Usability Testing in Medical Informatics: Cognitive Approaches to Evaluation of Information Systems and User Interfaces

Andre W. Kushniruk, M.Sc.[1], Vimla L. Patel, Ph.D.[1], and James J. Cimino, M.D.[2]
[1]Cognitive Studies in Medicine: Centre for Medical Education,
McGill University, Montreal, Quebec, Canada
[2]Department of Medical Informatics, Columbia University, New York, NY

This paper describes an approach to the evaluation of health care information technologies based on usability engineering and a methodological framework from the study of medical cognition. The approach involves collection of a rich set of data including video recording of health care workers as they interact with systems, such as computerized patient records and decision support tools. The methodology can be applied in the laboratory setting, typically involving subjects "thinking aloud" as they interact with a system. A similar approach to data collection and analysis can also be extended to study of computer systems in the "live" environment of hospital clinics. Our approach is also influenced from work in the area of cognitive task analysis, which aims to characterize the decision making and reasoning of subjects of varied levels of expertise as they interact with information technology in carrying out representative tasks. The stages involved in conducting cognitively-based usability analyses are detailed and the application of such analysis in the iterative process of system and interface development is discussed.

INTRODUCTION

There is currently a need for development of medical technologies based on a sound groundwork in the understanding of the cognitive processes involved in health care decision making and reasoning. In addition, there is a need for the development and application of improved methodologies for the assessment of medical systems and their user interfaces, both for providing input into the iterative design process and for end product testing. Problems with conventional methods of evaluation, such as questionnaires and interviews with users, include the limitations of user's recall of their experience in using a computer system. Such techniques may inform us of what users think they do in using a computer system, however this may be considerably different from their actual behaviour [1]. Outcome-based evaluations, which are also widely used in medical informatics, focus on examining effects of system use on clearly defined and pre-specified outcome measures, but do not allow us to carefully examine the actual process of system use by health care workers as they perform complex day-to-day activities. Recent work from both the study of human-computer interaction [2,3] and from the study of cognitive science in medicine [4], show considerable promise when integrated and applied to the problem of evaluating health care information systems and their user interfaces. In this paper we will describe some of the methods we have adapted and modified from these areas in order to develop more effective ways of providing system designers with information about the effects of their systems and how they can be improved. These methods can be applied in the study of systems in both the laboratory setting and in real-life settings and contexts.

BACKGROUND

Usability Testing

Usability of a computer system can be defined as the capacity of the system to allow users to carry out their tasks safely, effectively, efficiently and enjoyably [2]. In the field of medical informatics, issues of usability have come to the fore, with the ultimate acceptance or rejection of systems such as computerized patient records depending to a large extent on their degree of usability. To cope with the challenge of designing systems that provide desired functionality, and that are easy to learn and use, a variety of techniques from the study of human-computer interaction have emerged and become important in the general software industry [5]. Usability testing refers to the evaluation of information systems that involves participants (i.e. subjects) who are representative of the target user population. Over the last several years, the field of usability engineering has emerged and commercial usability laboratories have sprung up, where various forms of usability tests are conducted, typically involving recording of subjects as they interact with systems under study [6].

Cognitive Task Analysis

Our approach to the evaluation of medical information systems represents a novel integration of work from the field of usability engineering, described above, and cognitive research in medicine, particularly, cognitive task analysis. Cognitive task analysis is concerned with characterizing the decision making and reasoning skills of subjects, as they perform activities involving the processing of complex information [2]. In our recent applications of task analysis we have studied subjects of varied levels of medical expertise (e.g. medical students, residents, attending physicians) as they interact with complex information technology such as computerized patient record (CPR) systems. Subjects may be asked to "think aloud", or verbalize their thoughts, as they perform specific tasks (e.g. entering data into a CPR system [1]). We also apply principled methods for psychological analysis of such verbal data which have emerged from the study of medical cognition [4].

Research from the study of human expertise and medical cognition has provided a theoretical/methodological framework for the development of techniques that can be applied in conducting technology evaluation. For example, written descriptions of medical cases, can be used as stimulus material for subjects (e.g. physicians) as they interact with systems (e.g. subjects being requested to enter or summarize the essential findings of the case into a computer system). This type of approach allows for experimental control in the development and presentation of information to subjects, and draws on considerable experience in the collection and analysis of such data from the study of reasoning of health care workers [4]. More recently, our evaluations of technology have also drawn from the cognitive study of doctor-patient interaction (involving observation of physicians as they interview patients), extended to include study of the interaction of physicians with patients, while using a computer system (e.g. CPR or decision support) in realistic situations.

