Exposure and Order Effects of Misinformation on Health Search Decisions

Mustafa Abualsaud
School of Computer Science
University of Waterloo
m2abuals@uwaterloo.ca

Mark D. Smucker
Department of Management Sciences
University of Waterloo
mark.smucker@uwaterloo.ca

ABSTRACT

Online search engines can return inaccurate information of ineffective and sham treatments that may influence people into making harmful decisions. In a user study conducted by Pogacar et al. [9], they showed the impact of search result bias to people's ability to determine the correct efficacy of health interventions, and that people's accuracy is very low when search results are biased towards incorrect information. While their study provided evidence that search results have the potential to inflict harm on users, it is still unclear what caused people to make their final decisions. In this paper, we looked at click sequences and their relation to different types of health decisions. A clear pattern from the sequence data emerges. First, users exposed to a larger amount of correct information are more likely to make non-harmful and correct health decisions than those exposed to a larger amount of incorrect information. Second, it appears that a recency order effect is exhibited by users examining a sequence of documents, i.e. information presented in the last clicked document has more influence on their final decision than documents viewed early in the sequence.

1 INTRODUCTION

Pogacar et al. [9] conducted a controlled study to measure the effect of misinformation in search results on people's ability to determine the correct efficacy of health treatments. Their results show that people's decisions can indeed be negatively influenced by the misinformation presented to them. When users were presented with search results biased towards incorrect information, users' accuracy in determining the true efficacy of treatment was a low 23%. Their accuracy climbed to 65% when presented with search results biased towards correct information. While their experiment provided evidence that users can be influenced by misinformation from search results, little is known on what user interactions can be used as signals to predict their final judgments. Click sequences, as an example, can be rich in useful information that can be used to better understand overall exposure to correct and incorrect information, as well as biases that may have influenced a user's final decision.

In this paper, we examine click data from the controlled experiment by Pogacar et al. [9] to understand users' clicking behavior and its relation to their final decision. Our goal is to identify patterns in clicking behavior that can be indicative of a user's final decision. We measured the fraction of time a user was exposed to correct information to determine its relation to harmful and non-harmful health decisions.

We show that:

- The fraction of time people are exposed to misinformation can lead users to bad decisions. People exposed to documents with false information are more likely to make harmful decisions than those who spend their time on documents with correct information, i.e. more time spent exposed to correct information leads to better decisions. We also found that an order effect may have been exhibited by users.
- Order effects can be categories as primacy (where users are more influenced by items presented first in the sequence) and recency (where users are influenced by items presented last) [7]. We found that a recency order effect is more indicative of their final decision than a primacy order effect. In other words, users may be more influenced by documents read later in the sequence than documents read early on, and ending the search process with a document containing false information is more likely to produce a harmful decision than ending with a document with correct information.

We briefly describe the Pogacar et al. [9] study and data in Section 3. We show and discuss our results from the click data in more detail in Section 4 and conclude with future work in Section 5.

2 RELATED WORK

There have been a number of works in understanding health search queries in online search engines [10–13]. Of particular interest is the work of White and Horvitz [13] that investigated changes in users' beliefs about the efficacy of medical interventions to health issues. An example of a medical intervention for a health issue from [13] is "Does melatonin help jet-lag". White and Horvitz [13] found that people spent less time on documents that are found to be contradicting to their existing beliefs of the efficacy of the medical intervention. In another work by White and Hassan [12], they have shown that people often formulate their queries towards "helpful" information, entering keywords such as "help", "cure", or "treat" that can bias search results. While some medical intervention can be indeed helpful towards a health issue, others can be ineffective or may cause harmful unintended effects. Search result bias and people's existing beliefs can lead people to be exposed to incorrect information and may as a result induce people to harmful health decisions. Pogacar et al. [9] have directly investigated the impact of search result bias towards people's accuracy in determining the true efficacy of medical interventions. In a controlled user study where the authors purposely controlled the amount of correct information in a search engine result page (SERP), they have found that

ROME 2019, July 25, 2019, Paris, France

search results can have a statistically significant and strong effect on people's ability to make correct decisions. Accuracy of people's decisions when search results were biased towards incorrect information was only 23%. The work of White and Hassan [12] and others [9, 10] highlight the importance of building search engines that aid people in making better health decisions.

