

How do Clinicians Search For and Access Biomedical Literature to Answer Clinical Questions?

Annie Y.S. Lau, Enrico Coiera

Centre for Health Informatics, University of New South Wales, Australia

Abstract

This paper presents a retrospective data analysis on how 75 clinicians searched for and accessed biomedical literature from an online information retrieval system to answer six clinical scenarios. Using likelihood ratio measures to quantify the impact of documents on a decision, and a graphical representation to model clinicians' journeys of accessing documents, this analysis reveals that clinicians did not necessarily arrive at the same answer after having accessed the same document, and that documents did not influence clinicians in the same manner. A possible explanation for these phenomena is that people experience cognitive biases during information searching which influence their decision outcome. This analysis raises the hypotheses that people experience the anchoring effect, order effects, exposure effect and reinforcement effect while searching for information and that these biases may subsequently influence the way decisions are made.

Keyword:

evidence journey, information retrieval, cognitive bias, decision making, likelihood ratio, clinicians

Introduction

Increasingly, information searching plays an important part in health consumers' decision making [1] and clinicians' practice of evidence-based medicine [2]. Decisions are improved by better access to relevant information, and searching for documents on the Web is increasingly an important source of that information [3]. Yet, studies have confirmed that clinicians do not always achieve optimal results when using information retrieval systems [4]. While much research focuses on the design of retrieval methods that identify potentially relevant documents, there has been little examination of the way retrieved documents then shape real-life decision making [5]. To develop information retrieval systems that actively support decision making, it is necessary to understand the complex process of how people search for and review information when making decisions [6].

To understand how people search for and use information to make a decision, it is important to understand the "journey" in which people undertake to arrive at the decision. Taking the definition of evidence to be "the body of observed, reported or research-derived data that has the

potential to inform decision making" [7], an "evidence journey" is the process that describes an individual accessing different pieces of information retrieved from an online information retrieval system to make a decision. The notion of the evidence journey is similar to Bates' *berrypicking* metaphor, or *bit-at-a-time* retrieval, where "a query is satisfied not by a single final retrieved set, but by a series of selections of individual references and bits of information at each stage of the ever-modifying search" [8].

There is a body of literature looking at how people use information retrieved from search engines to inform their decision making. An example from the information science literature is the model of document use proposed by Wang and colleagues [9,10]. Based on a longitudinal study of academic researchers' use of documents retrieved from online databases during a research project, they proposed that document use is a decision making process and people do not necessarily use the same criteria to select, read or cite documents. Gruppen and colleagues, reporting in the medical decision making literature, suggested that information gathering and selection are more problematic than information integration and use [11]. Based on a study examining how first year house officers select information to make a diagnosis, they found that subjects selected the optimal information in only 23% of the cases but were able to use the selected information to make a diagnosis over 60% of the cases. They suggested that physicians appear to have difficulties recognising the diagnostic value of information, which results in making decisions based on diagnostically weak information.

This study analyses the evidence journeys that clinicians undertake to answer clinical questions. It graphically models the way clinicians searched for and accessed biomedical literature to make clinical decisions and identifies factors in the evidence journey that may have influenced the decision making process.

Methods

Data description

A retrospective analysis was constructed on a dataset of 75 clinicians' search and decision behaviours (44 doctors and 31 clinical nurse consultants), who answered questions for 8 real-life clinical scenarios within 80 minutes in a con-

Selected for best paper award.

trolled setting at a university computer laboratory [12]. Scenarios were presented in a random order, and subjects were asked to record their answers and level of confidence for each scenario. Subjects were then presented with the same scenarios and asked to use an Internet search engine to locate documentary evidence to support their answers.

Subjects recorded their pre- and post-search answers to each question, their confidence in these answers and their confidence in the evidence they had found using the search engine. There were four options to answer each question: yes, no, conflicting evidence and don't know. Confidence was measured by a 4 point Likert scale from "very confident" to "not confident". In addition, subjects recorded any change in answer or confidence from their pre-search response and identified which documents influenced their decision. They were asked to work through the scenarios as they would within a clinical situation and not spend more than 10 minutes on any one question.

