A Think-Aloud Study about Medical Misinformation in Search Results

Amira Ghenai[1] and Mark D. Smucker[2] and Charles L. A. Clarke[1]
[1] David R. Cheriton School of Computer Science, University of Waterloo, Canada
[2] Department of Management Sciences, University of Waterloo, Canada
{aghenai,mark.smucker,charles.clarke}@uwaterloo.ca

ABSTRACT

People increasingly rely on the internet in order to search for health related information. Searching for information about medical treatments is among the most frequent uses of search engines. While being a convenient and fast method to collect information, search engines have a content bias towards web pages stating that treatments are helpful, regardless of the truth. The presence of incorrect information in search results might potentially cause harm, especially if people believe what they read without further research or professional medical advice. In this paper, we aim to understand the decision making process of determining the efficacy of medical treatments using search result pages. We use a think-aloud study in order to gain insights on strategies people use during online search for health related topics. Results show that, even with verbalization, participants are still strongly influenced by a search results bias. Furthermore, people pay attention to majority, authoritativeness and content quality when evaluating online content. Rank and participants' bias towards treatments being helpful are potential subconscious biases influencing the decision making process while using search engines.

CCS CONCEPTS

• **Information systems → Users and interactive retrieval**; **Retrieval effectiveness**.

KEYWORDS

Health Search; User Study; Misinformation; Coding; Analysis; Think-aloud protocols

1 INTRODUCTION

The majority of US internet users rely on web search to look for information about a health issue or a medical treatment [5]. However, there is an increased concern over the lack of accountability and dubious quality of this online content. Prior research [16] has shown that search engines can be biased towards stating that medical treatments are helpful, regardless of the truth. Given the substantial impact of search engines on people's decision making, if results are biased towards incorrect information, people's accuracy reduces from 43% to 23%, and there is a potential harm in the case that people believe this incorrect information [10].

To better understand the decision making process while people use search engines for health related purposes, we report on a study using a think-aloud method. Collecting and analyzing think aloud protocols has been used in literature to build models of cognitive processes during a problem solving task[13]. Applying the think-aloud method, we aim to gain some insight on the strategies used by participants while using search engines to answer health related queries. The insights will be helpful to improve and build search engines that better support people's decision making.

In this paper, we ask participants to determine the effectiveness of four medical treatments. We provide participants with search result pages that help them answer the questions about the treatment's efficacy. While doing the task, we ask participants to say out loud what goes through their head by stating directly what they think. Later, we ask participants about their decisions during the task and about using search engines for health related purposes. We found that:

- Even with verbalization, participants are being heavily influenced by a search result bias. When biased towards correct information, participants' accuracy reached 67% while the accuracy was reduced to 32% when search results were biased towards incorrect information.
- Majority, authoritativeness and quality are among the important aspects people pay attention to when using search result pages to answer health related questions
- There are factors that effect people's decisions about the efficacy of the medical treatment that have been hidden during the think-aloud process. People have subconscious biases such as rank and helpful bias.

We next discuss related work. We then cover the details of the study and present the study's results, along with our conclusions.

2 RELATED WORK

The work proposed in this paper builds directly on Pogacar et al. [10]'s work. The major key finding of that work is that search results have a strong statistically significant effect on people's decision about the efficacy of medical treatments. The study showed that when search results are biased towards incorrect information, people's accuracy is reduced from 43% to 23% while, when biased towards correct information, people's accuracy increased from 43% to 65%. Further, the rank of the topmost correct result has some effect on people's accuracy. Additionally, knowledge of the medical treatment can protect people from the presence of incorrect information in search results. Finally, participants are generally

ROME 2019, July 25, 2019, Paris, France

ACM ISBN yyy…$15.00
https://doi.org/xxx

biased towards stating that the treatment is helpful, regardless of the truth.

These findings confirm that search engines may have a substantial impact on people's decisions about the efficacy of medical treatments. Results show that, even when there was always a correct answer in either rank 1 or 3, participants were not able to successfully find the correct answer. More importantly, search engines can potentially harm people with a mix of correct and incorrect information. In order overcome the current search engine limitations and build ones that better support people's decision making, it is paramount to further explore the reasons leading people to be heavily influenced by misleading information in search results.

In Pogacar et al. [10]'s study, limited information about the interaction of participants with search results (i.e., click behavior and search logs) is collected. In the work presented in this paper, we extend Pogacar et al. [10]' study by investigating, in more detail, the interactions between people and search engines. In doing this, we aim to shed the light on possible explanations of the impact of search results on people's decision making about the efficacy of medical treatments.