Another area of research explores helping users make better decisions. Decision-making research has long shown that cognitive biases can have an adverse effect on users' decisions [3, 4, 6]. Coiera and Lau [2] investigated debiasing techniques to increase users' accuracy in making correct health decisions. Their work is focused on anchoring and order cognitive biases experienced during online web search and whether they can be corrected using specifically designed debiasing interfaces. In a user study of over 200 participants, they show that a specifically designed interface for debiasing order effect was successful. However, there was not any significant improvement in users' accuracy using the interface. The anchoring bias persisted in the anchoring debiasing interface but increased the accuracy of subjects who had answered incorrectly prior to using the interface.

3 DATA

We use click data from [9]'s experiment to further understand the behavior of users and their decisions. Their experiment involved 60 participants interacting with a controlled SERP to determine the efficacy of medical treatments towards health issues. We briefly describe their experiment settings below and refer the reader to their full paper for the complete details.

3.1 Experiment Tasks and Topics

Pogacar et al. asked sixty participants to complete 10 decision tasks. Each decision task involved making a decision on the efficacy of a medical intervention towards a health issue. An example topic of a decision task is "Does Cinnamon help Diabetes?" (See Table 7 for all topics), where the medical intervention (or treatment) is cinnamon and the health issue is diabetes. The selected 10 topics were part of 249 topics used in a research study by White and Hassan [12]. White and Hassan [12] judged the effectiveness of each medical intervention in each topic by reading the Cochrane Review [1, 5] associated with it. A Cochrane review is a systematic review conducted by medical experts to promote evidence-informed health decision-making for patients and health practitioners and is internationally recognized as a source of high-quality medical information. White and Hassan [12] judged each topic as either helps, inconclusive, or does not help. Pogacar et al. modified and expanded on White and Hassan's definition of these categories and described them to their study participants as:

- **Helps**: The medical treatment **helps** if the treatment is effective and has a direct positive influence on the specified illness.
- **Inconclusive**: The effectiveness of a medical treatment is **inconclusive** if medical professionals are still unsure if the treatment will have a positive, negative or no influence on the specified illness.
- **Does not help**: The medical treatment **does not help** if the treatment is ineffective and either has no effect or has a direct negative influence on the specified illness.

Pogacar et al. [9] selected 5 *helpful* and 5 *non-helpful* for their experiment. A helpful topic is a topic judged as "Helps" by White and Hassan [12], and similarly for non-helpful. In [9], participants were asked to provide their answers by selecting one of the three categories mentioned above, indicating that their final decision of the efficacy of the treatment is of that selected category.

3.2 Search Results

Two of the 10 study tasks in [9] were control tasks. In the control tasks, participants were asked to provide their answers without any assistance, i.e. without using a search engine. In the remaining 8 tasks, participants were presented with the topic along with a controlled search engine result page of 10 documents. All 10 documents were related to the topic, but they were either biased towards providing correct or incorrect information. Correctness of the document information is determined by whether it matches the conclusion of the Cochrane review. For example, any document with information that matches or contradicts with the Cochrane review is considered to be a correct or incorrect document, respectively.

3.2.1 SERP Bias. The SERP in each task was controlled to either contain 8 or 2 documents with correct information, with the remaining documents being documents with incorrect information. Each topic had a pool of 8-10 correct and 8-10 incorrect documents that were used to generate the search results in a randomized fashion. The rationale behind this amount of bias is that it is actually similar to that found in actual search engines, as shown by White and Hassan [12] who investigated content bias in Bing search engine.

