Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

doi:10.3233/978-1-61499-678-1-95

Usability Evaluation of Clinical Guidelines on the Web Using Eye-Tracker

Soudabeh Khodambashi[1], Heidi Gilstad and Øystein Nytrø
Norwegian University of Science and Technology

Abstract. Publishing clinical guidelines (GLs) on the web increases their accessibility. However, evaluating their usability and understanding how users interact with the websites has been neglected. In this study we used Tobii eye-tracker to analyse users' interaction with five commercial and public GL sites popular in Norway (four in Norwegian and one English of US origin (UpToDate)). We measured number of clicks and usage rate for search functions, task completion time, users' objective and perception of task success rate. We also measured learning effect for inexperienced users. We found a direct correlation between participant's satisfaction regarding website usability and the time spent, number of mouse clicks and use of search function to obtain the desired results. Our study showed that users' perceived success rate was not reliable and GL publishers should evaluate their website regarding presentation format, layout, navigation bar and search function.

Keywords. Clinical guidelines, usability evaluation, Tobii eye-tracker, UpToDate

1. Introduction

Publishing clinical guidelines on the web increases their accessibility, but not necessarily their usability [1]. Research has shown that primary care physicians spent on average two minutes seeking information for their clinical questions and time was a major concern [2]. Therefore, it is crucial that websites give quick access to the right information with little navigation effort. In our previous study we highlighted the necessity of usability evaluation of published clinical guidelines on the web [3]. Usability by definition is 'the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use' [4]. Addressing pros and cons of evaluation methods, Jaspers [5] compared usability methods for testing interactive health technologies, and provided an overview of three categories of usability methods: heuristic evaluation, cognitive walkthrough, and think aloud. In addition, Asan and Yang [6] concluded that although eye-tracking technology is useful in discerning usability problems in health information technology, it has not been widely applied.

In this study our goal was to investigate users' interaction with different clinical guidelines published on the web, by measuring success rate in task completion, task completion time and users' feedback on website usability. In addition, we investigated the possible impact of task completion time and task success rate on a participant's satisfaction levels with regard to website usability. In order to record the above

[1] Corresponding Author: soudabeh@idi.ntnu.no

mentioned metrics accurately, we used an eye-tracking system for data collection. We compared the usability of UpToDate [7] and four widely used clinical guideline websites in Norway (Legehandboka.no – a clinical handbook for general practice, but also used in specialist healthcare [8], Helsebiblioteket.no – the national health library collection of guidelines, external resources and a federated search engine [8], Helsedirektoratet.no – the collection of clinical recommendations maintained by the directorate of health [9], and Oncolex.org – mainly for cancer diagnostics, treatment, and supportive care [10]). Although usability evaluation of online resources for health professionals have been published [7, 11], to our knowledge there are no scholarly publications on usability evaluation of web-published clinical guidelines using eye-tracking technology.

2. Methods

2.1. Sample selection and implementation of case study with the eye-tracker

The inclusion criterion for informants was that they should have experience in using clinical guidelines. We approached 5th and 6th grade medical students by email and asked for their willingness to participate in our case study. A total of 14 medical students participated. They were awarded a gift certificate valued at approximately 50 Euros.

To control for previous experience with the guideline websites or the clinical cases, all subjects had to complete a paper-based questionnaire before the usability test. In order to reduce the effect of different clinical competence among the participants, and reduce the risk of learnability of the scenario topics, we designed five different scenarios for each website. Each scenarios consisted of two tasks; task one (T1) was a broad question about one specific domain (i.e. presented a patient's signs and symptoms about one disease and asked the participants to explain further steps in diagnosis and treatment), and task two (T2) was a question about a patient statement that required more detailed searches in the guidelines (i.e. identify staging of rectum cancer based on provided information in the scenario). Scenarios were presented to participants on a single piece of paper. The participants read the scenario, accessed the assigned website through eye-tracker and searched through the guidelines to find answers to the assigned tasks. The participants were asked to write their answer on the paper after reading the guidelines.

We used the Tobii EyeTracker system [12] to record the participants' interaction with the websites including mouse clicks, task completion time, users' navigation and gaze plots. All of the tests were conducted in a usability lab. The computer was equipped with a video camera and microphone. The eye-tracking system enabled us to coach participants during the test by adding instructions to each Web element–containing unique scenarios and website URLs (five Web elements for five URLs in total for the five selected websites in our study). We had a brief introductory session for participants about the usability lab and eye calibration. We did not ask participants to think aloud as it might influence the recording time [13]. We observed participants during task completion. Our scenario-based task completion was followed by retrospective semi-structured interviews to enable participants to elaborate on the issues they faced and provide feedback.