In addition to this paper, White et al. [14–17] demonstrated that search engines have a strong content bias towards stating that medical treatments are helpful even when they actually are not. Tang et al. [12] compared the search results of Google to those of a domain-specific health and depression search engine. Results showed that, while Google returns more relevant documents, the domain specific search engine returns more correct search result pages.

Elsweiler et al.[3] designed a think-aloud user study in order to shed the light on how people access the credibility of search result pages. Results showed that people are not certain when accessing the credibility of online sources. People use ten different cues in order to access the credibility of sources and the usage of these cues differ for each participant and each topics.

3 MATERIALS AND METHODS

3.1 Study Design

First, we calibrated the eye tracking device to measure the participants eye movement. Next, participants signed the consent forms then filled out a questionnaire providing demographic information. Following the questionnaire, they read detailed instructions about the participation before proceeding with the study. After that, with the help of search results, participants had the chance to practice determining the effectiveness of a medical treatment. While doing the practice task, participants were asked to articulate and say their thoughts out loud. Later, we started video and audio recording of the participants. Then, participants began the main study where they had to determine the effectiveness of four medical treatments while thinking out loud (Concurrent think-aloud). While participants were doing this search task, we wrote down notes about the verbal and non-verbal interactions. After finishing the search task, we showed participants the video recording of the participation, with their eye movements to help them remember their thoughts, and asked them questions about their decisions (Retrospective think-aloud). Finally, we ask participants general questions about their usage of search engines for health related purposes (Questionnaire).

The study is designed as a web application and the search results are modelled as a traditional style of web search engine. We recreated the interface of Pogacar et al. [10] for our study. Refer to Pogacar et al.[10]"s work to get more details about the user interface design.

Table 1: This table shows the medical treatments with their corresponding efficacy

T	Medical Treatment	Efficacy
T1	Do antioxidants help female subfertility?	Unhelpful
T2	Do benzodiazepines help alcohol withdrawal?	Helpful
T3	Do probiotics help treat eczema?	Unhelpful
T4	Does caffeine help asthma?	Helpful
T5	Does cinnamon help diabetes?	Unhelpful
T6	Does melatonin help treat and prevent jet lag?	Helpful
T7	Does surgery help obesity?	Helpful
T8	Does traction help low back pain?	Unhelpful

3.2 Study Material

We use the study material from the publically available dataset[1]. In this section, we briefly explain the study material. Refer to Pogacar et al. [10]'s work for a detailed explanation.

We controlled search result content in terms of two levels. First, the search result bias which was either *correct* or *incorrect*. Second, the topmost correct search result where we place the first correct result at either rank 1 or 3. Further, we measure participants' performance by keeping track of the fraction of correct decision and the fraction of harmful decisions. Participants had to determine the efficacy of medical treatments as either *helpful* or *unhelpful* or *inconclusive*.

3.2.1 Medical Treatments. We use a list of 8 medical treatments from Pogacar et al.[10]'s study. Each medical treatment can either be: *helps* (the medical treatment has a direct positive influence on a specific illness), *inconclusive* (medical professionals are not sure about the effectiveness of the medical treatment) or *does not help* (the medical treatment has either a direct negative influence or no influence on a specific illness). Out of the 8 medical treatments, four were *helpful* and four were *unhelpful*. Table 1 shows the list of the medical treatments with their corresponding effectiveness.

3.2.2 Search Results. During the study, we asked participants to pretend they had a question about the effectiveness of a medical treatment and decided to use a search engine to help them answer the question. We showed participants a web page that had ten search results, with the general appearance of a standard search engine results page (SERP). The search results were either biased towards correct or incorrect information. When biased towards correct, we showed eight correct search result pages and 2 incorrect ones. When biased towards incorrect, we showed participants eight incorrect search result pages and two correct ones. We, further, controlled for the rank of the topmost correct result page to either

[1] https://cs.uwaterloo.ca/~aghenai/user_study_pages.html

be at rank 1 or 3. We randomly assigned the search results to the corresponding ranks from a pool of 8-10 correct and 8-10 incorrect documents.

3.2.3 Documents and Snippets. To build the SERP pages, we collected documents about efficacy of the medical treatment. We use the same 158 documents used in Pogacar et al. [10]'s work for this purpose. Every document is either correct (contains information about the treatment efficacy that agrees with the truth) or incorrect (contains information about the treatment efficacy that contradicts with the truth). For every search result, we show the document's title,url, and snippet. We use the same snippets generated from Pogacar et al.[10]'s study.