3.2.2 Rank of First Correct Document. In addition to controlling the number of correct documents in the SERP, Pogacar et al. [9] also controlled the position of the first correct document to be either at rank 1 or 3. The authors choose ranks 1 and 3 based on previous eye-tracking studies on attention behavior.

3.3 Study Design

3.3.1 Search Interface. Figure 1 shows an example of the SERP. The topic information is displayed at the top of the page and includes the definition of both the medical treatment and the health issue. The right-hand side includes the 3 categories the user is expected to select as their answer after having interacted with the search result. Users can click on a document to view its content and are allowed to click at as many documents as they wish.

3.3.2 Experimental Conditions Overview. In short, the experiment conditions are labeled and described below:

- **No SERP**: Participants were asked to provide their answer without any assistance, i.e. they were not provided with search results.
- **Good-1**: Participants were provided 10 search results, 8 of which contained correct information and 2 incorrect information. The rank of the first correct document is 1.
- **Good-3**: As above. The first correct document is 3, i.e. the first 2 documents contain incorrect information.

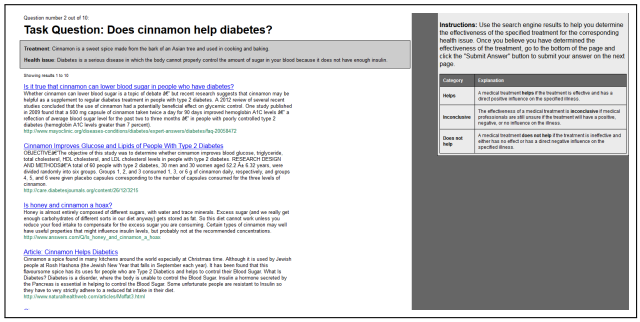

Figure 1: The user interface of the experiment. The top of the page includes information about the medical treatment and the health issue. The right hand side includes information on the definitions of treatment efficacy: Helps, Does not Help, and Inconclusive. Users can click on a document to view its content.

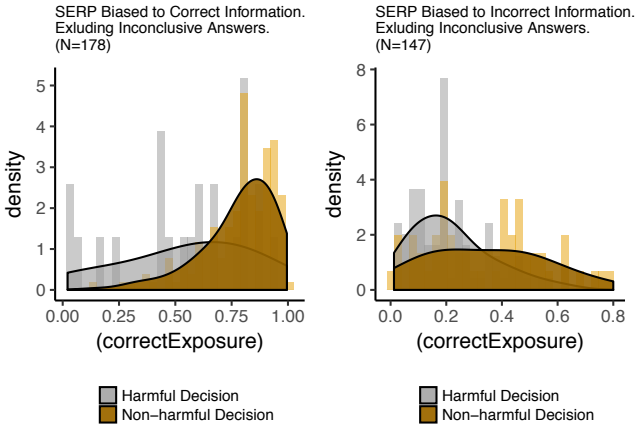

Figure 2: Density plot of amount of exposure to correct information. Plots split by type of decision.

- **Bad-1**: 8 of the documents contained incorrect information, and only 2 contained correct information. The rank of the first correct document is 1.
- **Bad-3**: As above. The rank of the first correct document is 3.

In our analysis, we exclude results from the **No SERP** condition.

3.3.3 Balanced Design. The study was fully balanced in terms of topics and experiment conditions using a 10x10 Graeco-Latin square. Each participant completed each experiment condition twice, and each of the helpful and non-helpful topics had an equal and systematic balance of the experimental conditions.

3.3.4 Decisions types. We follow the same method of categorizing user decisions as in Pogacar et al. [9]. Decisions are categorized in two categories:

Harmfulness

- **Harmful Decision**: a decision is considered harmful when it is opposite of the true efficacy of the medical intervention, i.e. when a user answers with "Helps" when the true answer is "Does not help", or vice-versa.
- **Non-harmful Decision**: a decision is considered non-harmful when it matches the true efficacy of the medical intervention, or when a participant chooses "Inconclusive" as their answer.

