

‘Televaluation’ of clinical information systems: an integrative approach to assessing Web-based systems

Andre W. Kushniruk^{a,b,*}, Camille Patel^b, Vimla L. Patel^b, James J. Cimino^c

^a *Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3*

^b *Cognitive Studies in Medicine: Centre for Medical Education, McGill University, Montreal, Quebec, Canada*

^c *Department of Medical Informatics, Columbia University, New York, USA*

Received 7 July 2000; received in revised form 1 November 2000; accepted 4 December 2000

Abstract

The World Wide Web provides an unprecedented opportunity for widespread access to health-care applications by both patients and providers. The development of new methods for assessing the effectiveness and usability of these systems is becoming a critical issue. This paper describes the distance evaluation (i.e. ‘televaluation’) of emerging Web-based information technologies. In health informatics evaluation, there is a need for application of new ideas and methods from the fields of cognitive science and usability engineering. A framework is presented for conducting evaluations of health-care information technologies that integrates a number of methods, ranging from deployment of on-line questionnaires (and Web-based forms) to remote video-based usability testing of user interactions with clinical information systems. Examples illustrating application of these techniques are presented for the assessment of a patient clinical information system (PatCIS), as well as an evaluation of use of Web-based clinical guidelines. Issues in designing, prototyping and iteratively refining evaluation components are discussed, along with description of a ‘virtual’ usability laboratory. © 2001 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Cognitive evaluation; Patient clinical information systems; Technology assessment; Telemedicine; World Wide Web

1. Introduction

With the increasingly widespread use of the World Wide Web for health applications, the development of effective methods for evaluating their effectiveness and usability is becoming

critical. Designing evaluations in health informatics can be difficult but becomes even more challenging when it involves evaluation from a distance. Given the increasingly distributed nature of current systems, this may include assessment of information resources accessed from various locations ranging from the clinic to the home. In this paper, we describe a framework for designing evaluations directed at assessing use of Web-based

* Corresponding author. Tel.: +1-416-7362100 ext. 33980; fax: +1-416-7365757.

E-mail address: andrek@mathstat.yorku.ca (A.W. Kushniruk).

information systems by both patients and providers. Illustrations will be provided in the context of the assessment of several systems, including PatCIS, a patient clinical information system (developed at Columbia University) as well as an assessment of the use of Web-based clinical guidelines by physicians.

The effects of changes being brought about by emerging technology, including Internet-based information resources for patients, must be considered in relation to patient understanding and provider therapeutic goals. In order to develop individualized, context-sensitive information systems that will end up being useful for physicians and patients, we must be able to evaluate how that information is understood, who is trying to understand it, and what problems occur in its comprehension and application. We have been involved in cognitive studies of reasoning and comprehension of medical information by both lay people and health-care professionals. Currently, this work has been extended to evaluation of the use of Web-based information resources to assess their usability and impact on health care. We are applying approaches to evaluation that borrow from advances in a number of areas, including cognitive science and the emerging field of usability engineering [1]. The objective is to develop a coherent framework for understanding the effects of advances in health-care information technology, in particular telemedicine.

The evaluation of Internet-based technologies, especially those designed for use by a wide range of users (i.e. both health professionals and lay people), raises a number of issues that go beyond conventional evaluations typically undertaken in medical informatics. These include: (1) determining the extent to which the end user's view of the system differs from that of the designers and how potential 'mismatches' influence system

effectiveness, especially as the range of users becomes increasingly varied; (2) going beyond assessment of user satisfaction to consideration of how the user's interaction with the system changes over time and how use of such technology changes the interaction between patients and health-care professionals; (3) characterizing the impact of systems on reasoning and thought processes as a result of continued use; and (4) identifying technical and methodological issues for performing evaluations remotely.

