, participant) in which the criterion of the dependent variable is satisfied. For example, in the row of correct decisions (3-level), the data displays the frequency of clicks made by participants. It is important to note that decisions classified as *inconclusive* are excluded from this count and mentioned in the next row. A 3-level correct decision is when the participant has chosen the exact stance of the topic as their decision. However, a harmed decision occurs when the person deliberately chose the very opposite option of the correct topic's stance on the given task. Based on this definition, an unharmed decision is taken when the participant's decision was either correct or *inconclusive*.

Based on the data shown in Table 4.6, it can be deduced that a lower quantity of clicks, namely less than 2 clicks during a search task, has resulted in participants making a harmful decision. In this scenario, their evaluation of the efficacy of a therapy was exactly opposed to the corresponding topic's stance. However, it should be noted that an increased amount of clicking has also resulted in users being perplexed and ultimately selecting the *inconclusive* alternative.

Furthermore, the data shown in Table 4.7 illustrates the cumulative count of clicks recorded across various search result ranks. The findings indicate that individuals have a proclivity

Table 4.7: Average number of total and unique clicks per search task on each SERP result rank.

Rank	Unique Clicks	Total Clicks
1	285	370
2	189	217
3	144	160
4	83	94
5	67	70
6	52	53
7	38	39
8	30	31
9	23	23
10	27	30
11+	18	18

for selecting higher-ranked search results, as seen by the fact that over 68% of the clicks made by the participants in the research were directed towards the top three search results.

4.3.2 Queries

In order to enhance comprehension of individuals' interactions with search engines while seeking health-related information, we systematically documented the queries entered by the study participants inside our built search engine environment. Based on our investigation of the query patterns shown by the study participants, it was found that they made efforts to use a diverse range of query types in order to get information that aligned

with their sought medical information with higher precision. The notes observed about each query characteristic is mentioned in the Table 4.8 with an example of a real query performed by a study participant for different types of observed queries. The complete list of queries per topic and the number of times that distinct participants have used them is attached to this thesis in Appendix A.

The average number of queries made by the study participants for each given topic is shown in Table 4.9. In contrast to the “Queries” column in the table, which displays the average number of search queries conducted by each user per topic, the “Unique Queries” column excludes redundant queries made by the same user for the same topic, ensuring that they are only counted once. Upon analysis of the provided table, it becomes apparent that some topics, including topics 136, 128, 101, and 103, required more queries in order to access articles that would aid participants in their final decision on the effectiveness of the respective treatment.

4.3.3 Decision Time

An additional significant aspect to consider when examining the user’s interactions with a search engine in this study is the decision time, which refers to the duration that the user spends on the search engine before locating the desired information. In order to conduct a comprehensive study of participants’ behavior in seeking health-related information, we meticulously logged all search engine interactions along with their corresponding timestamps. Consequently, we obtained an extensive set of timestamps encompassing the participant’s engagement with the task question, initiation of the search process, submission of each query, selection of search results by clicking on them, pagination over the search result

Table 4.8: Query type examples among different topics.

#	Topic Question	Query	Query Characteristics
101	Will wearing an ankle brace help heal Achilles tendonitis?	achilles tendonitis ankle brace	Keyword query
107	Does yoga improve the management of asthma?	Can yoga reduce symptoms of asthma condition?	Full sentence query
103	Can folic acid help improve cognition and treat dementia?	“folic acid” AND “dementia”	Boolean query
118	Can dupixent treat eczema?	what is eczema	Seeking more info about the health issue
109	Does selenium help prevent cancer?	What is selenium?	Seeking more info about the treatment
115	Can magnesium help prevent migraine?	A randomized prospective placebo-controlled study of intravenous magnesium sulphate vs. metoclopramide in the management of acute migraine attacks in the emergency department.	Text from previous search results as a new query.
149	Will at-home exercises manage hip osteoarthritis pain?	can calisthenics help with hip osteoarthritis	Similar or synonym terms e.g., calisthenics instead of at-home exercises.
118	Can dupixent treat eczema?	dupixent and eczema - mayo clinic	High quality source names e.g., Mayo Clinic
107	Does yoga improve the management of asthma?	is yoga helpful for asthma management from a medical standpoint	Search for medical studies

Table 4.9: Average number of unique and total queries made by each participant per topic.

Topic #	Subject	Queries	Unique Queries
101	ankle brace achilles tendonitis	1.93	1.70
103	folic acid dementia	1.85	1.75
107	yoga asthma	1.45	1.35
109	selenium cancer	1.40	1.33
112	evening primrose oil eczema	1.23	1.23
115	magnesium migraine prevention	1.60	1.40
118	dupixent eczema	1.70	1.45
128	steam shower croup	1.93	1.83
134	remove tick with vaseline	1.53	1.48
136	dates iron deficiency anemia	2.15	2.08
140	fungal cream athlete's foot	1.80	1.73
149	hip osteoarthritis at-home exercises	1.53	1.50
Overall		1.67	1.57

Table 4.10: Average time in seconds it took participants to make their decision with and without responding to the post-task questionnaire.

Dependent Variables	Avg. decision time w/o posttask (sec)	Avg. decision time w/ posttask (sec)
Correct Decisions (3 level)	131.2	138.5
Incorrect Decisions (3 level)	161.1	170.7
Unharmed Decisions	143.0	151.5
Harmed Decisions	146.6	152.9
Correct Decisions (y/n)	135.3	143.3
Incorrect Decisions (y/n)	170.4	179.5
Decision: Helpful	137.1	143.8
Decision: Inconclusive	168.1	179.3
Decision: Unhelpful	124.1	132.8

pages, decision submission, and completion of the post-task questionnaire. The average decision time in different situations base on the final decision of the participants, including and excluding the time allocated for the post-task questionnaire, is shown in table 4.10.

The data demonstrates that the mean response duration of post-task questionnaires across various states by the study participants, as shown in the “Dependent Variables” column, is around 8 seconds on average. So, currently, this column can be disregarded, and our attention is directed on the duration of time between the completion of the pre-task questionnaire and when the participants start their interaction with the embedded search engine and the submission of the participants’ 3-level choice, prior to the administration of the post-task questionnaire. As previously stated, a 3-level incorrect decision might manifest as an *inconclusive* option made by the participant. However, a harmed decision occurs when

the person deliberately chose the very opposite option of the correct topic's stance on the given task. Based on this definition, an unharmed decision is taken when the participant's decision was either correct or *inconclusive*. By examining the data in Table 4.10 we observe that, not only taking more time spend on each task in the search engine page has led the participants to come up with incorrect both 3-level and "Yes" and "No" answers, but also their decisions were more harmful. This suggests that spending additional time deliberating on the effectiveness of a treatment through the use of a search engine could potentially lead the user to make a final incorrect decision. This is because the act of reading various search results may result from the user's confusion in determining the accurate stance of the topic among the displayed search results. Thus, it is also observed that participants take longer to reach a decision of *inconclusive* than to decide a treatment's effectiveness to be *helpful* or *unhelpful*.

Moreover, in order to enhance our understanding of user behavior throughout the search for treatment efficacy, we have collected data on decision time for each topic, which is shown in Table 4.11.

In the present table, after a comparison made between the findings of Table 4.9 and the observed relationship between the number of queries generated by participants and the corresponding amount of time spent on those topics, it became evident that topics requiring a greater number of queries also required a longer duration of engagement with the search engine environment. However, in addition to the aforementioned challenging topics discussed in Section 4.3.2, topic 109 also has a controversial nature, as seen by the fact that participants required almost three minutes of searching before reaching out to a conclusive decision.

Table 4.11: Average time in seconds it took participants to make their decision per search task with and without responding to the post-task questionnaire.