2.2. Metrics and data analysis method

We evaluated the task success rate by reviewing participants' written responses. Based on records in eye-tracker, we analysed task completion time (elapsed time in seconds), number of mouse clicks, and number of tries using search-function to complete each task. All the metrics were measured from start to finish for each use of a website. We calculated a geometric mean to present overall performance of the participants in task completion time, since this is a better estimate for small samples (n <25) [14].

3. Results

Our pre-test survey showed that none of the participants had particular previous experience that would bias our usability measures. Thus, we did not take this into account. However, not all the participants had prior experience in using the websites. Hence, we first analysed the collected data without considering users' experiences using the website which are presented in this section. Detailed analysis of data including prior users' experience using the website is further presented in the Discussion. The case study results included: 1) the mean objective success versus users' perception of success rate in finding an answer for each task (T1 and T2) for each evaluated website (Figure 1), 2) geometric mean of task completion time (Figure 2a), and 3) the mean number of mouse clicks and mean number of using search-functionality classified by tasks (T1 and T2) for each website (Figure 2b).

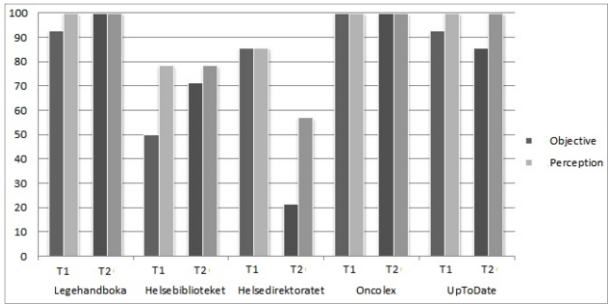

Figure 1. Objective vs. users' perception of success rate in task completion

Analysis of recorded users' navigation and gaze plots in Tobii EyeTracker showed that users looked for the navigation bar as a first choice when searching. We found that the structured navigation bar in Oncolex not only enabled users to find answers quickly with less effort, but also had a positive effect on the objective success rate in finding an answer. In contrast, none of the users could find the navigation bar on the Helsedirektoratet website and users clicked on the 'back' button excessively as they could not identify their current position. The Helsedirektoratet website has a small menu-button with classified guideline subjects, but only one user noticed it and browsed the website through the menu.

Our interviews with participants showed that they found Oncolex easier to use while Helsedirektoratet was the most challenging website for the users. They stated that they would use UpToDate for self-education, rather than clinical use, because of the compact text and layout which was time-consuming to read. Participants' feedback on Helsebiblioteket.no (it provides federated search to external sources) showed that

redirection to other webpages is challenging and they would rather use Google search engine instead. Participants stated that presenting search results as 'at-a-glance' interface (in Legehandboka) was easier to read than a list of webpages as is in Helsebiblioteket. Furthermore, participants preferred navigation bars to search functions.

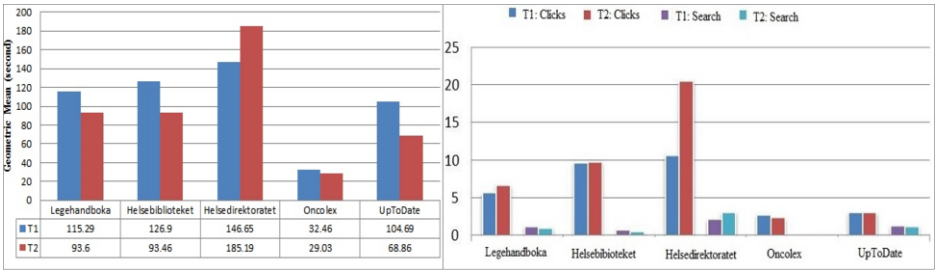

Figure 2. (a) Left: Geometric mean of task completion time, (b) Right: Mean number of mouse clicks and using search functionality.

4. Discussion and conclusion

For Oncolex, results in Figure 1 showed that users' perception and objective success in finding answers to clinical questions was similar while Figure 2 showed that task completion time was shorter compared to other webpages; the number of mouse clicks and use of the search function was also lower compared to other webpages. The Helsedirektoratet site had the opposite results: longer task completion time, higher number of clicks and use of the search function. As mentioned in the results, interviews with participants revealed that Oncolex was very easy to use. Based on Figure 1 and Figure 2, we found that the time spent (strongly correlated with users' mental effort [15]) and number of required steps (clicks and search function) to obtain the desirable results were directly correlated with users' satisfaction. Tsopra et al. found that 'the decrease in the number of clicks required seemed to be important to the physicians' [16] which was similar to our findings.