A number of studies have investigated the use of Internet-based information systems by health-care providers and described development of evaluation instruments (typically questionnaires) for assessing their use of information technology [2,3]. However, an in-depth understanding of the effects of use of the WWW in providing providers and patients with access to their medical data necessitates new approaches to evaluation. As health care rapidly moves towards widespread distribution of medical information via the WWW, such an understanding will become increasingly important. Over the past decade, we have worked on developing an integrated set of methods for evaluating the use and impact of information systems in health care [4]. This has included development of techniques for laboratory study of end users (in particular, video-based usability testing) that we are extending to the naturalistic study of technology use in settings such as clinics and doctor-patient interviews. Our work in the evaluation of systems has been motivated by a number of considerations including the observation that the majority of evaluations in medical informatics have focused around measurement of outcomes (e.g. the effect of the use of a computer system on number of pre-specified outcome measures) with far less work being conducted on examining the effects of systems on the critical

processes involved in health-care decision-making and reasoning. In addition, conventional assessment methods (such as exclusive use of questionnaires, or feedback forms for Web-based applications) have a number of limitations, including the reliance on users' recall of their previous experiences in using systems. It has been our experience that new approaches to evaluation are required to understand more fully the impact of the WWW as a new medium for access to health-care information. To address these issues, we have conducted a number of system evaluations in health care based on principles of cognitive task analysis.

Cognitive task analysis aims to characterize the decision-making of subjects of varying levels of expertise as they perform tasks involving information processing, for example, a physician entering a diagnosis into a computerized patient record system, or a patient interpreting information provided by a health-related Web site. In our laboratory-based studies, we have video- and audio-taped users' interactions with a variety of computer systems. The approach typically involves having subjects (e.g. patients or physicians) 'think aloud' as they interact with the system under study. The studies typically involve video recording of all user interactions (including all computer screens) and coding of the resulting video- and audio recordings using methods that we have adapted from both the study of medical reasoning and the study of human-computer interaction [4]. Based on this approach, we have been able to characterize the effects of use of health information systems on physicians' decision-making and interaction with patients. For example, we have found that small changes in the structure of information presented in information systems can strongly guide and alter physicians' basic reasoning and decision making strategies, as well as their requests for

information when interviewing patients [5]. Related work we have conducted has involved a study of lay reasoning of patients as they interact with information technologies over the WWW. This has included a combination of methods, including video recordings of user-system interactions and follow-up interviews with end users.

In this paper, we describe our recent extension of this line of work to the development of assessment methods for evaluating health-care systems at a distance, i.e. 'teleevaluation' of emerging health-care information technologies. It is argued that as the technology we assess changes (invariably becoming more advanced), so must our methods for tracking use of these systems in order to assess their impact. In addition, we feel that it is critical to evaluate technology in the user's real-life setting, in the context of other competing cognitive tasks and processes, in order to assess its full impact on decision-making and reasoning. The overall approach described in this paper attempts to integrate the advantages of a number of techniques, including strategic scripting and presentation of on-line questionnaires and Web-based forms to assess use of systems as close to the point of care as possible. In addition, we have extended our video-based usability testing and recording of end users to application across the WWW, with the objective of moving usability engineering and assessment from the laboratory to the real-world.

2. Evaluation objectives

In this paper, we describe work on the development and integration of evaluation instruments specifically designed for assessing interaction with Web-based information systems in health care. We will also describe a working model for conducting 'teleevalua-

tions' from a remote site. As in conducting other types of evaluations in medical informatics [6], the first step involves determining the objectives of the evaluation. Much of our current work focuses around evaluation of effects of Web-based clinical information systems designed to provide access to medical data and advice (based on guidelines) to both patients and providers. Specific questions we seek to answer regarding our investigation of these types of Web-based clinical information systems include the following:

- How do providers' and patients' perceptions of use of technology change over time as they begin to interact and use Web-based information systems?
- How does previous experience in using computers and expectations about using information technology affect actual use of such systems over time?
- Is the content and functionality provided by the system of value to users for their decision-making and provision of improved health care?
- What problems do providers and patients have in comprehending, understanding and applying information provided on-line over the WWW?
- How does the use of such technology affect the patients' interaction with their health-care providers?
- How does the use of information systems affect what patients do regarding daily management of their condition, in terms of reasoning, decisions and actions?
- What are the current limitations of Web-based information systems in providing user-specific information and how can they be improved from the point of view of usability (i.e. their efficiency, effectiveness and enjoyability)?