Topic #	Subject	Avg. decision time w/o posttask (sec)	Avg. decision time w/ posttask (sec)
101	ankle brace achilles tendonitis	166.7	175.4
103	folic acid dementia	189.2	197.3
107	yoga asthma	125.1	133.2
109	selenium cancer	172.7	182.8
112	evening primrose oil eczema	122.7	131.1
115	magnesium migraine prevention	143.6	151.6
118	dupixent eczema	117.5	124.5
128	steam shower croup	144.5	153.5
134	remove tick with vaseline	107.0	116.4
136	dates iron deficiency anemia	167.1	175.9
140	fungal cream athlete's foot	131.5	137.6
149	hip osteoarthritis at-home exercises	134.1	141.5
Overall		143.5	151.7

4.4 Confidence and Prior Knowledge

Prior to presenting participants with any search tasks pertaining to a specific medical treatment, in the pre-task questionnaire, we conducted individual inquiries on their prior knowledge of the health issue and the corresponding medical treatment. Subsequently, after the submission of decisions by the participants, they were sent to a post-task questionnaire wherein they were asked about their level of confidence in their decision about the treatment's effectiveness. This section will discuss the results obtained from the analysis conducted on the participants' prior level of knowledge of both health issues and suggested treatments and posterior level of confidence in their decisions, as well as their potential association with the accuracy of their decision-making.

4.4.1 Prior Knowledge

In advance of engaging in the search activity, participants were instructed to assess their level of knowledge, first with the treatment and then with its corresponding health issue using a 5-point rating scale with the following options:

1. **Nothing:** No prior familiarity at all with the health issue or the treatment.
2. **Heard:** Small prior knowledge. The participant has somehow just heard the names of the health issue or the treatment.
3. **General:** Basic knowledge. The participant has general knowledge about what the health issue or the treatment is referring to.
4. **Familiar:** More than general information. The participant knows at least one effective use of the treatment or possible effects of the health issue on the body.

5. **Extensive:** A profound understanding of the health issue or the treatment. The participants knows possible treatments for the health issue or knows almost all possible effects of the treatment on the human body.

For each task, participants were required to answer two prior knowledge questions in a pre-task questionnaire, as displayed in Figure 4.1.

Health Issue Knowledge

Upon recording all the replies from the pre-task questionnaire, we proceeded to collect and combine the participants' decision accuracy data with their previous knowledge. Our objective was to examine any potential association between their prior knowledge and their decision performance. Table 4.12 presents the quantitative data on the distribution of knowledge levels of health issue declared by the participants across various scenarios. No statistically significant link was seen between the participants' self-declared levels of previous knowledge about health issues and their 3-level decision accuracy or choices. But in contrast, for the participants decision's based on harmfulness and Yes/No topic question accuracy, we could observe a significant correlation between them and participants' prior health issue knowledge. To make this correlation in the participants' decision accuracy, more visible, we grouped Extensive and Familiar knowledge categories as *high* and General, Heard and Nothing knowledge categories as *low* knowledge levels. The percentage of participants who their decisions were unharmed with a *high* level of health issue knowledge is 30% compared to the harmful decisions which is 19%. Also, for the same comparison in Yes/No question accuracy, participants with a *high* level of prior knowledge in health issues were contributing to 29% of correct decisions and 24% of wrong decisions. Upon the analysis of Table 4.12, it becomes evident that the distribution of participants' previous knowledge on the health issue is consistent throughout both control and search tasks,

Task Question: Does selenium help prevent cancer?

Treatment: Selenium is a mineral essential for human health.

Health Issue: Cancer is a disease where the cells in the body start to grow and spread uncontrollably.

Pre-task Questionnaire:

1. How much do you know about "selenium"?

- Nothing
- Heard of the treatment
- Know generally about the treatment
- Quite familiar with the treatment
- Know extensive details about the treatment

2. How much do you know about "cancer prevention"?

- Nothing
- Heard of the health issue
- Know generally about the health issue
- Quite familiar with the health issue
- Know extensive details about the health issue

SUBMIT

Figure 4.1: Study Web Application's Pre-Task Questionnaire Example.

regardless of the different states of the dependent variables. Therefore, it can be inferred that there is no substantial association between participants' prior knowledge of health issues and their 3-level decision accuracy in both control and search engine tasks, but for both control and experimental conditions, the decision's harmfulness and also the topic question's Yes/No question accuracy was correlated to the participants prior knowledge of the health issues.

In order to get a more comprehensive understanding of the average self-declared knowledge of health issues in various situations among the participants, we used a weighted average approach. This included giving specific weights to each degree of knowledge, namely: Nothing (0), Heard (1), General (2), Familiar (3), and Extensive (4). Table 4.13 shows the average health issue knowledge scores in different dependent variable cases. The table shown provides a clearer indication that when people used a search engine, there is no noticeable disparity in knowledge levels between individuals who made 3-level correct decisions. Conversely, there is a significant disparity in the average scores between decisions that were harmful and those that were unhelpful, as well as between correct and incorrect replies to Yes/No questions. People with more prior knowledge of health issues were more likely to choose the *inconclusive* option rather than the *unhelpful* option when using a search engine, compared to when they did not use a search engine.

Treatment Knowledge

The distribution of knowledge levels on treatment stated by the participants across different dependent variable scenarios is shown in Table 4.14. In contrast to what mentioned for the prior health issue knowledge, it is apparent that participant's prior treatment knowledge

Table 4.12: Participants' prior knowledge of the health issue for search and control tasks.

Dependent Variables	Prior Health Issue Knowledge (Exp. Conditions)				
	Extensive	Familiar	General	Heard	Nothing
Correct Decisions (3 level)	20 (7%)	63 (22%)	92 (33%)	72 (25%)	36 (13%)
Incorrect Decisions (3 level)	15 (8%)	37 (19%)	69 (35%)	50 (25%)	26 (13%)
Unharmful Decisions	33 (8%)	90 (22%)	137 (33%)	109 (26%)	47 (11%)
Harmful Decisions	2 (3%)	10 (16%)	24 (38%)	13 (20%)	15 (23%)
Correct Decisions (y/n)	29 (8%)	79 (21%)	120 (33%)	98 (27%)	42 (11%)
Incorrect Decisions (y/n)	6 (5%)	21 (19%)	41 (37%)	24 (21%)	20 (18%)
Decision: Helpful	19 (7%)	61 (23%)	94 (35%)	56 (21%)	35 (13%)
Decision: Inconclusive	13 (10%)	27 (20%)	45 (34%)	37 (28%)	11 (8%)
Decision: Unhelpful	3 (4%)	12 (15%)	22 (27%)	29 (35%)	16 (20%)
Overall	35 (7%)	100 (21%)	161 (34%)	122 (25%)	62 (13%)

Dependent Variables	Prior Health Issue Knowledge (Ctrl. Condition)				
	Extensive	Familiar	General	Heard	Nothing
Correct Decisions (3 level)	6 (6%)	23 (24%)	35 (37%)	21 (22%)	9 (10%)
Incorrect Decisions (3 level)	3 (2%)	30 (21%)	51 (35%)	38 (26%)	24 (16%)
Unharmful Decisions	8 (4%)	43 (24%)	62 (34%)	46 (25%)	23 (13%)
Harmful Decisions	1 (2%)	10 (17%)	24 (41%)	13 (22%)	10 (17%)
Correct Decisions (y/n)	8 (5%)	33 (22%)	53 (36%)	35 (24%)	18 (12%)
Incorrect Decisions (y/n)	1 (1%)	20 (22%)	33 (35%)	24 (26%)	15 (16%)
Decision: Helpful	5 (5%)	19 (20%)	39 (42%)	21 (23%)	9 (10%)
Decision: Inconclusive	2 (2%)	20 (23%)	27 (31%)	25 (28%)	14 (16%)
Decision: Unhelpful	2 (3%)	14 (24%)	20 (34%)	13 (22%)	10 (17%)
Overall	9 (4%)	53 (22%)	86 (36%)	59 (25%)	33 (14%)

Table 4.13: Participants' average prior health issue knowledge score for control and search engine tasks. The weighted average scores are calculated by assigning following weights to each knowledge level: Nothing=0, Heard=1, General=2, Familiar=3, Extensive=4.