Data from only six scenarios for which a correct answer could be identified were analysed (scenario questions are described in Table 1). The unit of measure used in the analysis is a search session, which is "the entire series of queries by a user" [13] to answer one question.

Table 1 - Clinical questions in the scenarios presented to subjects [12]

Question (scenario name)	Expected correct answer
Does current evidence support the insertion of tympanostomy tubes in child with normal hearing? (Glue ear)	No, not indicated
What is the best delivery device for inhaled medication to a child during moderate asthma attack? (Asthma)	Spacer (holding chamber)
Is there evidence for the use of nicotine replacement therapy after myocardial infarction? (MI)	No, use is contraindicated
Is there evidence for increased breast and cervical cancer risk after IVF treatment? (IVF)	No evidence of increased risk
Is there evidence for increased risk of SIDS in siblings of baby who died of SIDS? (SIDS)	Yes, there is an increased risk
What is the anaerobic organism(s) associated with osteomyelitis in diabetes? (Diabetes)	Peptostreptococcus, Bacteroides

Document likelihood ratio

Since a document may be influencing some subjects to answer in different ways, a quantitative measure is needed to model the impact a document may have on a decision. One simple method is to associate a document with the fre-

quency of correct and incorrect decisions made after having accessed the document. This leads to the idea of using a likelihood ratio (LR) to calculate a ratio of the frequency that accessing a document is associated with a correct answer rather than with an incorrect answer, $P(\text{AccessedDoc}|\text{Correct}) / P(\text{AccessedDoc}|\text{Incorrect})$ (Equation 1). The LR thus measures the impact a document has in influencing a subject towards a specific answer. Documents with a LR > 1 are most likely to be associated with a correct answer and a LR < 1 with an incorrect answer.

To calculate the likelihood ratio, the sensitivity and specificity of a document with respect to an answer are calculated. The *sensitivity*, or true positive rate, of a document is the frequency with which the document being accessed correlated with a correct answer being provided post-search (Equation 2). The false negative rate, the frequency with which access of a document correlated with an incorrect answer, was also calculated. The *specificity*, or true negative rate, is one minus the false negative rate (Equation 3). The sensitivity and specificity values were calculated based upon the frequency with which a document was accessed for each scenario.

$$\text{Likelihood ratio} = \frac{P(\text{AccessedDoc} | \text{Correct})}{P(\text{AccessedDoc} | \text{Incorrect})} = \frac{\text{Sensitivity}}{1 - \text{Specificity}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{No. of correct post-search answers where document was accessed}}{\text{Total no. of post-search correct answers}} \quad (2)$$

$$1 - \text{Specificity} = \frac{\text{No. of incorrect post-search answers where document was accessed}}{\text{Total no. of post-search incorrect answers}} \quad (3)$$

Results

Overall, subjects made 1761 searches and accessed 1873 documents across the 400 search sessions for six scenarios. In a search session, subjects took on average 404.75 seconds (standard deviation (SD): 590.824), made 4.32 (SD: 4.002) searches and accessed 4.65 (SD: 3.670) documents to complete a question.

Across the six scenarios, most subjects improved their answers after searching. There are subjects, as reported in [12], who had a wrong answer before searching and a right answer after searching, wrong-right (WR: 37%), followed by those who never answered correctly, wrong-wrong (WW: 30%), then those who answered correctly before and after searching (RR: 26%), and those who went from right to wrong (RW: 8%).

Since some documents were accessed on only a few occasions, it was not possible to calculate meaningful sensitivity and specificity measures for all documents. Thus, LR was only calculated for the subset of documents that had been accessed by at least five subjects (each document was accessed by 4.7 subjects on average). A total of 725 documents were accessed across the 6 scenarios, with a range from 78 to 138 documents per scenario. After culling, 88 documents were kept in the pool of influential documents (i.e. accessed by more than 5 subjects), with a range from 10 to 19 documents per scenario.

Did clinicians who accessed the same document arrive at the same answer?