To address these questions, we have been involved in refining a number of evaluation instruments for obtaining baseline user profi-

les and then specifically addressing the issues raised above. This has also involved working with system design teams in refining system content, based on feedback from our evaluations in an iterative design cycle (driven by our formative evaluations). For example, we have iteratively modified the user interface of patient record systems (during the system design phase) based on the results of usability testing involving analysis of real user–system interactions. We have recently extended this to development of on-line questionnaires and forms scripted to appear at the point of use in order to capture context-sensitive information about the use of systems, which can be used in refining and improving these systems. The complexity of designing evaluations that address the above issues requires an integrated approach to both data collection and analysis. By emphasizing a methodological approach that continually takes advantage of the latest technological advances and a theory-based approach from cognitive science, evaluations can be designed that go beyond collection of conventional usage statistics and questionnaires.

Cognitive science provides theoretical frameworks from which our methods in the study in medicine have been drawn [7]. Cognitive science is an interdisciplinary endeavor that builds on several areas of research, including cognitive psychology, computer science, linguistics, anthropology and artificial intelligence. Our initial work in applying cognitive science to the study of medicine focussed on characterizing the nature of expertise in medical reasoning and decision-making and has included an examination of text comprehension and the representation of knowledge by health-care workers of varying levels of expertise [8]. We have extended this work to the study of reasoning by patients [9] as well as extending the methodological approach taken to examination of the effects of

use of information technologies on health-care decision making and reasoning. Methods that we have borrowed from cognitive psychology and incorporate in much of our work include a collection of ‘think aloud’ protocols from subjects as they interact with systems to solve complex tasks. We have employed a number of methods for analyzing this type of data (consisting of both audio and video recordings) using principled approaches to the analysis of process data from cognitive psychology and also the study of human discourse [10]. In borrowing from cognitive science, we emphasize a study of the *processes* involved in using systems, to complement more traditional outcome based measures and results from questionnaire based data. Such an integrated approach will be needed to address the critical issues in evaluation raised above.

3. Methodological components

The evaluations that we have developed and are currently employing involve several methods of on-line data collection (e.g. recording of patients’ usage of Web-based systems accessed from home), telephone interviews with users and in-depth video analysis of subjects using the system under study [6]. We feel that a multi-method approach is often needed in order to adequately address the multiple objectives typical of our evaluations. From our prior evaluation experiences, we have found that individual methods alone can provide valuable information (e.g. logging of user interactions, questionnaires, or interviews, etc.), but in order to gain an in-depth understanding underlying the use of a system, more than one complementary method may be required [11]. Fig. 1 depicts an overview of the evaluation methods that we are using, and these are described below.

3.1. On-line questionnaires in the study of the use of Web-based systems

For assessing the usability of Web-based systems, deployment of on-line questionnaires has been one of the most widely used approaches in evaluation of system use. However, questionnaires have a number of limitations, including their reliance on users’ recall of their interaction with a system when presented with questionnaires after a duration of time, and the low usage of ‘voluntary’ feedback forms, designed to provide Web-site developers with suggestions and comments. In this section, we describe the work that we have done in developing and extending application of on-line questionnaires and on-line Web-based forms to overcome some of these limitations for two objectives: (1) to collect baseline information about usage of systems, and then to present follow-up questionnaires, timed to appear over the Web over at key points during the period of an evaluation (e.g. after a specified number of logins, or invocations of a system function by a user); and (2) to design and deploy short forms/questionnaires to collect very specific usage information (timed to appear as close to the time of use of a system as possible, e.g. at