Dependent Variables	Avg. Health Issue Knowledge Score	
	Control	Engine
Correct Decisions (3 level)	1.96	1.86
Incorrect Decisions (3 level)	1.66	1.82
Unharmful Decisions	1.82	1.89
Harmful Decisions	1.64	1.55
Correct Decisions (y/n)	1.85	1.88
Incorrect Decisions (y/n)	1.66	1.72
Decision: Helpful	1.89	1.90
Decision: Inconclusive	1.67	1.95
Decision: Unhelpful	1.75	1.48
Overall	1.78	1.84

has a significant correlation with both chosen options and decision-making accuracy. This way, for the search tasks, 34% of the participants who made correct 3-level decisions had a *high* degree of prior knowledge, but 19% of participants with the same level of knowledge made incorrect 3-level decision. The same conclusions can be inferred from both the dependent variables of harmfulness and Yes/No decision accuracy. As such, 29% of unharmed decisions and 22% of harmed decisions are made by participants with a *high* level of prior treatment knowledge. The percentages for the “Yes” and “No” topic question accuracy is also 30% compared to 22%. Table 4.15 shows the average prior treatment knowledge scores in different dependent variable cases. The displayed table makes it clearer that participants with greater levels of prior knowledge of the treatment were more likely to make the correct choice regarding its efficacy on all 3-level scale decisions, Yes/No topic question responses, and also in terms of harmfulness decisions when using a search engine. It is noteworthy that there exists a notable discrepancy in the mean scores seen between harmed and unharmed decisions within the control condition. Specifically, when participants with a lower level of prior knowledge of the treatments who did not have access to a search engine, have shown an affinity for selecting the *inconclusive* option which is considered as an unharmed decision, thus decreased the average knowledge score for unharmed decisions.

4.4.2 Confidence

Following the completion of each task, we administered a post-task questionnaire to all participants, consisting of two inquiries: 1) How certain they were in their decision about the treatment’s effectiveness 2) The topic Yes/No question. An example of a post-task questionnaire is illustrated in Figure 4.2. Based on participants’ self-declared certainty in their decisions expressed in the first question of the post-task questionnaire, we analyzed the participants’ confidence in different situations. In the question participants were instructed

Table 4.14: Participants' prior knowledge of the treatment for search and control tasks.

Dependent Variables	Prior Treatment Knowledge (Exp. Conditions)				
	Extensive	Familiar	General	Heard	Nothing
Correct Decisions (3 level)	29 (10%)	69 (24%)	64 (23%)	52 (18%)	69 (24%)
Incorrect Decisions (3 level)	8 (4%)	30 (15%)	48 (24%)	46 (23%)	65 (33%)
Unharmful Decisions	35 (8%)	87 (21%)	94 (23%)	84 (20%)	116 (28%)
Harmful Decisions	2 (3%)	12 (19%)	18 (28%)	14 (22%)	18 (28%)
Correct Decisions (y/n)	34 (9%)	78 (21%)	78 (21%)	74 (20%)	104 (28%)
Incorrect Decisions (y/n)	3 (3%)	21 (19%)	34 (30%)	24 (21%)	30 (27%)
Decision: Helpful	22 (8%)	58 (22%)	68 (26%)	47 (18%)	70 (26%)
Decision: Inconclusive	6 (5%)	18 (14%)	30 (23%)	32 (24%)	47 (35%)
Decision: Unhelpful	9 (11%)	23 (28%)	14 (17%)	19 (23%)	17 (21%)
Overall	37 (8%)	99 (21%)	112 (23%)	98 (20%)	134 (28%)

Dependent Variables	Prior Treatment Knowledge (Ctrl. Condition)				
	Extensive	Familiar	General	Heard	Nothing
Correct Decisions (3 level)	11 (12%)	33 (35%)	21 (22%)	17 (18%)	12 (13%)
Incorrect Decisions (3 level)	7 (5%)	19 (13%)	44 (30%)	26 (18%)	50 (34%)
Unharmful Decisions	13 (7%)	41 (23%)	41 (23%)	38 (21%)	49 (27%)
Harmful Decisions	5 (9%)	11 (19%)	24 (41%)	5 (9%)	13 (22%)
Correct Decisions (y/n)	13 (9%)	37 (25%)	32 (22%)	31 (21%)	34 (23%)
Incorrect Decisions (y/n)	5 (5%)	15 (16%)	33 (35%)	12 (13%)	28 (30%)
Decision: Helpful	8 (9%)	35 (38%)	26 (28%)	10 (11%)	14 (15%)
Decision: Inconclusive	2 (2%)	8 (9%)	20 (23%)	21 (24%)	37 (42%)
Decision: Unhelpful	8 (14%)	9 (15%)	19 (32%)	12 (20%)	11 (19%)
Overall	18 (8%)	52 (22%)	65 (27%)	43 (18%)	62 (26%)

Table 4.15: Participants' average prior treatment knowledge score for control and search engine tasks. The weighted average scores are calculated by assigning following weights to each knowledge level: Nothing=0, Heard=1, General=2, Familiar=3, Extensive=4.

Dependent Variables	Avg. Treatment Knowledge Score	
	Control	Engine
Correct Decisions (3 level)	2.15	1.78
Incorrect Decisions (3 level)	1.36	1.34
Unharmful Decisions	1.62	1.62
Harmful Decisions	1.83	1.47
Correct Decisions (y/n)	1.76	1.63
Incorrect Decisions (y/n)	1.54	1.49
Decision: Helpful	2.14	1.68
Decision: Inconclusive	1.06	1.28
Decision: Unhelpful	1.85	1.85
Overall	1.67	1.60

to assess their level of confidence in their decision using a 5-point rating scale with the following options:

1. **Very Uncertain:** The participant had difficulty in obtaining a substantial number of conclusive results, hence opting for this choice may indicate that the participant relied only on intuition to assess the effectiveness of the treatment.
2. **Somewhat Uncertain:** Still, the participant had difficulties in obtaining sufficiently definitive results. This option demonstrates a greater tendency for uncertainty rather than certainty.
3. **Neutral:** The level of confidence and doubt seemed to be somewhat balanced, as the participant exhibited a lack of decisiveness about their decision.
4. **Somewhat Certain:** There exists a moderate amount of confidence in the choice that has been taken based on prior knowledge or obtained search results. This alternative exhibits a greater inclination towards certainty as opposed to uncertainty.
5. **Very Certain:** The participant exhibits a high level of confidence in their decision-making process, which is derived from either relying on their prior knowledge or consulting their search engine findings.

The data obtained from the participants' responses to the confidence question is shown in Table 4.16. The first finding derived from the table pertains to the notable rise in the percentages of confidence levels inside the "Somewhat Certain" and "Very Certain" levels in the experimental condition, as compared to the control condition. The degree of confidence has seen a substantial increase, resulting in a significant rise in the combined percentage of confidence in the categories of "Somewhat Certain" and "Very Certain," from 33% to 80%. This observation suggests that individuals tend to exhibit more confidence in their

Task Question: Does selenium help prevent cancer?

Treatment: Selenium is a mineral essential for human health.

Health Issue: Cancer is a disease where the cells in the body start to grow and spread uncontrollably.

Post-task Questionnaire:

On the previous page, you judged the effectiveness of "selenium" for "cancer prevention". How certain are you in your judgment of the treatment's effectiveness?

- Very uncertain
- Somewhat uncertain
- Neutral
- Somewhat certain
- Very certain

2. If you had to give a 'Yes' or 'No' answer to the task question: "Does selenium help prevent cancer?", what would your answer be?

- Yes
- No

SUBMIT

Figure 4.2: Study Web Application's Post-Task Questionnaire Example.

decision-making, particularly in relation to the success of health treatments, when they use a search engine as opposed to when they do not. Another notable finding in the tabulated data is that the proportion of participants who expressed the highest level of certainty in their decision increased almost thrice when they were correct, as compared to when they were incorrect in both control and search engine conditions. In the context of search engine conditions, it was observed that out of the 150 instances when a participant expressed the highest level of confidence (Very Certain) about their decision, a significant majority of 142 participants (95%) made unharmed decisions, in which 121 (81%) of them chose exactly the correct stance of the topic in the 3-level decision-making question. Furthermore, no substantial association was found between the confidence levels of the participants and the decisions they made.