3.3 Performance and Statistical Significance

We measure the participants' performance in the user study by computing two different measures: the fraction of correct decisions and the fraction of harmful decisions. A participant's decision is correct if it agrees with the truth. Note that, inconclusive is considered an incorrect decision as all medical treatments are either helpful or unhelpful. Further, a participants' decision is harmful if it is opposite to the truth where inconclusive is not considered a harmful decision.

The fractions of correct and harmful decisions are the dependent variables. The search result bias and the topmost correct are the independent variables. In order to measure the statistical significance of the independent variables on the fractions of correct and harmful decisions, we used generalized linear mixed effect model in R. More details about the modeling method can be found in Pogacar et al. [10]'s paper.

3.4 Think-aloud Protocol

We use the think-aloud protocol during the study in order to reveal the potential factors influencing the decision making process of people using search engines to answer health related questions. We believe that the study is suitable to apply the think-aloud protocol as the tasks are of intermediate level of difficulty [2]. We combine two types of the think-aloud protocol in the study: concurrent and retrospective think-aloud for different reasons. Concurrent think-aloud (CTA) is verbalization of thoughts as the task is being completed while retrospective think-aloud (RTA) is asking the participant about the thoughts after completing the task [11]. We choose to apply CTA as it is helpful in extracting immediate thoughts while doing the task. We further implement RTA as it is helpful when participants do not verbalize enough the ideas. It is also a chance to deeper thoughts and better interpret and validate the CTA (such as asking about pauses etc.) [8] .

In the CTA part, we ask the participant to say out loud their thoughts while doing the search task. There was no interaction between the participant and the searcher. During the process, a video recording was made of the screen and an audio recording recorded the think-aloud. In the meantime, we made notes of the participants verbal and non-verbal interactions. A "KEEP TALKING" sign was used to remind participants to talk without distracting the thinking process. In the RTA part, when all the four search tasks were finished, we asked the participant questions related to the think-aloud. In this workshop paper, we only report the results of the CTA part and the RTA results will be reported in a later full paper publication.

3.5 Transcription

We video recorded the concurrent and retrospective think-aloud process while participants interacted with the search task and audio recorded the questionnaire part. An outside vendor transcribed all the parts of the collected data. The reported results in this report are based on the transcribed data. The transcription service included timestamps for the transcribed scripts without verbatim (filler words are removed from the transcripts).

3.6 Coding Scheme

After transcribing the think-aloud recordings, we start the coding process in which we generated tags in order to quantify the observations during the think-aloud. The coding process was performed by one of the paper authors. We use QSR International's NVivo 12 qualitative data analysis software [9] for the coding process.

We perform qualitative analysis for the think-aloud data using a mixed methods research for both the bottom-up and the top-down approach [6]. Some of the codes were inspired by existing research about possible cognitive biases of using web search for health related purposes such as prior belief [15] and rank [1, 7, 10] (top-down). While other codes have been added and modified as we explore the think-aloud transcribed data such as advertisements, statistics and studies (bottom-up). Applying the mixed method approach, we aim to discover the possible strategies participants apply when using search engine to answer a health related question.

When coding, we kept track of each coding occurrence to compute the frequency counts. Table 5 shows the list of codes with corresponding frequency counts (references). We use this quantitative method in order to identify which of the codes are more and less important for participants during the decision making process.

3.7 Participants

We obtained ethics approval from the Office of Research Ethics at our university. Next, we recruited participants using posters and email announcements to different graduate student email lists. As the user study involved an English language think-aloud process and, in order for participants to be able share their thoughts easier, one of the recruiting requirements was to have only native English speakers. All participants gave their informed consent. Following their participation, we debriefed all participants and provided them with the correct answers regarding the efficacy of the medical treatments. We paid participants $15. Participants were 16 students (7 male, 9 female) from different majors (7 from engineering and mathematics, 8 from arts and sciences and 1 from environment) with an age between 18 and 28 years old (37.5% less than 20, 56.25% between 20 and 25 and 6.25% greater than 25, with an average age of 21).

4 RESULTS AND DISCUSSION

In this workshop paper, we only explain the preliminary results of the concurrent think-loud part of the study. Future work will report the retrospective think-aloud and the questionnaire part in a full paper publication.

Table 2: Main results. Based on the decisions the 16 participants made, we compute the fraction of correct and harmful decisions. Fractions are shown along with their standard errors.

