Development of Methods for Usability Evaluations of EHR Systems

Lars H. Lilholt, Signe S. Pedersen, Inge Madsen, Per H. Nielsen, Niels Boye, Stig K. Andersen and Christian Nørh

Abstract. Developing electronic health record (EHR) systems in Denmark is an ongoing, iterative process, where also a maturation process for clinical use should be considered. Convincing methodology for collecting and incorporating in the soft- and hardware knowledge and robustness for the clinical environments is not on hand. A way to involve the clinicians in the development process is conducting usability evaluations. The complexity of the clinical use of the systems is difficult to transmit to a usability laboratory, and due to ethical issues a traditional field study can be impossible to carry out. The aim of this study has been to investigate how it is possible to identify usability problems in an EHR system by combining methods from laboratory tests and field studies. The methods selected for the test design are: the think aloud method, video and screen recording, debriefing, a scenario based on an authentic patient record, and testing on the normal production system. The reliability and validity of the results is increased due to the application of method- and data-triangulation. The results of the usability evaluation include problems in the categories: system response time, GUI-design, functionality, procedures, and error messages. The problems were classified as cosmetic, severe, or critical according to a rating scale. The experience with each method is discussed. It is concluded that combining methods from laboratory test and field study makes it possible to identify usability problems. There are indications that some of the usability problems only occurred due to the establishment of an authentic scenario.

Keywords: method, usability, evaluation, participation, EHR, EMR, CPOE, data triangulation, method triangulation, health informatics.

1. Introduction

The national IT-strategy for the Danish health care sector 2003-2007 [1] sets a target for the implementation of electronic health records (EHR) systems on every hospital in Denmark based on common standards before the end of 2005. This goal has not yet been achieved and in recent years there has been an increasing criticism of the development and the implementation of EHR systems in Denmark [2], [3].

This criticism from the clinicians has also been reflected in the international literature and is pointing towards applications being unsuitable for clinical use, systems
not supporting the clinical work practice, and lack of involvement of clinicians in the system development process [4], [5], [6]. The CPOE module is the most widely implemented part of the coming full-scale EHR systems (in Denmark), but the CPOE systems provide so far medication order entries only, thus there is still some distance to go between the actual state of EHR’s and the vision of full implementation. It is therefore of importance to evaluate systems thoroughly with structured methods to provide qualified feed-back to the remaining development process.

Usability evaluation is a way of involving the end user in the system development process [7]. As the end user has to interact with the system during the test, possible usability problems can be identified, which can contribute to improvement of the system. Applying this method to the development of EHR systems can be an approach of involving clinicians in the system development.

The two archetypes of usability evaluation is laboratory test and field study. By applying the laboratory test, vital knowledge on the system-context-interaction can be missed. On the contrary it is very problematic to apply a full blown field study at a hospital due to ethical issues. Therefore the aim of this study has been to investigate how it is possible to identify usability problems in an EHR system by combining methods from laboratory tests and field studies.

The aim of the present study was to develop a structured, authentic evaluation method of clinical end user load and gains in interacting with the (medication-) CPOE-system. The method was developed on one HMO-CPOE (Aarhus County, Denmark) but will be applied in a comparative study of three to five CPOE-systems.

2. Methods

Methods from laboratory test and field study were combined utilizing experiences from Evaluation Laboratory (EVALAB) in Lille, France [8], and the EHR-center in Trondheim, Norway [9]. The focus has been on combining advantages from both types of tests to retain a large degree of control and at the same time creating an authentic situation for the clinician by including a rich clinical context.

The components in our methods were selected with the purpose that obtained experimental data could be triangulated to validate a result. Details of the specific methods described below can be found in [7], [10].

The think aloud method is a simple method of collecting the users’ thoughts throughout a test. The user simply explains his thoughts and reflections while using the system. The think aloud method was considered as vital for the later data analysis. Video recording of the contextual work situation as well as screen capture of the CPOE system was used as a method for collecting data from the test. These recordings are essential documentation of the occurrence of usability problems during the test. Directly after the test a debriefing was performed, to increase the level of information concerning the test. Since the main target of this study was to evaluate the combination of methods and not to conduct a full usability evaluation itself, two CPOE experienced physicians participated as users. A scenario was scripted based on an authentic, complex patient trajectory to ensure the use of a wide range of functionalities of the CPOE system. The case concerned a 39 year old female patient suffering from cancer, admitted to a long-term medical treatment involving several stakeholders and multiple drug ordinations. In an evaluation laboratory, Skej-Lab, at Aarhus University Hospital, Skejby Sygehus, a ward was set up with a bed, a table, a chair, and a medicine chest.
and various other accessories. Besides the ward an office was set up with a computer from where the user should access the CPOE system, see Figure 1.