In order to conduct an analysis of the participants' behavior throughout their use of our search environment for the purpose of evaluating the effectiveness of the treatments, we employed the weighted average of their confidence levels across various scenarios. So as to compute the weighted average of the categorical values representing the participants' confidence levels, as submitted in the post-task questionnaire for each task, we applied specific weights to each confidence level as shown below: The scale used to measure uncertainty is as follows: Very Uncertain is assigned a value of 1, Somewhat Uncertain is assigned a value of 2, Neutral is assigned a value of 3, Somewhat Certain is assigned a value of 4, and Very Certain is assigned a value of 5. The weighted average confidence ratings in different scenarios based on each dependent variable are presented in Table 4.17. Based on the overall average confidence scores, it is evident that people expressed much higher levels of confidence in their decision-making while they were using a search engine as opposed to when they were not as the mean of 2.66 confidence score is even less than "Neutral" level but the average score of 4.04 is between "Somewhat Certain" and "Very Certain". Moreover, it is evident that participants who generated the *helpful* and *unhelpful* selections

Table 4.16: Participants' decision confidence for search and control tasks.

Dependent Variables	Confidence (Experimental Conditions)				
	Very Uncertain	Somewhat Uncertain	Neutral	Somewhat Certain	Very Certain
Correct Decisions (3 level)	2 (1%)	7 (2%)	18 (6%)	135 (48%)	121 (43%)
Incorrect Decisions (3 level)	4 (2%)	18 (9%)	44 (22%)	102 (52%)	29 (15%)
Unharmed Decisions	6 (1%)	21 (5%)	51 (12%)	196 (47%)	142 (34%)
Harmed Decisions	0 (0%)	4 (6%)	11 (17%)	41 (64%)	8 (13%)
Correct Decisions (y/n)	5 (1%)	10 (3%)	38 (10%)	179 (49%)	136 (37%)
Incorrect Decisions (y/n)	1 (1%)	15 (13%)	24 (21%)	58 (52%)	14 (13%)
Decision: Helpful	0 (0%)	11 (4%)	24 (9%)	127 (48%)	103 (39%)
Decision: Inconclusive	4 (3%)	14 (11%)	33 (25%)	61 (46%)	21 (16%)
Decision: Unhelpful	2 (2%)	0 (0%)	5 (6%)	49 (60%)	26 (32%)
Overall	6 (1%)	25 (5%)	62 (13%)	237 (49%)	150 (31%)

Dependent Variables	Confidence (Control Condition)				
	Very Uncertain	Somewhat Uncertain	Neutral	Somewhat Certain	Very Certain
Correct Decisions (3 level)	10 (11%)	17 (18%)	17 (18%)	37 (39%)	13 (14%)
Incorrect Decisions (3 level)	51 (35%)	37 (25%)	31 (21%)	22 (15%)	5 (3%)
Unharmed Decisions	48 (26%)	40 (22%)	37 (20%)	41 (23%)	16 (9%)
Harmed Decisions	13 (22%)	14 (24%)	11 (19%)	18 (31%)	2 (3%)
Correct Decisions (y/n)	35 (24%)	31 (21%)	26 (18%)	40 (27%)	15 (10%)
Incorrect Decisions (y/n)	26 (28%)	23 (25%)	22 (24%)	19 (20%)	3 (3%)
Decision: Helpful	8 (9%)	19 (20%)	15 (16%)	38 (41%)	13 (14%)
Decision: Inconclusive	38 (43%)	23 (26%)	20 (23%)	4 (5%)	3 (3%)
Decision: Unhelpful	15 (25%)	12 (20%)	13 (22%)	17 (29%)	2 (3%)
Overall	61 (25%)	54 (23%)	48 (20%)	59 (25%)	18 (8%)

Table 4.17: Participants' decision confidence score for control and search engine tasks. The weighted average scores are calculated by assigning following weights to each confidence level: Very Uncertain=1, Somewhat Uncertain=2, Neutral=3, Somewhat Certain=4, Very Certain=5.

Dependent Variables	Avg. Confidence Score	
	Control	Engine
Correct Decisions (3 level)	3.28	4.29
Incorrect Decisions (3 level)	2.27	3.68
Unharmed Decisions	2.65	4.07
Harmed Decisions	2.69	3.83
Correct Decisions (y/n)	2.79	4.17
Incorrect Decisions (y/n)	2.46	3.62
Decision: Helpful	3.31	4.22
Decision: Inconclusive	1.99	3.61
Decision: Unhelpful	2.64	4.18
Overall	2.66	4.04

had higher levels of confidence in their decisions in comparison to individuals who opted for the *inconclusive* alternative. Another notable fact to consider is that those who made unharmed decisions or correctly answered the Yes/No or the 3-level question had higher levels of confidence compared to those who made the wrong choices.

Chapter 5

Conclusion

This thesis explored the impact of modern commercial search engines on users' decision-making processes in assessing the effectiveness of medical treatments. A controlled experiment was conducted to evaluate the decision-making ability of 60 recruited participants when answering health-related questions, comparing their performance while utilizing current commercial search engines to scenarios in which search engines were not employed. The participants were given a series of 12 tasks aimed at addressing inquiries regarding the effectiveness of medical treatments. It is also noteworthy to mention, these topics used in the tasks were selected from the topics covered in the [TREC 2021 Health Misinformation](#) track.

The key findings of this thesis are:

- Using a search engine facilitates consumers in making more accurate decisions about the helpfulness of medical treatments in general, particularly those treatments that

have been proven to be effective. Nevertheless, in cases when the treatment does not demonstrate clear effectiveness for the specific health issue, search engines have the potential to misguide people about the effectiveness of the treatment.

- Despite the use of modern commercial search engines, the issue of negative effects on some unhelpful treatments by these search engines remains unresolved. Hence, when a search engine is used, the accuracy of participants in identifying helpful treatments significantly improves from 49% to 85%. Conversely, the accuracy in identifying unhelpful treatments remains low, with a rate of 33% compared to 29% when participants did not use a search engine.
- Reduced time allocation for assessing search results, along with a limited number of clicked search results throughout the search process, might potentially lead to a higher likelihood of making wrong decisions.
- Possessing higher levels of prior knowledge of a health issue or medical treatment may enhance individuals' ability to accurately evaluate the effectiveness of a medical treatment when using search engines to answer health questions.
- The utilization of a search engine has the capacity to bolster users' confidence in their decision. Nevertheless, it has been shown that people who exhibit a significant level of confidence in their decisions tend to exhibit a higher level of accuracy in their ultimate decisions.

Our research demonstrates that the use of a modern commercial search engine, in comparison to its absence, can significantly enhance individuals' accuracy in their decisions about the effectiveness of medical treatments. In comparison to the previous studies of [Pogacar et al. \(2017\)](#) and [Ghenai et al. \(2020\)](#), our study showed that even interacting with a modern commercial search engine does not fully eliminate the negative impacts of search

engines on people's decisions, especially when it comes to deciding about the effectiveness of medically unhelpful treatments. Furthermore, our study demonstrated that individuals with a greater degree of prior knowledge on health issues or treatments are less likely to make incorrect or potentially harmful medical decisions when using a modern commercial search engine. Moreover, interacting with a commercial search engine can boost people's confidence in their decisions, regardless of those decisions being correct or not.

Further investigation is required to examine the characteristics and nature of the participants' queries and the search results they clicked on, in order to get insights into user behavior while using search engines while they seek health-related information. This may also aid in comprehending the specific categories of search results that exhibited varying levels of relevance to the users' search queries, particularly in cases when the treatments lack demonstrated effectiveness, therefore misleading users.

References

- Abualsaud, M. and M. D. Smucker (2019). Exposure and order effects of misinformation on health search decisions. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Rome*.
- Allam, A., P. J. Schulz, K. Nakamoto, et al. (2014). The impact of search engine selection and sorting criteria on vaccination beliefs and attitudes: two experiments manipulating google output. *Journal of Medical Internet Research* 16(4), e2642.
- Azzopardi, L. (2021). Cognitive biases in search: a review and reflection of cognitive biases in information retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pp. 27–37.
- Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM* 61(6), 54–61.
- Best, P., E. Gil-Rodriguez, R. Manktelow, and B. J. Taylor (2016). Seeking help from everyone and no-one: Conceptualizing the online help-seeking process among adolescent males. *Qualitative health research* 26(8), 1067–1077.
- Blythe, J. M., E. Sillence, and P. Briggs (2017). The role of trust in ehealth websites: The enduring power of impartiality. In *Frontiers in Public Health. Centre for Behaviour Change Digital Health Conference, University College London*, pp. 22–23.