Results Bias	Fraction of Decisions	
	Correct	Harmful
Correct	0.67 ± 0.08	0.06 ± 0.03
Incorrect	0.32 ± 0.06	0.28 ± 0.06

Table 3: Statistical significance of independent variables.

Independent Variable	Dependent Variable	Pr(>Chisq)
Search Results Bias	Correct Decision	$\ll 0.001$
Search Results Bias	Harmful Decisions	$\ll 0.01$
Topmost Correct Rank	Correct Decision	0.8
Topmost Correct Rank	Harmful Decisions	0.05

Table 2 reports the fraction of correct and harmful decisions of the 16 participants corresponding to the Search Results Bias. We see that, similar to Pogacar et al. [10], results with bias towards correct information leads to an increased accuracy up to 67% while lowering harmful decisions to 6%. Conversely, results biased towards incorrect information reduces accuracy to 32% while increasing harmful decisions to 28%.

Table 3 reports the statistical significance of the search results bias and topmost correct rank on the correct and harmful decisions. Similar to Pogacar et al. [10], we find that the search result bias has a statistically significant effect on the fraction of correct decisions and harmful decisions. Due to the smaller sample, we find that the topmost correct rank has less of an effect on the correct and harmful decisions.

As verbalization makes people take longer time doing the search tasks (39 minutes average participation time in [10] compared to 65 minutes in current study), we expect people to be more conscious about their decisions and search results bias to have less or no effect on people's decisions. However, results demonstrated once again that search results have a potentially strong effect on people's decisions.

Pogacar et al. [10] as well as White and Hassan [16] demonstrated that participants have a strong bias towards believing that treatments are helpful. Looking at the current think-aloud data, we split the medical treatment types into "helpful", "unhelpful" and "inconclusive" treatments to further investigate this trend. Similar to prior work [10, 16], the results in table 4 show that helpful is the most frequent option people tend to answer during the study. Furthermore, participants are more likely to answer inconclusive more frequently than what Pogacar et al. [10] observed i.e., when thinking out loud, people tend to respond inconclusive more frequently than when not thinking out loud.

The coding process gives us some insights of the potential reasons why people are influenced with the search results even when the correct answer is always placed in higher ranks. Table 5 shows the number of participants mentioning each code and the total number of references for that corresponding code. The codes are

Table 4: Confusion matrices. This table shows the decisions made by the study participants regarding the efficacy of the 2 helpful and 2 unhelpful medical treatments.

Truth	Participants			Total
	Unhelpful	Helpful	Inconclusive	
Unhelpful	13	6	13	32
Helpful	5	18	9	32
Total	18	24	22	64

arranged in a descending ordered by the number of participants then references.

First, from the transcribed data, 14 out of 16 participants mentioned *Majority* with a total number of 36 mentions. Majority means that participants try to find out what most websites state about the treatment effectiveness or try to look for an agreement between them. If participants are exposed to results geared towards a specific direction, they end up being influenced by what the majority of the search results state. This finding explains why search result bias (in both this study and [10]) has a significant effect on people's decisions. Here, we provide examples of the majority effect from the think-aloud transcript with the participant number in parentheses:

> **(Participant 5)** *I'm going to say helps because a lot of people, like it was just, the vast number were in agreement.*
> **(Participant 6)** *So I'm seeing a lot of doctors recommending the melatonin pill. Yeah, I think this helps.*
> **(Participant 9)** *I think that's the common trend that we're seeing. So I'm going to submit and say that it does help.*

It is important to note that some people look at search results as individuals having opinions (Participants 5 & 6 in the above examples). They lean towards a specific direction because they believe that the majority of search results reflects the majority of opinions in real life which is a potentially dangerous misconception.

Further, we find that 45% of the total codes are about *authoritativeness* with 13 participants talking about it and a total of 153 references. Authoritativeness refers to the amount of reliability and trustworthiness towards specific content. We observed that participants talk about authoritativeness in three different ways: 40% of the time, people state that the content is not authoritative (negative authoritativeness), 34% of the mentions state that the content is trustworthy (positive authoritativeness) and the remaining 26% are about not being sure whether or not to trust the content (neutral authoritativeness). Bellow, we show some examples of each case from the think-aloud transcript:

> **(Participant 17)** *Health.com, I've seen it before, not really ... I don't really rely on it for information the first time I see it.*
> **(Participant 10)** *WebMD. It's a more trust worthy source, I think.*
> **(Participant 14)** *Okay. I don't really know what this website is. Medications for management of alcohol withdrawal.*

The high percentage of mentions about authoritativeness show the importance of this factor to participants when evaluating the effectiveness of the treatments. When Pogacar et al. [10] designed the user study, authors did not control for the authoritativeness of search results i.e., correct answers might be in non-authoritative web pages. Doing this, they potentially harm participants' performance especially with an incorrect search results bias. This might be another possible reason why people have been heavily influenced during the study.