Analysis of different evidence journeys taken by study participants reveals that subjects did not necessarily arrive at the same answer after having accessed the same document. In one scenario (MI), around 25% of WW subjects (i.e. 10 out of 39 subjects) cited the same source as WR subjects to support the opposite post-search answer. Across the six scenarios, subjects often produced different answers to questions despite having accessed the same document. The majority of frequently accessed documents were seen *both* by subjects who answered a question correctly after searching as well as those who answered incorrectly (Figure 1).

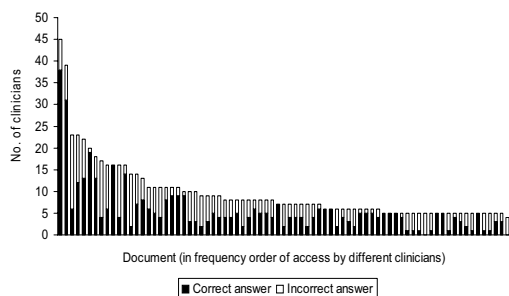


Figure 1 - Frequency of subjects answering a question correctly or incorrectly for each document accessed [14]

In addition, Figure 2 shows that documents accessed were almost equally distributed between those more likely to be associated with a correct answer (51% of accessed documents had a $LR > 1$) or incorrect answer (49% of accessed document had $LR < 1$). There was a clear variation in the likelihood that accessing different documents was associated with a subject providing a correct or incorrect post-search answer.

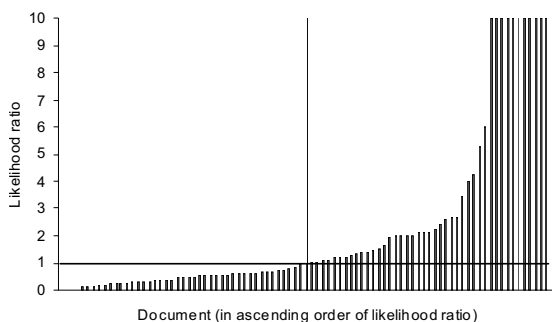


Figure 2 - Variation in influence of accessed documents in obtaining a correct post-search answer, as measured by likelihood ratio (Note: Likelihood ratio = 10 refers to documents having a likelihood ratio ≥ 10) [14]

Are there typical patterns in the evidence journey?

To better understand the way that accessing a sequence of documents might have influenced an individual in making a decision, a qualitative analysis was conducted to look for typical patterns amongst the dataset. In a search session, a *positive document* is a document with $LR > 1$ and is represented by a closed circle; a *negative document* is a document with $LR \leq 1$, represented by an open circle; an *indeterminate document* is a document with a LR that cannot be established and it is represented with a strip-patterned circle; and each query submitted to the search engine is represented with a line.

The following examples demonstrate the evidence journeys of subjects in four different categories: RR, RW, WR and WW.



Figure 3 - Example of a subject with a right answer pre- and post-search (RR)

The subject in Figure 3 was correct before and after searching. This subject expressed being very confident in the pre-search answer. The subject made only one search and accessed one document, which is a positive document (a). One possible interpretation of this evidence journey is that the first document confirmed the subject's pre-search answer; hence, the subject stopped searching and provided the correct answer.



Figure 4 - Example of a subject with a right answer pre-search and a wrong answer post-search (RW)

The subject in Figure 4 gave a correct answer before searching but changed to an incorrect answer after searching. The subject made four searches. On the first search, the subject accessed two documents; the first document was a positive document (a), followed by a negative document (b). The subject then performed more searches, viewed the titles and summaries of documents retrieved on the results pages but did not access any document until the last search, which was a document with an indeterminate likelihood ratio (c). One possible interpretation is that as the subject spent more time on the negative document (b: 5 min) than the other two documents (a: 2 min, c: 30 sec); the extra time spent on the negative document may have contributed to the subject giving an incorrect answer

after searching. (Note: Time spent on a document was measured as time elapsed between the commencement of accessing the document and the subject's next action).