- Bondarenko, A., E. Shirshakova, M. Driker, M. Hagen, and P. Braslavski (2021). Misbeliefs and biases in health-related searches. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pp. 2894–2899.
- Carr, N. (2008). Is google making us stupid? *Yearbook of the National Society for the Study of Education* 107(2), 89–94.
- Cho, J. and S. Roy (2004). Impact of search engines on page popularity. In *Proceedings of the 13th International Conference on World Wide Web*, pp. 20–29.
- Clarke, C. L., M. Maistro, and M. D. Smucker (2021). Overview of the trec 2021 health misinformation track. In *TREC*.
- Clarke, C. L., A. Vtyurina, and M. D. Smucker (2020). Offline evaluation without gain. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, pp. 185–192.
- Cochran, W. G. and G. M. Cox (1948). Experimental designs. Technical report, North Carolina State University. Dept. of Statistics.
- Dai, E., Y. Sun, and S. Wang (2020). Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In *Proceedings of the International AAAI Conference on Web and Social Media*, Volume 14, pp. 853–862.
- Draws, T., N. Tintarev, U. Gadiraju, A. Bozzon, and B. Timmermans (2021). This is not what we ordered: Exploring why biased search result rankings affect user attitudes on debated topics. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 295–305.
- Finney Rutten, L. J., K. D. Blake, A. J. Greenberg-Worisek, S. V. Allen, R. P. Moser, and B. W. Hesse (2019). Online health information seeking among us adults: measuring progress toward a healthy people 2020 objective. *Public Health Reports* 134(6), 617–625.

- Fox, S. and M. Duggan (2013). Health online 2013. *Health 2013*, 1–55.
- Geldsetzer, P. (2020). Knowledge and perceptions of covid-19 among the general public in the united states and the united kingdom: a cross-sectional online survey. *Annals of internal medicine* 173(2), 157–160.
- Ghahramani, F. and J. Wang (2020). Impact of smartphones on quality of life: A health information behavior perspective. *Information Systems Frontiers* 22(6), 1275–1290.
- Ghenai, A. (2017). Health misinformation in search and social media. In *Proceedings of the 2017 International Conference on Digital Health*, pp. 235–236.
- Ghenai, A., M. D. Smucker, and C. L. Clarke (2020). A think-aloud study to understand factors affecting online health search. In *Proceedings of the 2020 conference on human information interaction and retrieval*, pp. 273–282.
- Jimmy, J., G. Zuccon, B. Koopman, and G. Demartini (2019). Health cards for consumer health search. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 35–44.
- Joachims, T., L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay (2007). Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems (TOIS)* 25(2), 7–es.
- Kattenbeck, M. and D. Elsweler (2019). Understanding credibility judgements for web search snippets. *Aslib Journal of Information Management* 71(3), 368–391.
- Kayhan, V. O. (2013). Seeking health information on the web: positive hypothesis testing. *International Journal of Medical Informatics* 82(4), 268–275.
- Kelly, D. et al. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval* 3(1–2), 1–224.

- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* 30(1/2), 81–93.
- Kitchens, B., C. A. Harle, and S. Li (2014). Quality of health-related online search results. *Decision Support Systems* 57, 454–462.
- Kumpulainen, S., H. C. Huurdeman, et al. (2015). Shaken, not steered: The value of shaking up the search process. In *SCST@ ECIR*.
- Lambert, S. D. and C. G. Loiselle (2007). Health information—seeking behavior. *Qualitative health research* 17(8), 1006–1019.
- Lau, A. Y. and E. W. Coiera (2007). Do people experience cognitive biases while searching for information? *Journal of the American Medical Informatics Association* 14(5), 599–608.
- Lau, A. Y. and E. W. Coiera (2009). Can cognitive biases during consumer health information searches be reduced to improve decision making? *Journal of the American Medical Informatics Association* 16(1), 54–65.
- Odgers, D. J., R. Harpaz, A. Callahan, G. Stiglic, and N. H. Shah (2014). Analyzing search behavior of healthcare professionals for drug safety surveillance. In *Pacific Symposium on Biocomputing Co-Chairs*, pp. 306–317. World Scientific.
- Pogacar, F. A., A. Ghenai, M. D. Smucker, and C. L. Clarke (2017). The positive and negative influence of search results on people’s decisions about the efficacy of medical treatments. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, pp. 209–216.
- Popat, K., S. Mukherjee, J. Strötgen, and G. Weikum (2017). Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 1003–1012.

- Popat, K., S. Mukherjee, J. Strötgen, and G. Weikum (2018). Credeye: A credibility lens for analyzing and explaining misinformation. In *Companion Proceedings of the The Web Conference 2018*, pp. 155–158.
- Pugachev, A., E. Artemova, A. Bondarenko, and P. Braslavski (2023). Consumer health question answering using off-the-shelf components. In *European Conference on Information Retrieval*, pp. 571–579. Springer.
- Purcell, K., L. Rainie, and J. Brenner (2012). Search engine use 2012. <https://www.pewresearch.org/internet/2012/03/09/search-engine-use-2012/>. Accessed: 2023-09-30.
- Rekabsaz, N., O. Lesota, M. Schedl, J. Brassey, and C. Eickhoff (2021). Tripclick: the log files of a large health web search engine. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2507–2513.
- Song, Q. and J. Jiang (2022). How misinformation density affects health information search. In *Proceedings of the ACM Web Conference 2022*, pp. 2668–2677.
- Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science* 185(4157), 1124–1131.
- Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 297–323.
- Vakili Tahami, A. (2022). Answering consumer health questions on the web. Master’s thesis, University of Waterloo.
- Voorhees, E. M., D. K. Harman, et al. (2005). *TREC: Experiment and evaluation in information retrieval*, Volume 63. MIT press Cambridge.

- White, R. (2013). Beliefs and biases in web search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pp. 3–12.
- White, R. W. (2014). Belief dynamics in web search. *Journal of the Association for Information Science and Technology* 65(11), 2165–2178.
- White, R. W. and A. Hassan (2014). Content bias in online health search. *ACM Transactions on the Web (TWEB)* 8(4), 1–33.
- White, R. W. and E. Horvitz (2013). Captions and biases in diagnostic search. *ACM Transactions on the Web (TWEB)* 7(4), 1–28.
- White, R. W. and E. Horvitz (2015). Belief dynamics and biases in web search. *ACM Transactions on Information Systems (TOIS)* 33(4), 1–46.
- White, R. W., S. Wang, A. Pant, R. Harpaz, P. Shukla, W. Sun, W. DuMouchel, and E. Horvitz (2016). Early identification of adverse drug reactions from search log data. *Journal of Biomedical Informatics* 59, 42–48.
- Windfeld, A. C. and F. M. Meier (2021). “does vinegar kill coronavirus?”-using search log analysis to estimate the extent of covid-19-related misinformation searching behaviour in the united states. In *iConference 2021: Diversity, Divergence, Dialogue*.
- Yom-Tov, E., E. Gabrilovich, et al. (2013). Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. *Journal of medical Internet research* 15(6), e2614.
- Zhang, D., A. Vakili Tahami, M. Abualsaud, and M. D. Smucker (2022). Learning trustworthy web sources to derive correct answers and reduce health misinformation in search. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2099–2104.

APPENDICES

Appendix A

Participants' Queries

Table A.1: Queries made by participants per topic sorted by number of users put the query.