Figure 1- Illustration of the test facility setup. To the right the ward, in the middle the office, and to the left the control and debriefing area can be seen.

An actress was playing the role of the patient, partly due to ethical issues, but also to secure a similar act in each test. Since the test facility was physically located at the hospital, it was possible to use the normal production IT-system instead of a test version prototype. This should contribute to the realism of the test exposing authentic system response time etc.

After the tests the video and screen recordings were analyzed and log files were prepared. These log files were then analyzed with the purpose of identifying usability problems.

3. Results

Classification system

The data analysis of the log files identified a number of usability problems. These were classified according to Molich’s rating scale for classifying severity of a usability problem into the categories: cosmetic, severe, and critical [11].

Inspired by Skov and Stage [12] a diagram to assist the quantification of usability problems according to Molich’s rating scale was developed in an iterative process. The diagram shown in Figure 2 is not generic, but specifically targeted at usability problems in a CPOE system.

<table>
<thead>
<tr>
<th>Response Time</th>
<th>GUI Design</th>
<th>Functionality</th>
<th>Procedure</th>
<th>Error Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosmetic</td>
<td>The GUI design is too simple or does not support the user. The user has to make many clicks and therefore is aware of which task it is not fitting.</td>
<td>The functionality causes temporary problems but it is possible to perform the function, taking a few seconds extra.</td>
<td>The system gives rise to an incorrect procedure.</td>
<td>The user message is not enough, and causes a few seconds delay.</td>
</tr>
<tr>
<td>Severe</td>
<td>The functionality is difficult to carry out, taking several seconds extra. The user makes a failure.</td>
<td>The system gives rise to an incorrect procedure.</td>
<td>The user message is not explicit, and causes a few seconds delay.</td>
<td></td>
</tr>
<tr>
<td>Critical</td>
<td>The functionality is not possible to perform or a system crash occurs.</td>
<td>The system gives rise to an incorrect and critical procedure.</td>
<td>The user message is not explicit, and causes a few seconds delay.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 - The classification diagram developed in the study.
Experiences with the methods

Think aloud: Both users expressed that the think aloud method was straightforward and simple to use. During the medication process physicians are normally supported by nurses in discussing the patient situation and the specific medication, so they felt no interference in having a test monitoring person beside them during the test.

Recording: The recordings of the screen display of the CPOE system were essential for the analysis, but the video recordings of the ward and the office were less useful. The observer’s notes taken during the tests were quite adequate to capture the essential information from these locations.

Debriefing: The debriefing session immediately after the test gave useful results, as the users could be more specific about and comment on some of the experienced problems during the test. The users also gave their experience of having participated in the test, both expressing that a nurse would promote the realism in the test set-up.

Scenario: The scenario play based on the authentic patient record was the key element in the test to promote the realism. The scenario play on the “ward” helped the users to acquaint themselves with the patient case and it helped them to act realistic. Both users expressed that they made considerations about the patient’s record and gave serious considerations to the ordinations described in the scenario description. The name of the patient and the date of birth in the CPOE-system did not correspond to the data in the scenario. The patient name in the system was “CPOE test Woman”, being twice as old as the female patient in the scenario. These system data did not promote the realism.

Scenario settings: The arrangement of the ward and the clinical staff office worked well. A pill bottle at the patient’s bedside table should illustrate the patient’s home medication. The label on the pill bottle did not match the drug prescription in the scenario which confused one of the users. One of the users did bring his own white coat, including reference books, stethoscope, etc., and he did use one of his reference books during the test to check some medicine.

Production system: Performing the test at the hospital using the production system definitely had an impact on the results of the tests. Halfway through one of the tests a system crash occurred, and the test could not continue. Both users experienced several times long response times.