Computer-Supported Video Analysis

Video recording subjects as they interact with user interfaces (in carrying out specific tasks), provides a source of data that is rich in physical, temporal and social context. Video recordings can also be made of the actual computer screens, as subjects either think aloud or conduct an interview with patients [1]. Recently, computer tools for the analysis of video data have made this form of data collection practical and greatly facilitate the reliable coding and analysis of video recordings of subjects' actions, verbalizations and problems. As described below, using computer-supported coding, textual annotation can be directly linked to the corresponding video sequences on a VCR interfaced to the computer, allowing for computer control of the video tape and automatic access and play-back of time-stamped video sequences [7].

METHODOLOGICAL APPROACH

In this section, the steps we employ in carrying out cognitive evaluations of health care systems and user interfaces are detailed. Although there may be some variation in the stages, our work typically involves consideration of each of the steps.

Step 1. Development of Test Plan

One of the most critical steps in conducting cognitively-based usability testing is the development of a sound test plan. This includes identification of the overall objectives of the evaluation, for example, describing problems in human-computer interaction, or evaluating the effects of a system on physician decision making.

Step 2. Selection of Representative Users/ Study Design

This stage involves identification and selection of target subjects for the evaluation. Subjects should be representative of end users of the system under study, and several levels of subjects may be studied (e.g. to investigate how effective a computer system or user interface is for different types of users). In our studies, groups of subjects are typically selected based on level of medical expertise (e.g. residents and attending physicians) or level of computer literacy (as determined from a pre-test questionnaire). As evaluation involving video analysis provides a rich source of data, a considerable amount of information may be obtained from a small number of subjects (e.g. three or four in a group), however if a goal of the evaluation is to produce statistical analyses of, for example user errors, a minimum of eight subjects per group is recommended.

Study designs may consist of within-group or between-group designs. Between-group testing might involve, for example, comparison of the usability of two different prototype versions of a user interface, involving testing with two groups of subjects. As another example, testing may involve use of a CPR system by two group of subjects - one group who are highly computer literate, and another group of subjects who have had little experience with computer systems. Within-group studies we have conducted have focused on longitudinal study of how health care workers learn how to use and master medical information systems over time, with testing occurring at intervals following initial training in use of a system [7].

Step 3. Selection of Representative Tasks/Contexts

The studies we have conducted can be located on a continuum from controlled laboratory studies, e.g. involving use of constructed medical cases, to naturalistic studies of doctor-patient interaction involving use of computer systems in real contexts (e.g. tasks consisting of subjects being asked to interview a patient while entering data into a computerized patient record system [7]). In laboratory-based work, we have employed written medical case descriptions, which are used as stimulus material (e.g. subjects may be asked to develop a diagnosis for the case, using a decision support system). Naturalistic studies of real-doctor patient interaction sacrifice ability to experimentally control the study for an increase in ecological validity (i.e. collection of data on use of a system in a hospital clinic). In either case, the selection of tasks should be based on the overall objectives of the study as identified in step 1. In addition, in both experimentally-controlled and naturalistic studies, tasks should be chosen to be representative of real uses of the information technology under study.

Step 4. Setting up the Testing Environment

The testing environment can vary considerably, depending on the monetary investment in the testing facility and the nature of the evaluation. Currently, commercial usability laboratories are growing in number throughout North America. For the work we describe in this paper, we have adopted a cost-effective approach to conducting cognitively-based usability testing in medical environments that does not require an expensive usability laboratory (usability laboratories typically consist of testing rooms containing computer systems subjects interact with, and observation rooms with one-way mirrors, for experimenters to watch subjects from [6]). For most of our studies we have adopted a portable, "discount usability engineering" approach [3], involving video recording of subjects in the most convenient setting possible, or in some cases even within clinics and actual hospital settings. Equipment for video recording subjects consists simply of a camcorder and a microphone for recording subjects' verbalizations. In addition to videotaping subjects as they interact with systems, all the computer screens from the subject-computer interaction are recorded for later detailed analysis. This is done either by splitting the output of the subject's computer display to a second monitor and video recording that screen, or by employing a PC-to-video converter which converts the output of the computer directly to video.

Step 5. Conducting the Usability Test

In our studies participants include the subject (e.g. health care worker), a test administrator whose responsibility is to ensure that the session proceeds properly, and in the case of studies involving doctor-patient interaction, either a real patient, or a "simulated patient" (i.e. a research collaborator, who "plays" the part of a patient). The test administrator's instructions to subjects vary, depending on the nature of the study. In some studies (e.g. evaluating the user interface of a CPR system), subjects may be asked to "think aloud", or verbalize their thoughts as they interact with the system to perform a task (e.g. enter data into the system). This methodology provides a powerful way of recording subject's thoughts during the process of using and interacting with a computer system, with comments and verbalizations being linked to the subject's corresponding actions on the computer (as described in step 6). The test administrator must also ensure that during the testing session all video-recording proceeds properly, including both recording of the computer screens, and recording of subject-computer interactions. The end of the testing session may involve presenting the subjects with a questionnaire on their subjective impressions of the system's usability (for comparison with the video-based process data) and debriefing of the subject about the study in general.