Correctness

- **Incorrect Decision**: a decision that does not match the true efficacy of the medical intervention.
- **Correct Decision**: matches the true efficacy of the medical intervention.

4 RESULT AND DISCUSSION

4.1 General Statistics

Table 1 provides general statistics of the experiment that were not previously reported. The average number of clicks is around 3.8 documents in conditions where the first correct document is placed at rank 1. The average number of clicks is slightly higher in conditions where the first correct document is at rank 3. As the rank of the first correct document decreases from 1 to 3, the average time spent on correct documents also decreases regardless of the experimental condition. Whether the search result is biased towards correct or incorrect information, users are less exposed to correct information when it is lower in the search page. This has important implications on allowing a user to make better decisions, as less exposure to correct information could influence users into believing incorrect information as the truth, thus making harmful or incorrect decisions.

4.2 Exposure to Correct Information

Although eye-tracking data is unavailable, we do know the number of incorrect and correct documents for each condition: 80% correct documents in the conditions where the SERP is biased towards correct information, and 20% when biased towards incorrect information. We use these percentages along with the time spent reading correct documents to measure the fraction of time a user was exposed to correct information, defined as:

$$correctExposure = \frac{(time@SERP \times B) + time@CorrectDocs}{TotalTime}$$

where B is the percentage of correct items in the SERP.

Figure 2 shows the density plot of the amount of correct exposure grouped by the type of decision. It is clear that when the search results are biased towards correct information, more users are exposed to correct information. Users exposed to a larger amount of correct information appear to have made more non-harmful decisions than users who are exposed to a smaller amount of correct information. The opposite is true in conditions where the search results are biased towards incorrect information. Users who are exposed to a larger amount of incorrect information are more likely to make a harmful decision than those who were exposed to a larger amount of correct documents.

4.3 Number of Clicks

Table 2 shows the fraction of decisions based on the total number of clicks. Although a little noisy, we observe that the fraction of

	All	Condition			
		Good-1	Good-3	Bad-1	Bad-3
# Clicks	1673	400	422	414	437
# Unique clicks	1652	396	417	412	427
Avg. # clicks	3.92	3.81	4.06	3.80	4.01
Avg. # clicks - correct docs.	1.77	2.72	2.45	1.03	0.88
Avg. # clicks - incorrect docs.	1.71	0.62	1.07	2.42	2.76
Avg. time on docs.	108.97	104.26	105.11	110.28	115.88
Avg. time on correct docs.	54.51	87.01	66.79	35.86	30.14
Avg. time on incorrect docs.	54.46	17.25	38.32	74.42	85.74
Avg. time at SERP	29.00	26.88	33.98	28.01	27.15

Table 1: General Statistics of the dataset (60 participants).

# Clicks	Cov.	N	% Harmful	% Non-harmful	% Incor.	% Correct
0	0.11	53	0.28	0.72	0.68	0.32
1	0.13	62	0.27	0.73	0.47	0.53
2	0.18	84	0.30	0.70	0.58	0.42
3	0.19	89	0.24	0.76	0.57	0.43
4	0.12	58	0.21	0.79	0.62	0.38
5	0.09	45	0.24	0.76	0.60	0.40
6	0.06	31	0.16	0.84	0.45	0.55
7	0.03	14	0.07	0.93	0.50	0.50
8	0.02	11	0.09	0.91	0.45	0.55
9	0.01	4	0.25	0.75	0.25	0.75
10	0.06	29	0.17	0.83	0.45	0.55

Table 2: Fraction of decisions based on total amount of unique clicks. N indicates the total number of observations.

correct decisions is higher when users click more items (5 or more) than when they click fewer items or click nothing at all. When users do not click on any documents or click on 2 or fewer documents, the fraction of harmful decisions are higher than when they click at 3 or more documents. A possible explanation for this result, as also noted by Pogacar et al., is that when users click on more documents, they are exposed to more information that can lead them to the correct answer.