Participants talk about many factors that define the quality of search results during the think-aloud. Concepts C3-6, C8, C10 and C13-14 in table 5 are all about quality. For example, 12 participants mention 20 times the statistical analysis and detailed research studies during the think-aloud process (C3) in order to evaluate the quality of information in the search results. Examples of such beliefs can be found in these bellow transcribed participations:

> **(Participant 12)** *...so this is explaining a study. Who had been given cinnamon reduced their blood sugar by 18 to 29 percent. Well that seems like some good numbers. So that's interesting. I think, based on that, I'd probably say that it helps because it had really evidence from a study.*
> **(Participant 15)** *So this looks like a research study, so I think it's pretty reliable.*

We, further note some notion about prior beliefs during the think-aloud (C9) where 5 participants mentioned this concept a total of 8 times. Bellow, we show some examples:

> **(Participant 16)** *And I was also taught from school that benezenes are harmful to health so though I might be bias I have this thought that benzene would not exactly help with certain health concerns.*
> **(Participant 3)** *So this Kurt Donsbach, PhD ... He will claim that it has no positive function at all, but I've heard different, so right away I'm not convinced by this page.*

We also coded the concept of rank where Table 5 shows that only 2 participants out of 16 mentioned rank a total of 6 times. We show an example from the think-aloud transcript bellow:

> **(Participant 19)** *I'll just go to the first link, even though it's wikiHow, it is the first link. I don't really know much about search engines, but I feel like the first link ... they're trying to give you the most helpful link. So I'll just open it, but still.*

Looking at the results, people rarely talk about rank when, in Pogacar et al. [10]"s study, authors showed that rank has a potential effect on people's decisions. A possible explanation is that people are unconsciously influenced with the higher ranked search results, however, they are not aware of its effect.

We know from Pogacar et al. [10]'s work that people have a bias towards believing that treatments are helpful. We further know, from the work of White et al. [15, 17], that people have a strong confirmation bias when using search engines. However, in the think-aloud transcription data and coding process, we fail to find any mention of such biases. Participants are not aware of these influences and are subconsciously being biased with such factors during the search.

5 CONCLUSION

Search result pages contain a large number of incorrect information when people perform online search about the effectiveness of medical treatments. With the effect of content bias, people are being influenced and potentially harmed. In order to build systems that better support people's decisions, we need to gain insights about strategies people use during this decision making process. Understanding the cognitive biases while using search engines to answer a health related question is a complex phenomenon, mainly because there is a large number of biases and unconscious factors effecting the decision making process. In this paper, we perform a think-aloud study where we ask participants to verbalize their thoughts while using search results to decide about the effectiveness of a medical treatment. Results revealed some strategies people use doing online search for health related topics.

We expected that the think-aloud process would lessen the effect of the search result bias, for people were asked to be conscious of their decision making process and carefully perform the task in front of a researcher. However, people were still significantly influenced by the misinformation, which shows how much search bias can affect people's decisions

Additionally, biased content leads people to believing this reflects real life opinions. The implications are profound when, for example, searching for cancer treatments on today's popular web search engines that might return a mix of correct and incorrect results.

Finally, when people use search engines to answer health questions, there are many factors that, unconsciously, effect their decision making. The results of the think-aloud study showed some examples of such biases (the helpfulness bias, the confirmation bias as well as rank).

ACKNOWLEDGMENTS

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada through its Discovery and CRD programs and in part by the Qatar Foundation for Education.

REFERENCES

[1] Ahmed Allam, Peter Johannes Schulz, and Kent Nakamoto. 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 (2014), e100.

[2] Elizabeth Charters. 2003. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal OLD* 12, 2 (2003).

[3] David Elsweiler and Markus Kattenbeck. 2019. Understanding credibility judgements for web search snippets. *Aslib Journal of Information Management* (2019).

[4] Lijun Feng, Martin Jansche, Matt Huenerfauth, and Noémie Elhadad. 2010. A comparison of features for automatic readability assessment. In *Proceedings of the 23rd international conference on computational linguistics: Posters*. Association for Computational Linguistics, 276–284.