Topic#	Topic Question	Query	#
101	Will wearing an ankle brace help heal Achilles tendonitis?	achilles tendonitis	6
		ankle brace for achilles tendonitis	5
		achilles tendonitis + ankle brace	2
		achilles tendonitis ankle brace	2
		ankle brace and achilles tendonitis	2
		“ankle brace” AND “Achilles tendonitis”	1
		“ankle brace” Achilles tendonitis	1
		“ankle brace” achilles tendonitis	1
		Can ankle brace help heal achilles tendonitis	1
		Does ankle brace help in the treatment of achilles tendonitis	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		Will an ankle brace help achelles tendonitis	1
		Will wearing an ankle brace help heal Achilles tendonitis?	1
		achilles ankle brace	1
		achilles tendonities treatment	1
		achilles tendonities treatment with ankle brace	1
		achilles tendonitis how to heal	1
		achilles tendonitis reduced using ankle braces?	1
		achilles tendonitis treatment	1
		achilles tendonitis treatments	1
		ankle brace heals achilles tendonitis	1
		ankle brace Achilles tendonitis	1
		ankle brace achilles tendonitis	1
		ankle brace for Achilles tendonitis	1
		ankle brace for achilles tendonitis mayo	1
		ankle brace for achilles tendonitis ncbi	1
		ankle brace for tendonitis	1
		ankle brace resolve achilles tendonitis	1
		ankle brace to help with achilles tendonitis	1
		ankle brace vs athletic wrap	1
		ankle braces and achilles tendonitis	1
		ankle place to help with achilles tendonitis	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		are ankle brace recommended for achilles tendonitis	1
		can an ankle brace help with pain from achilles tendonitis?	1
		can ankle brace help heal Achilles tendonitis	1
		can wearing ankle brace help with achilles tendonitis	1
		cause of achilles tendonitis	1
		do ankle brace heal tendonitis	1
		do ankle braces help achilles tendonitis	1
		do ankle braces help with achilles tendonitis?	1
		do ankle braces work for achilles tendonitis	1
		does ankle brace help healing achilles tendonitis?	1
		does ankle brace help in healing Achilles tendonitis?	1
		is ankle brace useful for achilles tendonitis	1
		medical studies ankle braces effect on achilles tendonitis	1
		methods to treat achilles tendonitis	1
		should I use ankle brace for achilles tendonitis	1
		should i wear an ankle brace for achilles tendonitis	1
		treatments for achilles tendonitis	1
		treatments for achilles tendonitis ankle brace	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		wearing an ankle brace heal achilles tendonitis	1
		what does an ankle brace do	1
		what is achilles tendonitis	1
		what is achilles tendonitis?	1
		what us achilles tendonities	1
		will ankle brace help heal achilles tendonitis	1
		will wearing an ankle brace heal achilles tendonitis	1
103	Can folic acid help improve cognition and treat dementia?	folic acid and dementia	7
		folic acid for dementia	5
		folic acid dementia	4
		folic acid	3
		can folic acid help dementia	2
		dementia folic acid	2
		what is folic acid	2
		“folic acid” AND “dementia”	1
		Can folic acid help in improving cognition and treating dimentia	1
		Can folic acid treat dementia	1
		Can folic acid treat dementia?	1
		b12 folic acid	1
		can folic acid help improve cognition and treat dementia?	1
		can folic acid help treat dementia	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		can folic acid help with dementia	1
		can folic acid improve cognition and dementia	1
		can folic acid treat dementia	1
		can you treat dementia	1
		dementia	1
		dementia treatment with folic acid	1
		dementia treatments	1
		does folic acid help cognition	1
		does folic acid help dementia	1
		does folic acid help treat dementia	1
		does folic acid improve cognition and treat dementia	1
		does folic acid treat dementia	1
		effectiveness of folic acid slowing down dementia	1
		explain dementia	1
		folic acid and dementia treatment	1
		folic acid benefits	1
		folic acid cognition dementia	1
		folic acid dementia	1
		folic acid dementia treatment	1
		folic acid for cognitive health	1
		folic acid for dementia treatment	1
		folic acid health benefits	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		folic acid improve cognition	1
		folic acid in dementia	1
		folic acid on dementia	1
		folic acid supplements for dementia	1
		folic acid to treat dementia	1
		folic acid treatment on cognition and dementia	1
		folic acid use in dementia patients	1
		folic acid use in dementia patientws	1
		folic acid used to treat dementia	1
		folic acid, dementia	1
		foods rich in folic acid	1
		how is memory loss related to folic acid	1
		improve dementia+folic acid	1
		what is folic acid “brain”	1
		what is folic acid brain	1
		what is folic acid used for	1
107	Does yoga improve the management of asthma?	yoga for asthma	5
		asthma and yoga	3
		can yoga help with asthma	2
		do doctors support yoga for asthma patients	1
		“yoga” + “asthma”	1
		Can yoga help asthama	1
		Can yoga help asthma	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		Can yoga reduce symptoms of asthma condition?	1
		Does yoga help with Asthma?	1
		Does yoga improve the management of asthma	1
		Is Yoga helpful for treating Asthma?	1
		asthma	1
		asthma and toga	1
		asthma management	1
		asthma treatment yoga	1
		asthma yoga	1
		can yoga help asthma	1
		can yoga help in reducing issues such as asthma	1
		can yoga help managing asthma	1
		can yoga help to manage asthma	1
		can yoga help with managing the symptoms of asthma	1
		does yoga help asthma	1
		does yoga help in asthma	1
		does yoga help to manage asthma?	1
		does yoga help with asthma	1
		does yoga help with asthma studies	1
		efficacy of yoga for asthma	1
		efficacy of yoga for asthma 2023	1
		how yoga can heal people with asthma?	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		how yoga effective on healing people with asthma?	1
		how yoga effectiveness on healing people with asthma?	1
		is yoga helpful for asthma management	1
		is yoga helpful for asthma management from a medical standpoint	1
		yoga	1
		yoga and asthma	1
		yoga and asthma management	1
		yoga and benefits for treating asthma	1
		yoga asthma	1
		yoga benefits asthma	1
		yoga effect	1
		yoga effectiveness for asthma	1
		yoga helpful for asthma	1
		yoga to help asthma	1
		yoga to manage asthma	1
		yoga used to treat asthma	1
		yoga's effect on asthma	1
		yogas effect on asthma	1
109	Does selenium help prevent cancer?	selenium and cancer	4
		does selenium prevent cancer	3
		selenium cancer	3
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		selenium to prevent cancer	3
		what is selenium	3
		can selenium prevent cancer	2
		selenium cancer treatment	2
		Can the use of selenium prevent cancer?	1
		Does selenium help in cancer prevention ?	1
		Does selenium help prevent cancer	1
		Does selenium help prevent cancer?	1
		Is Selenium effective in preventing Cancer?	1
		Is selenium effective in preventing cancer?	1
		WHat is selenium mineral ?	1
		What is selenium?	1
		aetiological	1
		cancer prevention selenium	1
		cancer selenium	1
		cancer treatment selenium	1
		cancerl treatment selenium	1
		does selenium help prevent cancer	1
		does selenium prevent cancer?	1
		efficacy of selenium use in cancer prevention	1
		how does selenium help prevent cancer?	1
		how selenium prevent cancer spread	1
		prevent cancer with selenium	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		selenium	1
		selenium alternatives for cancer treatment	1
		selenium and cancer prevention	1
		selenium and cancer prevention	1
		selenium cancer prevention	1
		selenium effect in prevent cancer	1
		selenium for cancer prevention	1
		selenium help prevent cancer	1
		selenium help with cancer	1
		selenium in cancer prevention for women	1
		selenium preventing cancer	1
		selenium research for cancer prevention	1
		uses of selenium	1
		uses of selenium in health industry	1
112	Can evening primrose oil help treat eczema?	can evening primrose oil help treat eczema	2
		evening primrose oil for eczema	2
		evening primrose eczema	2
		“evening primrose” AND “eczema”	1
		Can we use primrose oil to treat eczema	1
		benefits of evening primrose oil	1
		can evening primrose oil help with eczema	1
		can evening primrose oil treat eczema?	1
		can primrose oil help with eczem	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		does evening primrose irritate skin	1