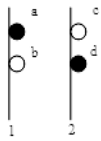


Figure 5 - Example of a subject with a wrong answer pre-search and a right answer post-search (WR)

The subject in Figure 5 gave an incorrect answer before searching, made two searches, accessed four documents and answered the question correctly after searching. The first document was a positive document (a), followed by two negative documents (b and c), and then a positive document (d). One possible interpretation is that the first and the last documents, which are positive documents, influenced the subject to change opinion and make a correct answer.

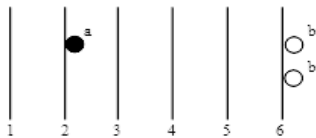


Figure 6 - Example of a subject with a wrong answer pre- and post-search (WW)

This subject in Figure 6 answered the question incorrectly before and after searching. Although the subject only accessed two documents, a positive document (a) and a negative document (b), the subject accessed the negative document twice (b). Perhaps, revisiting the negative document (b) led the subject to retain an incorrect pre-search opinion and provide an incorrect answer after searching.

Discussion

This study provides a snapshot on how clinicians searched for and accessed online biomedical literature to answer clinical questions. The likelihood ratio analysis shows that clinicians did not necessarily arrive at the same answer after having accessed the same document and that documents did not influence clinicians in the same manner. The graphical representation of evidence journeys illustrates that people take different journeys to arrive at an answer to a question and that the way documents were accessed and interpreted collectively may influence the way decisions are formed. It also reveals the possibility that the following components of an evidence journey influence the shaping of a decision outcome:

- an individual's *belief and confidence before searching*, e.g. Figure 3: RR
- the *order* of accessing documents, especially the first and last accessed documents, e.g. Figure 5: WR

- the amount of *time* spent on the documents, e.g. Figure 4: RW
- the *number of visits* the searcher made to the same document, e.g. Figure 6: WW

Each of these phenomena has been identified in the general decision making and information retrieval literature in different guises. Specifically, the literature identifies the existence of *cognitive biases* in the way people use information to make decisions [15-18]. For example, decision makers are biased by their prior beliefs and the order of information presentation when making decisions and forming impressions [19, 20].

The concept of decision biases may provide us with a theoretical model with which to understand search behaviours. Different subjects have different knowledge and levels of confidence before searching; their evidence journeys are different in search length, number and types of documents accessed; the way documents were accessed and how these documents may have interacted are possible elements that influence the way people process and use information to make decisions. Specifically, subjects may be experiencing the following cognitive biases in their evidence journeys:

- *anchoring effect*: the phenomena where one's prior belief exerts a persistent influence on the way new information is processed and subsequently affects the way beliefs are formed [21]
- *order effects*: the way in which the temporal order that information is presented or accessed affects the final judgement of an event [22]
- *exposure effect*: the phenomenon where the amount of time exposed to the information affects the final judgement of an event; e.g. [23]
- *reinforcement effect*: the phenomenon where the level of repeated exposure to information affects the final judgement of an event, which is best related to "mere exposure" discussed in [24]

Impact of these biases have been tested in a preliminary analysis that uses a Bayesian model to predict the impact of information searching on decision making [14]. The Bayesian model that predicts decision outcomes most accurately is the one that incorporates all the above-mentioned cognitive biases during information searching (without biases: 52.8% (95% CI: 47.85 to 57.59); with biases: 73.3% (95% CI: 68.71 to 77.35)).

Analysis limitations

The assumption that subjects read a document based on having accessed the document was a potential limitation in the study. Subjects may not have read documents they accessed, or only partially read them, modifying the likelihood that the document influenced them. Similarly, subjects may have been influenced by documents without accessing them, e.g., looking at the title or the abstract of the document only on the search engine results page, but not accessing the document itself.

Conclusion

This study presents a retrospective data analysis on how clinicians searched for and accessed biomedical literature to answer clinical questions. The analysis suggests that people take different journeys to answer questions, and that the way documents were accessed could contribute to different interaction effects between pieces of information, which influences the way evidence was evaluated and subsequently the decision making process. It also raises the hypotheses that people may experience cognitive biases during information searching that influence their decision making, and calls for further investigation to test these hypotheses.