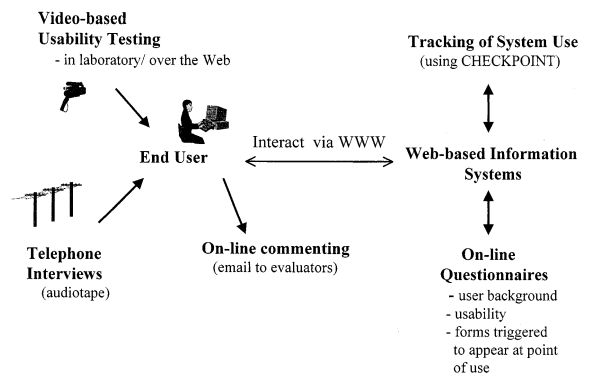


Fig. 1. ‘Teleevaluation’ of Web-based information systems, depicting evaluation methods employed for distance evaluation of Web-based information systems.

invocation of on-line guidelines at the point of care).

As an example, for our study of the use of PatCIS (a patient information system available over the WWW that allows patients to view and enter their own patient data over the Web [12]), several on-line questionnaires were developed using HTML (the standard hypertext markup language for developing Web applications). PatCIS was designed to handle presentation of questionnaires to subjects at login by invoking a function that determines if the user (i.e. the patient) has any outstanding questionnaires to fill out (which are presented to the user after login but prior to entering PatCIS, as described in Ref. [12]). The baseline questionnaire scales (presented to all users on their initial login) focus on assessing several major aspects of patient care and patient interaction with clinical information systems (see Fig. 2, for examples of the scales described below), including the following:

- **User Demographics:** Items regarding demographics of users of the system under study include standard questions about their age, sex, and educational background. This questionnaire is typically filled out at the beginning of the evaluation (upon first login by the user to the system being evaluated) to obtain basic demographic information.
- **Relationship with Health Care Providers:** This questionnaire contains items to assess patients' interaction with health-care professionals, including who they interact with and how often. Subsequent to an initial baseline presentation of this questionnaire, patients beginning to use an information system may also be asked whether they perceive their interaction with health-care professionals as changing (using a five-point scale ranging from 'no change' to 'considerable change').
- **Expectations About System Use:** The objective of this questionnaire is to assess users' expectations about the computer system under study, prior to actually interacting with the system. For example, this type of questionnaire was presented to patient users at the beginning of our study of PatCIS to assess the extent to which patients believed that use of the system would affect how they manage their illness (on a five-point scale, ranging from 'considerable change' to 'no change'), as well as how willing they were to use the system. A modified version of the questionnaire was presented to subjects later in the study to determine how subjects' expectations compare to their perceived experience. In addition, we are investigating to what extent the results of this questionnaire can predict (along with the results of the questionnaire on computer literacy, described below) how subjects may fare in interacting with innovative health-care technology.
- **Prior Computer Experience:** The accurate assessment of prior computer experience is an important aspect in understanding users' interactions with innovative information systems. Questions regarding prior use of computers are typically given at the beginning of our evaluations to determine the subject's level of computer literacy. For example, in our study of PatCIS, analyses were conducted to see how patients' level of prior computer experience predicted the use and problems encountered in learning and using patient clinical information systems. Scales were designed to assess how often the patient uses a computer, when he/she first started using a computer, and what type of computer systems (e.g. IBM-compatible systems running Windows, or Mac) and programs (e.g. word processing, email, etc.) are typi-

Background Questionnaire**A. Demographic Questions**

1. Check off the category which indicates how old you are:

0-19 years 20-39 years 40-64 years 65+ years

2. Check the box that indicates your sex

Male Female

3. Which of the following choices best describes who you live with?

Alone Family Extended Family Friends Significant Other

4. Check off the following that best describes the highest level of education you have achieved

Elementary School High School Technical School College (Arts) College (Science)
 University Bachelor Degree (Arts) University Bachelor Degree (Science) Professional Degree
 Postgraduate Arts Degree (MA or PhD) Postgraduate Science Degree (MS or PhD)