		does evening primrose help eczema	1
		does evening primrose oil treat eczema	1
		does primrose oil help relieve eczema	1
		eczema	1
		eczema primrose oil	1
		evening pimrose oil for eczema	1
		evening primrose applied to treat eczema	1
		evening primrose eczema	1
		evening primrose for eczema	1
		evening primrose eczema symptoms	1
		evening primrose for ecxema	1
		evening primrose for eczema	1
		evening primrose for eczema treatment	1
		evening primrose oil and eczema	1
		evening primrose oil and treat eczema	1
		evening primrose oil eczema	1
		evening primrose oil foe eczema	1
		evening primrose oil for eczema	1
		evening primrose oil on eczema	1
		evening primrose	1
		evening primrose for eczema	1
		how can primerose help to treat eczema?	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		how to get rid of eczema	1
		how to treat eczema	1
		is evening primrose a good treatment for eczema	1
		is primrose oil a solution for eczema?	1
		primrose oil and eczema	1
		primrose oil and exzema	1
		primrose oil eczema	1
		primrose oil for eczema	1
		primrose oil treatment for eczema	1
		treating exzema with evening primrose oil	1
		treatment for eczema	1
		what is evening primrose for	1
		what is evening primrose used for	1
		what type of oil does evening primrose plant seeds produce?	1
115	Can magnesium help prevent migraine?	magnesium to prevent migraine	3
		can magnesium help prevent migraines	2
		magnesium and migraine	2
		magnesium for migraines	2
		A randomized prospective placebo-controlled study of intravenous magnesium sulphate vs. metoclopramide in the management of acute migraine attacks in the emergency department.	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		Can magnesium help in treating migranes	1
		Can magnesium help prevent migraines	1
		Can taking magnesium prevent a migraine?	1
		Intravenous magnesium sulphate in the acute treatment of migraine without aura and migraine with aura	1
		The acute treatment of migraine in adults: the American Headache Society evidence assessment of migraine pharmacotherapies	1
		can magnesium help migraines	1
		can magnesium help prevent migraine	1
		can magnesium help with migraines	1
		can taking magnesium help with migraines	1
		does magnesium help in preventing migrain?	1
		does magnesium help prevent migraine	1
		does magnesium help prevent migraines	1
		does magnesium prevent mirgraines	1
		effect of magnesium on migraine	1
		healing migrain with intake magnesium	1
		how to prevent migraines	1
		magneisum and migraine	1
		magnesium and migraines	1
		magnesium for migraines treatment	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		magnesium for migranes	1
		magnesium migraine treatment	1
		magnesium on migraines	1
		magnesium prevent migraine	1
		magnesium preventing migraine	1
		magnesium to help migranes	1
		magnesium use in migriaine	1
		magnesium use in migriaine healthline	1
		magnesium, migraine treatment	1
		magneusim and migraine prevention	1
		migrain and magnesium	1
		migraine remedy	1
		migraine remedy vitamins	1
		migraine treatment	1
		migraine treatments	1
		migraine with aura	1
		migraines and magnesium	1
		migraines treatment	1
		prevent migraine	1
		prevent migraine + minerals	1
		prevent migraine + vitamin	1
		prevent migrane with magnesium	1
		prevention of migraines	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		prevention of migraines magnesium	1
		treating migraines	1
		treatments to prevent migraines	1
		what is magneisum good for?	1
118	Can dupixent treat eczema?	can dupixent treat eczema	6
		dupixent eczema	6
		dupixent	4
		dupixent and eczema	3
		dupixent for eczema	3
		does dupixent treat eczema	2
		dupixent to treat eczema	2
		eczema treatments	2
		what is dupixent	2
		“dupixent” + “eczema”	1
		Can I take dupixent for eczema?	1
		Can dupixent treat eczema?	1
		Dupixent effect on eczema	1
		Is dupixent treat eczema	1
		What is dupixent	1
		can dupixent trean eczema	1
		can dupixent treat eczema?	1
		can dupixent treat exzema	1
		can eczema be treated by dupixent?	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		does dupixent help eczema	1
		duoixent	1
		dupimulab	1
		dupixent and eczema - mayo clinic	1
		dupixent and eczema treatment	1
		dupixent benefits	1
		dupixent for eczema studies	1
		dupixent monoclonal antibody to treat eczema	1
		dupixent on eczema	1
		dupixent success eczema	1
		dupixent treat eczema	1
		eczema	1
		eczema dupixent treatment	1
		eczema dupixent treatment effectiveness	1
		eczema treatments by dupixent	1
		what is dupixebt	1
		what is dupixent for	1
		what is eczema	1
128	Does steam from a shower help croup?	croup steam	4
		croup	3
		croup treatments	3
		what is croup	3
		can steam help croup	2
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		croup hot shower	2
		does steam shower help croup	2
		remedies for croup	2
		shower steam and croup	2
		is steam helping to treat croup	1
		steam and croup	1
		wikipedia croup	1
		“shower” + “croup”	1
		“steam” + “croup”	1
		Can we use steam from our shower to treat croup	1
		Can we use steam to treat croup	1
		Croup	1
		Does shower steam help with croup	1
		Does steam from a shower help croup?	1
		Steam effect on Croup	1
		Will taking a shower steam help with croup?	1
		can a hot shower help croup	1
		can hot steam help croup	1
		can i use steam to help treat croup	1
		can shower steam help croup	1
		can steam help with croup	1
		can steam inhalation alleviate croup cough	1
		croup and hot steam	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		croup and hot steam from shoer==wer	1
		croup and hot steam from shower	1
		croup be treated using steam from shower?	1
		croup home remedies	1
		croup hot shower treatment	1
		croup humidifier	1
		croup humidity treatment	1
		croup release method	1
		croup shower	1
		croup steam	1
		croup steam treatment	1
		croup treatment	1
		does hot shower help croup	1
		does steam from a shower help croup	1
		does steam from shower help in croup	1
		does steam help croup	1
		does steam help with croup?	1
		effect of steam on croup	1
		fix croup in shower	1
		hot shower for croup	1
		how steam could help the croup	1
		humidifier croup	1
		is shower steam an effective remedy for croup?	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		is steam effective to cure croup?	1
		is steam effective to help croup?	1
		nemours	1
		nemours wiki	1
		steam breating for croup treatment	1
		steam from a shower help croup?	1
		steam treatment for croup	1
		what causes croup	1
134	Can I remove a tick by covering it with Vaseline?	tick removal vaseline	3
		removing ticks with vaseline	2
		vaseline for removing ticks	2
		vaseline ticks	2
		Can I remove a tick by covering it in vsline	1
		Can I remove a tick by Vaseline?	1
		Can I remove a tick by covering it with Vaseline	1
		Can I remove a tick by covering it with Vaseline?	1
		Can vaseline help in the treatment of Ticks ?	1
		Can we use vaseline to remove a tick	1
		Vaseline to remove ticks	1
		best methods to remove ticks	1
		can I remove ticks with vaseline	1
		can i remove a tick by covering it with vaseline	1
		can i remove ticks with vaseline	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		can i use vaseline to remove a tick?	1
		can vaseline be used to remove ticks?	1
		can vaseline remove a tick	1
		can you use vaseline for tick removal?	1
		can you use vaseline to get rid of ticks	1
		can you use vaseline to remove ticks	1
		common tick remedies	1
		does vaseline help in removing ticks from body?	1
		does vaseline help to remove ticks?	1
		how to get rid of ticks	1
		how to get rid of ticks on the skin	1
		how to remove a tick	1
		how to remove tick	1
		how to remove ticks	1
		negative effects of vaseline for tick	1
		removal of tick from skin	1
		remove tick by vaselin?	1
		remove tick using vaseline	1
		remove tick with only vaseline	1
		remove tick without tweezers	1
		removing tick with vaseline	1
		removing ticks using vaseline or petroleum jelly	1
		removing ticks vaseline	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		tick bite heal	1
		tick bite vaseline	1
		tick removal	1
		tick removal vaseline danger	1
		tick remove with Vaseline	1
		tick vaseline	1
		ticks in humans	1
		tuck remove with Vaseline	1