4.4 Order Effect

Order effect (or order bias) is the effect of the temporal order of information presented to a user and its influence on their final judgment [8]. Search results are often presented in a ranked list in which users process from top down, and thus an order effect from search results can occur. The order effect can be subdivided into primacy and recency effects [7]. In primacy, a user's final judgment is more influenced by information presented earlier in the sequence. In the recency effect, users are more influenced by the information presented later in the sequence.

We found that the effect of incorrect information on people's decision is stronger when it comes at the end of a click sequence than when it comes at the start. We measure the effect size of incorrect information as the change in the accuracy of judgments compared to correct information. For example, in Table 3, we see that when users interact with a SERP biased towards correct information, if the first click is on correct information, the accuracy is 0.71 but when the first click is on incorrect information, the accuracy drops to 0.62, which is an effect size of $0.71 - 0.61 = 0.09$. The effect size of incorrect information on accuracy is between 0.08-0.09 when it occurs as the first click. When incorrect information is the last click, it has an effect size of between 0.15-0.23. The larger effect size of the last vs. first click shows that incorrect information has a larger recency effect than primacy effect.

Although the data indicates that recency effect is a possible factor, we cannot eliminate prior beliefs and knowledge influence in final decisions [13]. A user whose prior beliefs matches with their last click can be the reason they have made their particular decision. Unfortunately, Pogacar et al. [9] did not collect any information about users' prior beliefs on the treatment and so we are unable to investigate this further. Another possible influencing factor is their time reading incorrect snippets from the search result. Without eye-tracking data, we are unsure which document snippets were read by the user and how much is spent reading them, as the sequence of items read in the snippets may also influence their final decision.

	Click	N	% Harmful	% Correct
SERP Biased to Correct Information (Good-1, Good-3)				
First Click	✔ (correct information)	122	0.07	0.71
	✗ (incorrect information)	87	0.14	0.62
Last Click	✔ (correct information)	178	0.05	0.70
	✗ (incorrect information)	38	0.32	0.55
	No Clicks	31	0.10	0.45
SERP Biased to Incorrect Information (Bad-1, Bad-3)				
First Click	✔ (correct information)	94	0.29	0.29
	✗ (incorrect information)	124	0.42	0.21
Last Click	✔ (correct information)	59	0.20	0.41
	✗ (incorrect information)	159	0.42	0.18
	No Clicks	22	0.55	0.14

Table 3: Fraction of decisions based on first and last click, or no clicks. Table split by whether condition is biased towards correct/incorrect information. ✔ and ✗ indicate a click on a document with correct or incorrect information, respectively.

4.5 Clicks Sequences

Our goal in this section is to identify which click sequences are more likely to result in a harmful decision. We define a click sequence to be a sequence of document clicks with either correct or incorrect information.

4.5.1 Possible sequences of specific lengths. Table 4 shows all possible sequences of length 0 to 3 sorted in a descending order by their fraction of correct decisions. When users click on a single correct document, the top sequence resulting in the most correct decisions is simply a single correct document, which gives users a decision accuracy of 0.69. Its fraction of harmful decisions is a low 0.08. The fraction of harmful decisions increases to 0.56 when users click on a single incorrect document. While the coverage is low, these fractions indicates that users are more likely to make harmful decisions when they are exposed to an incorrect document as opposed to a correct document.

When users click on 2 documents, the sequence with two correct clicks (✔→ ✔) has the highest fraction of users making a correct decision (0.71). Interestingly, the second most highest accuracy

(0.56) is for the sequence that starts with an incorrect document and ends with a correct document (✗→ ✔). As we observed in Section 4.4, ending with a correct document can lead users to better decisions. The third most accurate (0.28) two-click sequence is ✔→ ✗and the lowest accuracy (0.20) two-click sequence is the sequence with only incorrect document clicks (✗→ ✗). Both sequences end with an incorrect document and have the highest fraction of harmful decisions among the four sequences.