[5] Susannah Fox and Maeve Duggan. 2013. Health online 2013. *Health* 2013 (2013), 1–55.

[6] Yongqi Gu. 2014. To code or not to code: Dilemmas in analysing think-aloud protocols in learning strategies research. *System* 43 (2014), 74 – 81. https://doi.org/10.1016/j.system.2013.12.011 Language Learning Strategy Research in the Twenty-First Century: Insights and Innovations.

[7] Alexander Haas and Julian Unkel. 2017. Ranking versus reputation: perception and effects of search result credibility. *Behaviour & Information Technology* 36, 12 (2017), 1285–1298.

[8] Hannu Kuusela and Paul Pallab. 2000. A comparison of concurrent and retrospective verbal protocol analysis. *The American journal of psychology* 113, 3 (2000), 387.

Table 5: This table shows the list of codes with their corresponding description. Each code is assigned a label C1-C16 that we use throughout the paper to refer to specific codes. The table also shows the number of participants mentioning a particular code, and the total number of references assigned to the code.

No	Name	Description	Participants	References
C1	*Majority*	The majority of the search results stating that the treatment helps or that the treatment does_not_help or looking for a consensus of different search results.	14	36
C2	*Authoritativeness*	The trustworthiness and reliability in the content of the search results page.	13	153
C3	*Statistics & Studies*	The presence of statistics, numbers and detailed research studies in the search results page.	12	20
C4	*Advertisements*	The presence of messages to promote or sell a product, service or idea in a search results page.	7	16
C5	*Date*	The date and time the search results page was first published to the public or the dates mentioned in the page content reflecting how old the information is.	7	15
C6	*References*	Having a list of sources that have been cited to support the information in the search results page.	7	12
C7	*Negative information*	Mentioning negative information about the treatment in the search results page such as listing the side effects or explaining the dangers of using the treatment etc.	6	15
C8	*Information representation*	The information related to the style of the content presented in the search results page such as list versus grid representation, colors, the page layout, capital letters and special characters etc.	5	18
C9	*Prior belief*	Trusting the information that agrees with our prior knowledge and disregarding facts that contradict with it, regardless of the actual truth [17].	5	8
C10	*Readability*	The style of writing and the quality of content being easy to read [4].	4	8
C11	*Relevance*	The relevance to the topic about the effectiveness of the medical treatment.	4	7
C12	*Past experience*	Having a prior experience with the topic (either the medical condition or the treatment) that may effect how much we trust the information in the search results page regardless of the factual correctness.	3	3
C13	*Text length*	The amount of text content in the search results page which might impact the reliability. For example, longer explanations might lead to higher levels of trust.	3	3
C14	*Images*	The presence of visuals in the search results page. The intuition behind this is that images might help better remember the information which may interfere with the decision making process.	2	6
C15	*Rank*	The order of search results in the SERP page that might effect the trustworthiness and reliability of the sources.	2	4
C16	*Social factor*	Relate the information about the topic to people we know. For example, whether a friend or a family member's opinion effects our preferences and decision making.	1	2
		Overall	16	326

[9] QSR International Pty Ltd. Version 12, 2018. *NVivo qualitative data analysis software.*

[10] Frances A Pogacar, Amira Ghenai, Mark D Smucker, and Charles LA 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.* ACM, 209–216.

[11] Neil J Salkind. 2010. *Encyclopedia of Research Design.* Vol. 1. SAGE.

[12] Thanh Tin Tang, Nick Craswell, David Hawking, Kathy Griffiths, and Helen Christensen. 2006. Quality and relevance of domain-specific search: A case study in mental health. *Information Retrieval* 9, 2 (2006), 207–225.

[13] MW Van Someren, YF Barnard, and JAC Sandberg. 1994. *The think aloud method: a practical approach to modelling cognitive.* Citeseer.

[14] Ryen White. 2013. Beliefs and biases in web search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval.* ACM, 3–12.

[15] Ryen W White. 2014. Belief dynamics in Web search. *Journal of the Association for Information Science and Technology* 65, 11 (2014), 2165–2178.

[16] Ryen W White and Ahmed Hassan. 2014. Content bias in online health search. *ACM Transactions on the Web (TWEB)* 8, 4 (2014), 25.

[17] Ryen W White and Eric Horvitz. 2015. Belief dynamics and biases in web search. *ACM Transactions on Information Systems (TOIS)* 33, 4 (2015), 18.