		vaseline as tick treatment	1
		vaseline for tick removal	1
		vaseline for tick treatment	1
		vaseline for ticks	1
		vaseline remove tick	1
		vaseline side effects for tick	1
		vaseline to remove tick	1
		ways of removing tick	1
136	Can eating dates help manage iron deficiency anemia?	can dates help with anemia	3
		dates and anemia	3
		dates and iron deficiency anemia	3
		“dates” AND “iron deficiency”	1
		Can dates help in treating anemia	1
		Can dates help iron deficiency anemia	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		Can eating dates help manage iron deficiency anemia?	1
		Dates for iron deficiency anemia	1
		Does consuming dates help with managing iron deficiency anemia?	1
		Does eating dates cure anemia	1
		Does eating dates regularly help with iron deficiency?	1
		anemia	1
		anemia management by eating dates	1
		anemia treatment	1
		are dates good for anemia	1
		are dates high in iron	1
		are dates high in iron?	1
		are dates rich in iron?	1
		benefits of dates	1
		benefits of eating dates	1
		can dates help iron deficiency anemia	1
		can dates help reduce iron defficiency	1
		can dates help to treat anemia?	1
		can dates help with iron deficiency anemia	1
		can dates solve iron deiciency anemia	1
		can eating dates help with anemia	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		consumption dates remedy for anemia peer reviewed	1
		date fruit to treat anemia	1
		date nutrition	1
		date nutrition iron	1
		dates anemia scholar google	1
		dates anemia studies	1
		dates for anemia	1
		dates for anemia treatment	1
		dates for iron deficiency anemia	1
		dates for treating anaemia	1
		dates fruit	1
		dates fruit image	1
		dates health benefits	1
		dates help manage iron deficiency anemia?	1
		dates image	1
		dates iron deficiency	1
		dates nutrients	1
		dates remedy for anemia peer reviewed	1
		dates to treat anemia	1
		dates, iron deficiency anemia	1
		dies eating dates help manage iron deficiency and anemia	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		dietary sources of iron for treating anemia	1
		do dates contain iron	1
		do dates have iron	1
		do dates have iron?	1
		do dates help anemia	1
		do dates help anemia?	1
		do dates help iron deficiency	1
		do dates help treat iron deficiency	1
		do dates help with anemia	1
		does dates fruit help iron deficiency anemia	1
		does eating dates help manage iron deficiency and anemia	1
		does iron in date help anemia	1
		does iron in date help iron deficiency anemia	1
		eating dates benefits	1
		eating dates help iron deficiency anemia	1
		eating dates iron deficiency anemia	1
		eating dates would treat anemia	1
		food rich in iron	1
		foods that help anemia	1
		foods to help anemia	1
		google scholar	1
		how to treat anemia?	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		iron content of dates	1
		iron deficiency anemia treatment	1
		iron in dates	1
		is dates full of iron	1
		list of vegetables and fruits that help iron deiciency anemia	1
		pubmed dates for anemia treatment	1
		scientific correlation between date fruit for anemia treatment	1
		scientific correlation between dates for anemia treatment	1
140	Can fungal creams treat athlete's foot?	athlete's foot treatment	4
		athlete's foot	3
		fungal cream for athletes foot	3
		fungal creams for athlete's foot	3
		fungal cream	2
		fungal cream to treat athlete's foot	2
		fungal creamsfor athelete foot treatment	1
		Can athletes foot be treated with fungal creams?	1
		Can fungal creams be used for athletes foot?	1
		Can fungal creams help with athlet's foot?	1
		Is fungal cream effective in treating athlete's foot	1
		anti-fungal cream athlete's foot	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		antifungal cream for athlete's foot	1
		antifungal cream to treat athlete's foot	1
		antifungal creams	1
		antifungal creams for athlete's foot	1
		are fungal creams effective athelete foot	1
		atheletes foot fungal creme	1
		athlere's foot treatment using fungal creams	1
		athlete's foot antifungal cream	1
		athlete's foot causes	1
		athlete's foot cure	1
		athlete's foot how to treat	1
		athlete's foot prescribed treatments	1
		athlete's foot remedies	1
		athlete's foot treatment anti-fungal cream	1
		athlete's foot treatment fungal cream	1
		athlete's foot treatment using fungal creams	1
		athlete's foot treatments	1
		athlete's foot treatments fungal cream	1
		athlete's foot wikipedia	1
		athlete's foot, bacteria or fungus	1
		athletes foot can be treated with antifungal cream	1
		can fungal creams treat athlete's food	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		clotrimazole (Lotrisone), econazole (Ecoza, Spectazole) or ciclopirox (Loprox, Penlac)	1
		clotrimazole cream	1
		clotrimazole cream fungal cream	1
		cream treatments for athlete's foot	1
		does fungus cream get rid of athlete's foot	1
		fungal cream athlete's foot	1
		fungal cream for athlete's foot	1
		fungal cream on athlete's foot	1
		fungal cream treatment athlete's foot	1
		fungal cream vs antifungal cream	1
		fungal cream wikipedia	1
		fungal creams and athletes foot	1
		fungal creams athlete's foot treatment	1
		fungal creams for athlete's foot	1
		fungal creams to treat athlete's foot	1
		fungal creams to treat athletes foot	1
		how to treat athlete's foot? using fungal creams	1
		how to treat athlete's foot	1
		treatment for athlete's foot	1
		treatments for athlete's foot	1
		what causes athlete's foot	1
		what is a fungal cream	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		what is athlete's foot	1
		what is fungal creams	1
149	Will at-home exercises manage hip osteoarthritis pain?	osteoarthritis	2
		osteoarthritis treatments	2
		exercise and osteoarthritis	1
		"At-home exercises" AND "Osteoarthritis"	1
		Can at home exercises help manage hip osteoarthritis pain	1
		arthritis	1
		at home exercise and osteoarthritis	1
		at home exercises for osteoarthritis	1
		at home exercises hip osteoarthritis	1
		at-hom excercise to help with osteoarthritis pain	1
		at-home excercise to help with osteoarthritis pain	1
		at-home exercises for hip arthritis	1
		at-home exercises for hip osteoarthritis pain	1
		at-home exercises for osteoarthritis	1
		at-home exercises for osteoarthritis pain	1
		at-home exercises to reduce osteoarthritis pain	1
		at-home exercises+hip osteoarthritis pain	1
		can at-home exercises help manage hip osteoarthritis pain?	1
		can calisthenics help with hip osteoarthritis	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		do at-home excercises help with osteoarthritis?	1
		does at-home exercise help with hip osteoarthritis pain?	1
		does exercise for hip osteoarthritis need spicific equipment?	1
		does exercise help osteoarthritis	1
		does exercise help with hip osteoarthritis	1
		does exercizing at home manage hip osteoarthritis	1
		does home excercises help with hip osteoarthritis	1
		excecise for hip osteoathritis pain study	1
		excercises osteoarthritis	1
		exercise and hip osteoarthritis pain	1
		exercise hip osteoarthritis	1
		exercises for Osteoarthritis	1
		get rid of hip osteoarthritis pain with home exercise	1
		hip osteoarthritis pain	1
		hip osteoarthritis pain at home excercises	1
		hip osteoarthritis pain treatment	1
		hip osteoarthritis pain treatment exercise	1
		hip osteoarthritis with at home excercises	1
		home excecise for hip osteoathritis pain	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		home excersices enough to get rid of hip osteoarthritis?	1
		home exercise for hip osteoarthritis	1
		home remedies for osteoarthritis	1
		home treatments for hip osteoarthritis	1
		how do at-home exercises help manage hip osteoarthritis pain	1
		in home exercise for hip osteoarthritis	1
		managing hip osteoarthritis pain at home by exercise	1
		managing osteoarthritis hip pain at home	1
		osteoarthiritis	1
		osteoarthritis exercise	1
		osteoarthritis pain	1
		osteoarthritis treatment	1
		osteoarthritis with at home exercises	1
		osthearthritis exercises	1
		reduce osteoarthritis pain	1
		remedies for osteoarthritis pain	1
		what are the best ways to manage hip osteoarthritis pain?	1
		what can manage hip osteoarthritis	1
		what is hip osteoarthritis	1
⋮	⋮	⋮	⋮

Table A.1 continued from previous page.

Topic#	Topic Question	Query	#
⋮	⋮	⋮	⋮
		what is osteoarthritis	1