A clear pattern emerges from analyzing the two-click sequences that is worth mentioning: 1) as more correct documents are viewed, the chances of making a correct (or non-harmful) decision increases, and 2) the last document viewed has more influence on the decision than the first document viewed. This pattern also seems to appear in sequences of length 3. We clearly see that the top 3 sequences for producing accurate decisions are sequences that have more correct than incorrect documents. Although the sequence "✔→ ✗→ ✔" has two correct documents and is not ranked higher than some sequences with a single correct document, its fraction of harmful decisions is 0, i.e. most users have either provided a correct answer or answered with "inconclusive". We also observe that 4 out of the 5 top sequences for producing accurate decisions are sequences ending with a correct document, and these sequences have accuracies of 43% or greater.

While the click sequences have an effect on the decisions made, the search engine result page itself has an effect. As shown in Table 5 and 6, the user interactions with no clicks show that exposure to a larger amount of incorrect information in the search results snippets can lower decision accuracy. Thus, the effect of these click sequences is somewhat overstated, for clicks sequences with a larger amount of correct information also come from search engine result pages biased towards correct information, and vice versa for click sequences of incorrect information.

4.5.2 Most common sequence in each experimental condition. Table 5 and 6 display the most common sequences exhibited by participants in each experimental condition. Sequences are sorted by their coverage within the condition's dataset. In all conditions, an empty sequence (i.e. No Clicks) appears to be always in the top two most common sequences with respect to coverage. When people are presented with search result biased towards incorrect information, making a decision without clicking on any documents will most likely result in a harmful decision. The opposite is true when presented with search results that are biased towards correct information. This again shows the importance of being exposed to correct information for making good decisions.

We also notice the effect of controlling the first correct item in the search result. In conditions where the first correct document is at rank 3 (i.e. first 2 documents are incorrect) most users start by clicking at an incorrect document. Users clicked at many correct documents when the search results were biased towards correct information, increasing their exposure to correct information. In the experimental condition where the results are biased towards incorrect information and the first correct document is at rank 3 (Bad-3), the average number of clicks at correct documents is 0.88 (Table 1) and 2.76 at incorrect documents. Compared to the experimental condition Bad-1, users in Bad-1 were exposed to a little more correct information than in Bad-3, which explains how the fraction of correct decision is higher in Bad-1 than in Bad-3 (Table 1 in [9]).

5 CONCLUSION

We used click data from the Pogacar et al. [9] study to look into clicking behavior that may have influenced people to make harmful decisions. By measuring the fraction of time users spent viewing incorrect and correct information, we found that users who are exposed to a larger amount of correct information make a larger fraction of correct and non-harmful decisions than users that are exposed to a larger amount of incorrect information. This indicates the importance of reducing users exposure to misinformation in search results. Our second finding show that a recency order effect (influence of documents viewed last in the search process) has a stronger influence on users decisions than a primacy effect (influence of documents shown first in the search process). In other words, ending the search process with a documenting containing misinformation is more likely to cause a harmful decision than ending with a document with correct information. While the data points towards such effect, prior belief and knowledge may also be a contributing factor. For future work, we plan to use eye-tracking technology to look more into eye-sequences and reading behavior of users making harmful and non-harmful decisions, and investigate how a user's prior belief may play a role in influencing their sequence of interactions and their final decisions.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (Grants CRDPJ 468812-14 and RGPIN-2014-03642).