User profile and number of users: Both users were experienced users giving the advantage that they were both able to enhance positive aspects and application-specific usability problems at the same time. Due to the system crash the second user did not complete the whole scenario.

4. Discussion

Data collection

If a system is to be used while a dialogue is going on between e.g. a clinician and a patient, an alternative to the think aloud method has to be considered. An alternative could be retrospective testing [10] where the test monitor and the user immediately after the test go through the recordings of the test together. On the basis of the recordings and the questions from the test monitor, the user should be able to recall some of his considerations when using the system. There is a risk that some information will be lost, depending on the user’s ability to recall his considerations.
The video recordings of the ward and the office were of little use, and it should be considered if it is adequate to use the test observer’s notes. Though, it is recommended to keep video recordings as a back-up, and as a documentation of the test issues. The triangulation of the think aloud method, the recordings, and the debriefing, has increased the reliability of the results.

The use of experienced users limited the number of false usability problems, as they have experienced the problems before and they have better knowledge of the possible functionalities in the CPOE system. This enhances the validity of the results. In the light of the system crash, resulting in only one-and-a-half scenario was tested, three or maybe more users should be considered. Increasing the number of users increase the amount of empirical data to be analyzed, so this should be a balance between resources available and the aspiration for robustness and validity of the results.

**Realism parameters**

The loads and gains that the CPOE system demands/give from/to the clinical end-user are evaluated through the test persons acts - as recorded by video - and through the thoughts - as reflected in the verbalisation during the think aloud part of the test. Loads and gains to other stakeholders in the patient trajectory and to the health care providing organisation are not evaluated by this method. The use of a script based on an authentic patient record and the use of a physical patient- artefact - utilizing an actress as patient - were essential for the clinical richness of the test-situation and hence the validity of the system evaluation, since it promoted tasks in realistic spatial and temporal relations to be acted upon by the end-user test-persons.

It is of importance that the clinical information is realistic and realistic complicated. It should not be too specialized or advanced, since it is not the medical issues in the story that are subject of test and discussions. Both test-end-users expressed that the case was realistic and that the identified usability problems during the test are so general that they could have been identified with other cases. This increases the reliability of the results.

Accessories as e.g. pill bottles have to be correct, in order to support the realism and the user’s ability to play the scenario. It is recommended to ask users to bring their own white coat and other items they normally wear at their clinical work. One of the users in the test requested more office equipment which could easily be arranged with a few tables, paper and pens, and telephones. Including a nurse in the test would also promote the realism of the test, as a physician normally discusses ordinations, prescriptions, etc. with a nurse. This would at the same time open up the possibility to evaluate more applications and the support of the application to cooperative clinical work.

A prototype of the CPOE system would have given less credible results. Identified problems such as system crash and response times would not have been recognized using a test system. The use of the production system and the arrangement of the test at the hospital premises are considered to be very important parameters to promote the realism in the test.

**Further development of a classification system**

The diagram in Figure 2 was developed to classify the identified usability problems in this project. A more exhaustive and exclusive classification system should be developed through additional iterations. Such a classification system could be used as a support in identifying usability problems in a CPOE system, as documentation for
having classified a usability problem, support a better communication and procurement to the developers of the system, and finally it could serve as a benchmarking tool for CPOE systems.

5. Conclusion

This small study indicates a path towards better methods for capturing and communicating information about the overall clinical performance and influence on the clinical process of utilizing information technology - as well as a method of quantitative estimation of the “clinical impact” of specific clinical systems. To capture both end-user acts and thoughts to the activity and performance of an IT-system, a combination of methods from laboratory test and field study should be employed in an artificial clinical milieu with sufficient complexity and richness to mimic the clinical environment. Some of the usability problems only occur due to the establishment of an authentic scenario and performing the test on the production system, instead of a test version prototype. The classification system for usability problems has proven to be supportive to the identification and communication of usability problems. The methods applied should still be fine tuned and more tests should be conducted to consolidate an optimal triangulation and to prove the usefulness of the method in comparative evaluation of clinical systems.

6. Acknowledgements

We thank Jens Kelsen and Søren Tang Knudsen from Randers Hospital for participating as clinical users in the usability tests.

References