Acknowledgements

The authors would like to thank Professor Johanna Westbrook, Dr Sophie Gosling and Miss Nerida Creswick for providing the data in this retrospective data analysis. Annie Lau was financially supported by the Australian Research Council during this research.

References

- [1] Eysenbach G, Jadad AR. Evidence-based Patient Choice and Consumer health informatics in the Internet age. *J Med Internet Res*. 2001;3(2):e19.
- [2] Hersh WR. Evidence-based medicine and the Internet. *ACP J Club*. 1996;5(4):A14-A6.
- [3] Morrison JB, Pirolli P, Card SK. A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. Conference on Human Factors in Computing System; 2001; New York: ACM Press; 2001.
- [4] Hersh WR, Hickman DH. How Well Do Physicians Use Electronic Information Retrieval Systems? A Framework for Investigation and Systematic Review. *JAMA*. 1999;280(15):1347-52.
- [5] Hersh WR. Ubiquitous but unfinished: on-line information retrieval systems. *Med Decis Making*. 2005;25(2):147-8.
- [6] Spink A, Cole C. *New Directions in Cognitive Information Retrieval*. 1st ed. London: Springer 2005.
- [7] Ford N, Miller D, Booth A, O'Rourke A, Ralph J, Turnock E. Information retrieval for evidence-based decision making. *JD*. 1999;55(4):385-401.
- [8] Bates MJ. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Rev*. 1989;13(5):407-24.
- [9] Wang P, Soergel D. A cognitive model of document use during a research project. Study I. document selection. *J Am Soc Inf Sci*. 1998;49(2):115-33.
- [10] Wang P, White MD. A cognitive model of document use during a research project. Study II. Decisions at the reading and citing stages. *J Am Soc Inf Sci*. 1999;50(2):98-114.
- [11] Gruppen LD, Wolf FM, Billi JE. Information gathering and integration as sources of error in diagnostic decision making. *Med Decis Making*. 1991;11(4):233-9.
- [12] Westbrook JI, Coiera EW, Gosling AS. Do Online Information Retrieval Systems Help Experienced Clinicians Answer Clinical Questions? *J Am Med Inform Assoc*. 2005;12(3):315-21.
- [13] Jansen BJ, Spink A. An Analysis of Web Documents Retrieved and Viewed. Fourth International Conference on Internet Computing; 2003; Las Vegas, Nevada; 2003. p. 65-9.
- [14] Lau AYS, Coiera EW. A Bayesian model that predicts the impact of Web searching on decision making *J Am Soc Inf Sci Tec*. 2006;57(7):873-80.
- [15] Kahneman D, Slovic P, Tversky A. *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press 1982.
- [16] Elstein AS. Heuristics and Biases: Selected errors in Clinical Reasoning. *Acad Med*. 1999;74(7):791-4.
- [17] Croskerry P. Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias. *Acad Emerg Med*. 2002;9(11):1184-204.
- [18] Kaptchuk TJ. Effect of interpretive bias on research evidence. *BMJ*. 2003;326(7404):1453-5.
- [19] Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science*. 1974;185(4157):1124-31.
- [20] Asch SE. Forming impressions of personality. *J Abnorm Soc Psychol*. 1946;41:258-90.
- [21] Wickens CD, Hollands JG. *Decision Making. Engineering Psychology and Human Performance*. 3rd ed. Upper Saddle River, NJ: Prentice Hall 2000:293-336.
- [22] Wang H, Zhang J, Johnson TR. Human belief revision and the order effect. Twenty-second Annual Conference of the Cognitive Science Society; 2000; Hillsdale, NJ: Lawrence Erlbaum; 2000.
- [23] Varey CA, Kahneman D. Experiences extended across time: Evaluation of moments and episodes. *J Behav Decis Making*. 1992;5:169-85.
- [24] Zajonc RB. Attitudinal effects of mere exposure. *J Pers Soc Psychol*. 1968;9(2):1-27.

Author for correspondence

Annie Lau, a.lau@unsw.edu.au
Centre for Health Informatics,
University of New South Wales,
UNSW Sydney NSW 2052, Australia