REFERENCES

[1] A Cipriani, TA Furukawa, and C Barbui. 2011. What is a Cochrane review? *Epidemiology and psychiatric sciences* 20, 3 (2011), 231–233.
[2] Enrico W. Coiera and Annie Y.S. Lau. 2009. Can Cognitive Biases during Consumer Health Information Searches Be Reduced to Improve Decision Making? *JAMIA* 16, 1 (01 2009), 54–65. https://doi.org/10.1197/jamia.M2557
[3] Pat Croskerry. 2002. Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Academic Emergency Medicine* 9, 11 (2002), 1184–1204.
[4] Arthur S Elstein. 1999. Heuristics and biases: selected errors in clinical reasoning. *Academic Medicine* (1999).
[5] Julian PT Higgins, Sally Green, et al. 2008. Cochrane handbook for systematic reviews of interventions. (2008).
[6] Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. 1982. *Judgment under uncertainty: Heuristics and biases.* Cambridge university press.
[7] Abraham S Luchins. 1958. Definitiveness of impression and primacy-recency in communications. *The Journal of Social Psychology* 48, 2 (1958), 275–290.
[8] Nancy Pennington and Reid Hastie. 1988. Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, 3 (1988), 521.
[9] Frances A. Pogacar, Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. 2017. The Positive and Negative Influence of Search Results on People's Decisions About the Efficacy of Medical Treatments *(ICTIR)*. ACM, New York, NY, USA, 209–216. https://doi.org/10.1145/3121050.3121074
[10] Georg P. Schoenherr and Ryen W. White. 2014. Interactions Between Health Searchers and Search Engines *(SIGIR)*. ACM, New York, NY, USA, 143–152. https://doi.org/10.1145/2600428.2609602
[11] Ryen White. 2013. Beliefs and Biases in Web Search *(SIGIR)*. ACM, New York, NY, USA, 3–12. https://doi.org/10.1145/2484028.2484053
[12] Ryen W. White and Ahmed Hassan. 2014. Content Bias in Online Health Search. *ACM Trans. Web* 8, 4 (Nov. 2014), 25:1–25:33. https://doi.org/10.1145/2663355
[13] Ryen W. White and Eric Horvitz. 2015. Belief Dynamics and Biases in Web Search. *ACM Trans. Inf. Syst.* 33, 4, Article 18 (May 2015), 46 pages. https://doi.org/10.1145/2746229

# Clicks	Coverage	Click Sequence	N	Sequence Coverage	% Harmful	% Non-Harmful	% Incorrect	% Correct
0	0.11	-	53	0.11	0.28	0.72	0.68	0.32
1	0.13	✔	36	0.08	0.08	0.92	0.31	0.69
		✗	25	0.05	0.56	0.44	0.72	0.28
2	0.17	✔→ ✔	24	0.05	0.00	1.00	0.29	0.71
		✗→ ✔	16	0.03	0.12	0.88	0.44	0.56
		✔→ ✗	18	0.04	0.44	0.56	0.72	0.28
		✗→ ✗	25	0.05	0.56	0.44	0.80	0.20
3	0.19	✔→ ✔→ ✔	14	0.03	0.07	0.93	0.36	0.64
		✗→ ✔→ ✔	19	0.04	0.16	0.84	0.47	0.53
		✔→ ✔→ ✗	9	0.02	0.22	0.78	0.56	0.44
		✗→ ✗→ ✔	14	0.03	0.36	0.64	0.57	0.43
		✔→ ✗→ ✔	11	0.02	0.00	1.00	0.64	0.36
		✗→ ✔→ ✗	6	0.01	0.33	0.67	0.67	0.33
		✔→ ✗→ ✗	12	0.02	0.50	0.50	0.75	0.25
		✗→ ✗→ ✗	6	0.01	0.5	0.5	1.00	0.00

Table 4: Fraction of decisions based on number of clicks and sequence. Sequences sorted by fraction of correct decisions. Coverage indicates the percentage of observations found in the dataset. N indicates the total number of observations.