Step 6. Data Analysis

The analysis of data collected from can range from informal analysis, based on the experimenter's impressions gained from watching the subjects or viewing of the video recorded data, to more formal analyses. In our laboratory we have worked on developing novel methods for rigorous scientific analysis of such data, which are both objective and cost-effective. In order to cope with the complexity and density of the video record of subject's actions and computer screens, we have refined a number of approaches to developing principled coding schemes and practical approaches to video coding, described below.

Prior to analyzing the data from our studies, representative tapes of subject-computer interaction are reviewed by the experimenters to identify "hot spots", indicating major usability problems, or aspects of interaction that should be targeted for further detailed analysis. In addition, we also employ, as a preliminary form of analysis, usability inspection methods, which refer to a set of techniques for principled inspection of a computer system and its interface [8]. Usability inspection methods, such as the cognitive walkthrough, involves "bench-testing" of a system, with the analysts or experimenters methodically stepping through use of a system, identifying possible user problems, goals and actions. This type of evaluation, which does not involve direct testing of end users, can be conducted as a preliminary step in identifying aspects of usability that may be of interest. The results of such analysis, can be input into the selection of categories for coding video data of actual computer use.

Categories we have employed for analyzing video data include categories for identifying for subjects' comments and verbalizations regarding the following: *information content* (e.g. whether the information system provides too much information, too little etc.), *comprehension of graphics and text* (e.g. whether a computer display is understandable to the subject or not), *problems in navigation* (e.g. subject had difficulty in finding desired information or computer screen), and *overall system understandability* (e.g. understandability of icons, required computer operations and system messages). In addition to these categories, which focus on classical aspects of human-computer interaction, we have also extended our analyses to allow for the identification of higher level cognitive processes. Thus, we can examine both human-computer interaction problems and the effects of systems on higher level reasoning and decision making processes. For example, in some studies we code each occurrence of the generation of a diagnostic hypotheses by a subject, or request for information from a patient, in the case of studies of doctor-patient-computer interaction.

Coding of the video and audio recordings first involves having the audio portion of the testing session (e.g. subjects' thinking aloud) transcribed and entered into a word processing text file. The next step involves the experimenters watching the tape from beginning to end and identifying occurrences of coding categories. To facilitate this process, we employ a commercially available computer-based video annotation tool known as CVideo. This tool allows the experimenters to link annotations and transcripts (e.g. of subjects' verbalizations) in a computer text file, with the corresponding video sequences on a VCR. Using CVideo, sections of a computer-based text file containing the transcript of the verbalizations can be "time stamped" to the corresponding section of the video tape. This allows for computer-based control of the video tape, automatic searching of the tape for video sequences corresponding to text annotations and coding, and more generally facilitates improved inter-rater reliability in video coding.

Step 7. Recommendations to Designers

Having completed the collection and analysis of data, as described in the steps outlined above, the results are transformed into recommendations for system improvement and modifications to the user interface. This can involve summarizing the results of the analyses, in terms of frequency and importance of identified user problems. Based on the analysis, recommendations (e.g. regarding improvement of display of information, or improvement of system messages) can be made for each of the most important usability problems identified. This information should be communicated to system designers in the most expedient manner (e.g. production of a final report or presentation of findings to the design team).

Step 8. Iterative Input to Design

After implementation of changes to a system, based on the recommendations to the programming team in step 7, testing may be repeated to determine how the changes now affect the system's usability. In this way, usability testing can become integrated in the process of the design and development of information systems, iteratively feeding information back into their continual improvement.

EXPERIENCES TO DATE

Cognitively-based usability testing can be applied throughout the life cycle of information systems, i.e. from testing and evaluation of early prototypes (i.e. formative evaluation) to final, or summative, evaluation to determine if a computer system has met usability criteria. From our experience, the greatest benefits of the approach described above, can come from formative analysis, where there is an opportunity for results to be communicated to the designers of the information system and appropriate improvements to a system can be made on the basis of these results. In one study we recently conducted along these line [9], the frequency of all coded user problems were determined from video analysis of nine subjects (physicians) interacting with the user interface to a CPR system and its underlying medical vocabulary. The transcripts of subjects' interactions were analyzed and all coded categories of user problems were ranked, in order of their frequency of occurrence in testing sessions. The identification of particular problems from transcripts of user's thinking aloud and computer actions included need for greater consistency in the user interface (e.g. for data entry procedures and selection methods), as well as need for streamlining data entry. This summary information was then presented to the programming team. Based on recommendations for dealing with each of the identified problems, the user interface was strategically modified by the programming team at Columbia University, and a follow-up study of the system (conducted after the recommended changes were implemented) has indicated a ten-fold decrease in the average number of user problems.