We also compared the geometric mean of T1 completion time between users with prior experience versus users without experience in using the websites (Figure 3(a)), and found that Oncolex users with and without prior experience spent almost the same amount of time finishing the task while just the opposite for Helsedirektoratet. Between users with prior experience, Helsedirektoratet was the most time consuming website. In addition, we also measured learnability of the websites by comparing the geometric mean of task completion time (T1 and T2) for users without prior experience using the websites (Figure 3(b)). We found that Oncolex users did not increase speed with practice, while Helsedirektoratet users spent longer time overall, and no increase in speed with experience.

Our study showed that usability evaluation of web-published clinical guidelines is necessary in order to make access, navigation and search more efficient for clinicians. In addition, evaluation of the success rate in finding answers is important as users' perception of success is not reliable and we need to make sure guidelines with the right information can be accessed and used by clinicians. Different factors such as structured navigation bar, menu at the top of every page, guideline content with proper format and

layout which highlights headings, sub-headings and separated from the text on one page should be considered in presentation format. This study did not focus on content assessment, such as information exchange or communication potential, but this would be interesting to examine in a future study.

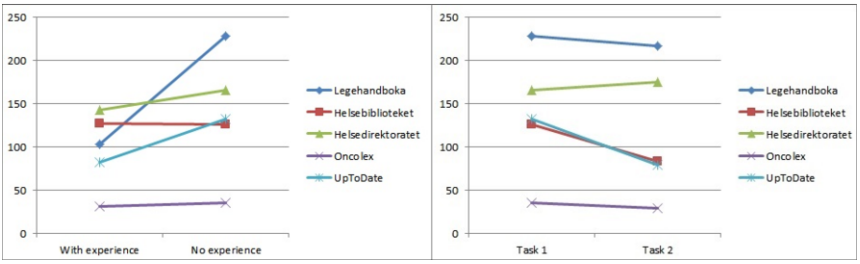

Figure 3. (a) The geometric mean of completion time for T1, (b) The geometric mean of task completion time (users without prior experience)

References

[1] R. D. Zielstorff, "Online practice guidelines," *Journal of the American Medical Informatics Association,* vol. **5**, pp. 227-236, 1998.

[2] H. C. Coumou and F. J. Meijman, "How do primary care physicians seek answers to clinical questions? A literature review," *Journal of the Medical Library Association*, vol. **94**, p. 55, 2006.

[3] S. Khodambashi, Z. Wang, and Ø. Nytrø, "Reality Versus user's Perception in Finding Answer to Clinical Questions in Published National Guidelines on the Web: An Empirical Study," *Procedia Computer Science,* vol. **63**, pp. 268-275, 2015.

[4] W. ISO, "9241-11. Ergonomic requirements for office work with visual display terminals (VDTs)," The international organization for standardization, 1998.

[5] M. W. Jaspers, "A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence," *International journal of medical informatics*, vol. 78, pp. 340-353, 2009.

[6] O. Asan and Y. Yang, "Using Eye Trackers for Usability Evaluation of Health Information Technology: A Systematic Literature Review," *JMIR Human Factors*, vol. **2**, p. e5, 2015.

[7] G. N. Fox and N. S. Moawad, "UpToDate: a comprehensive clinical database," *Journal of family practice,* vol. **52**, pp. 706-710, 2003.

[8] Legehandboka [Norwegian Medical Handbook 2015 (NEL Online)] Available: http://legehandboka.no

[9] Helsedirektoratet. Norwegian Directorate of Health [Helsedirektoratet.no]. Available: https://helsedirektoratet.no/om-oss

[10] Oncolex.org. Oncolex.org-a web based cancer encyclopedia for health care providers. Available: Oncolex.org

[11] S. E. Rosenbaum, C. Glenton, and J. Cracknell, "User experiences of evidence-based online resources for health professionals: User testing of The Cochrane Library," *BMC Medical Informatics and Decision Making*, vol. **8**, p. 1, 2008.

[12] T. Pro. (24.04.2015). Tobii EyeTracker. Available: http://www.tobii.com/en/eye-tracking-research/global/

[13] M. Van Den Haak, M. De Jong, and P. Jan Schellens, "Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue," *Behaviour & Information Technology*, vol. **22**, pp. 339-351, 2003.

[14] J. Sauro and J. R. Lewis, "Average task times in usability tests: what to report?," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 2347-2350.

[15] A. Skarlatidou and M. Haklay, "Public web mapping: preliminary usability evaluation," 2006.

[16] R. Tsopra, J.-P. Jais, A. Venot, and C. Duclos, "Comparison of two kinds of interface, based on guided navigation or usability principles, for improving the adoption of computerized decision support systems: application to the prescription of antibiotics," *Journal of the American Medical Informatics Association*, vol. **21**, pp. e107-e116, 2014.