Click Sequence	N	Coverage	% Harmful	% Correct
✔->✔	19	0.16	0.0	0.74
No Clicks	15	0.12	0.0	0.47
✔	11	0.09	0.18	0.82
✔->✔->✔	10	0.08	0.0	0.7
✔->✔->✗->✔	6	0.05	0.0	0.67
✔->✔->✔->✔	5	0.04	0.0	0.8
✔->✔->✗	4	0.03	0.5	0.5
✔->✗->✔	4	0.03	0.0	0.5
✔->✗	4	0.03	0.25	0.75
✗	3	0.03	0.33	0.67

(a) Rank of top most correct is 1 (Good-1).

Click Sequence	N	Coverage	% Harmful	% Correct
No Clicks	16	0.13	0.19	0.44
✗->✔->✔	14	0.12	0.07	0.57
✔	10	0.08	0.1	0.7
✗->✗->✔->✔->✔	8	0.07	0.12	0.62
✗->✔	8	0.07	0.12	0.62
✗->✗->✔->✔->✔->✔->✔->✔->✔->✔	7	0.06	0.0	0.86
✗->✗->✔	5	0.04	0.2	0.6
✗->✔->✔->✔	5	0.04	0.0	0.8
✔->✔->✔	4	0.03	0.25	0.5
✗->✔->✔->✔->✔	4	0.03	0.0	0.25

(b) Rank of top most correct is 3 (Good-3).

Table 5: Top 10 most common click sequences and their resulting fraction of decisions. SERP biased to correct information.

Click Sequence	N	Coverage	% Harmful	% Correct
No Clicks	11	0.09	0.45	0.09
✔->✗->✗->✗	11	0.09	0.45	0.09
✔->✗->✗	10	0.08	0.5	0.3
✔	10	0.08	0.0	0.5
✗->✗	9	0.07	0.44	0.22
✔->✗	8	0.07	0.62	0.25
✗	8	0.07	0.38	0.12
✔->✗->✔	5	0.04	0.0	0.2
✗->✔	4	0.03	0.0	0.25
✔->✗->✗->✗->✗	3	0.03	0.33	0.33

(a) Rank of top most correct is 1 (Bad-1).

Click Sequence	N	Coverage	% Harmful	% Correct
✗->✗	11	0.09	0.55	0.18
No Clicks	11	0.09	0.64	0.18
✗	10	0.08	0.8	0.2
✗->✗->✔	7	0.06	0.57	0.43
✗->✗->✗	5	0.04	0.4	0.0
✔	5	0.04	0.0	0.8
✔->✗	5	0.04	0.2	0.0
✗->✔->✔	4	0.03	0.5	0.25
✗->✔	3	0.03	0.33	0.67
✗->✗->✗->✔	3	0.03	0.33	0.0

(b) Rank of top most correct is 3 (Bad-3).

Table 6: Top 10 most common click sequences and their resulting fraction of decisions. SERP biased to incorrect information.

Topic (Cochrane ID)	Health Efficacy
Do benzodiazepines help alcohol withdrawal? (14651858.CD005063.pub3)	Helpful
Do sealants prevent dental decay in the permanent teeth? (14651858.CD001830.pub4)	Helpful
Does caffeine help asthma? (14651858.CD001112.pub2)	Helpful
Does melatonin help treat and prevent jet lag? (14651858.CD001520)	Helpful
Does surgery help obesity? (14651858.CD003641.pub3)	Helpful
Does traction help low back pain? (14651858.CD003010.pub5)	Unhelpful
Do insoles help back pain? (14651858.CD005275.pub2)	Unhelpful
Does cinnamon help diabetes? (14651858.CD007170.pub2)	Unhelpful
Do probiotics help treat eczema? (14651858.CD006135.pub2)	Unhelpful
Do antioxidants help female subfertility? (14651858.CD007807.pub2)	Unhelpful

Table 7: List of topics used in Pogacar et al. [9]. Health efficacy is judged by White and Hassan [12] by reading the Cochrane review associated with the topic. The Cochrane source ID is in parentheses.