A variety of other studies involving the methodologies outlined above have been conducted in our laboratory, ranging from controlled experimental study of subjects as they enter data into CPR systems, to the study of subjects as they interview a "simulated patient" while using a CPR system [7]. In one study, we tested a new CPR user over time, from a baseline evaluation of her interview style prior to using the CPR, through her training period on the system, and over several testing sessions as she became more familiar with the system [7]. Results from this, and related studies we have conducted, indicate that use of the system over time, resulted in essential changes in the subjects' diagnostic reasoning and decision making strategies. For example, as subjects became more familiar and comfortable with the use of the CPR, they began to be guided to a greater extent by the system's sequence and organization of information in conducting patient interviews, eventually following a "screen-driven" strategy, where questions posed to patients largely matched the sequence and order of medical findings displayed on the computer's screen.

We have also conducted a comparison of our approach to usability testing with questionnaire data on usability collected from subjects (physicians enrolled in continuing education) immediately after their interaction with a computer-based tutorial. Results indicated that subjects tended to rate the system in a very positive light on the questionnaire, despite the fact that video recording of their interaction with the system showed that they had encountered considerable problems in using the system, ranging from inability to navigate through the information contained in the program, to comments indicating that the program's content was out of date [10]. This is consistent with recent studies indicating the limitations of conventional methodologies, such as use of questionnaires, in assessing usability of medical technology [11].

CONCLUSION

In this paper we have described our work in the evaluation of medical technology which integrates ideas and techniques from a number of emerging fields, including usability engineering and the psychological study of medical cognition. The approach leads to rich data, that can be collected and analyzed in an efficient manner, using new approaches like computer-supported video analysis. Such data can greatly extend and complement traditional techniques for evaluating information systems in medical informatics. In addition, as described, the results of such analyses can be fed directly back into the iterative design of systems and user interfaces. Furthermore, by employing a "portable" approach to collecting video data, involving use of simple recording equipment that can be taken into the field, we have been able to extend our laboratory-based testing to testing in environments such as clinics in assessing complex interactions between physicians and computers. Our future directions include continuing with both controlled laboratory studies, involving presentation of case descriptions to health care workers as they interact with systems, and the assessment of the interaction between provider, patient and system in clinical environments, as well as in the home-care setting.

Acknowledgments

This work was supported in part by a grant from the National Library of Medicine as the EMR Cooperative Agreement (LM05857) to James Cimino and in part by the Medical Research Council of Canada (MA-13439) to Vimla Patel.

References

1. Kushniruk AW, Patel VL. Cognitive computer-based video analysis. In Greenes R, et al, eds. Proceedings of the Eighth World Conference on Medical Informatics, 1995:1566-1569.
2. Preece J, Roger Y, Sharp H, Benyon D, Holland S, Carey, T. Human-computer interaction. Addison-Wesley Publishing Company: 1994.
3. Nielsen J. Usability engineering. New York: Academic Press, 1993.
4. Patel, VL, Arocha, JF, Kaufman DR. Diagnostic reasoning and medical expertise. In Medin D, ed. The Psychology of Learning and Motivation (pp. 187-252), New York: Academic Press, 1994.
5. Rubin J. Handbook of usability testing. New York: John Wiley & Sons, 1994.
6. Wiklund ME. Usability in practice. New York: Academic Press, 1994.
7. Kushniruk AW, Kaufman DR, Patel VL, Levesque Y, Lottin P. Assessment of a computerized patient record system: a cognitive approach to evaluating an emerging medical technology, MD Computing, 1996:13(5), 406-415.
8. Nielsen J, Mack RL. Usability inspection methods. John Wiley & Sons, 1994.
9. Kushniruk AW, Patel VL, Cimino JJ, Barrows RA. Cognitive evaluation of the user interface and vocabulary of an outpatient information

system. In Cimino JJ, ed. Proceedings of 1996 AMIA Annual Fall Symposium, 1996: 22-26.

10. Kaufman D., Kushniruk, A., Bouchard R., Patel V., Yale J. Physicians' knowledge and practices in diagnosis and treating hypercholesterolemia and the effects of a CD-I program on learning and behavioural change. Technical report, Centre for Medical Education, McGill University, 1994.

11. Mackenzie CF, Jefferies NJ, Hunter WA, Bernhard WN, Xiao Y. Comparison of self-reporting of deficiencies in airway management with video analysis of actual performance. Human Factors, 1996:38(4), 623-635.