B. Computer Background Questions

1. How often do you use a computer?

Several times every day Once a day Several times a week Once a month or less often Never

2. When did you first start using computers?

In the past year In the past two years More than two years ago

3. Check off all the computer systems you currently use

IBM Compatible Macintosh Mainframe computer Other

5. Check off all the types of computer programs which you use or have used

Word processing Email Internet Databases Spreadsheets Programming

6. How often do you use the World Wide Web?

Several times a day Once a day Several times a week Once a month or less

7. Have you used the World Wide Web to obtain health care information?

Fig. 2. Excerpt from background questionnaire to assess users' computer experience and demographics.

cally used. The questionnaire also contains an item asking patients to indicate how computer literate they perceive themselves with regard to computer use (on a five-point scale). This questionnaire is filled out once at the beginning of the evaluation, to

obtain baseline data on patients' computer experience and literacy (a portion of this questionnaire is given in Fig. 2).

In addition to providing questionnaires over the Web to obtain baseline information about users at the start of a study, we have

also deployed questionnaires to users at different key points throughout the course of study of their interactions with health-care systems. For example, in our evaluation of PatCIS, approximately 9 months from the start of the study, we emailed all subjects who had signed up to participate a questionnaire (Fig. 3 illustrates several of the scales in this questionnaire) regarding their use of the system. Email was employed in this study as some of the subjects were no longer regular users of the system, but all could be reached by email. As will be described, the results from this type of questionnaire can provide insight into explaining the pattern of usage that may be obtained from the logging of user interactions. The questionnaire items include the following:

- Use of System Functions: Subjects are asked to indicate which functions they used and which they had not (along with a text entry box for them to indicate any comments). In addition, subjects were asked to indicate how often they used the system and from what locations.
- Usability Problems: The intent of this questionnaire is to obtain information about problems users may be encountering in interacting with an information system (i.e. problems in performing tasks that the system is designed to accomplish and general usability issues). The questions are based on standard usability scales for assessing user interfaces [13] and also contain text-entry boxes where subjects can enter their responses to the following questions: What features do you like? What features don't you like? What new features would you like? Were the reminders, alerts and guidelines provided helpful? What other types of on-line resources do you think would be useful? Users are also able to enter text-entry boxes to elaborate on any of the features and issues raised by the

questions. Our prior work has indicated that questionnaire data on usability alone may not provide sufficient information for determining how specifically to improve system and interface design [6]. Therefore, the results of this type of questionnaire should ideally also be complemented by other assessment instruments described below. In our current studies, we typically email the questionnaire to all participants in a study, including those who do not end up adopting use of a system, in which case, it is important to obtain information about why long-term use of a system is not adopted.

In our most recent work, we are using a variation of on-line questionnaires by presenting users with short Web-based forms that are triggered to appear at the time of specific interactions with systems. The objective is to assess the usage of system features and to assess how successful information systems are in meeting user needs in the context of specific uses of the system. For example, as will be described in detail in Section 4 on examples of our evaluations, we are currently collecting data on physicians' use of on-line clinical guidelines at the point of care. As a baseline, we are collecting usage statistics (using a tracking system we have developed that is described below, known as CHECK-POINT) on which guidelines are accessed and how often. While this is invaluable information, it does not tell us if the physicians are finding the guidelines useful, or whether such guidelines are applicable to their particular clinical situation. To address this, we have also developed brief questionnaires (i.e. deployed as Web-based forms) that appear *at the time* a user (e.g. a physician or patient) invokes a particular feature of a system over the Web (these forms are scripted to appear using a Perl program that triggers their invocation). As an example (described in detail

below), we wished to determine in which situational contexts physicians access clinical

guidelines from a particular site, i.e. guidelines from the American College of Physi-

1. How often do you use PatCIS?

- a. every day
- b. several times a week
- c. once a month or less
- d. less than once a month
- e. never

2. For what purposes do you use PatCIS?

[enter as many as applicable]

- a. to see my own patient data
- b. to enter my own data
- c. for educational purposes
- d. to get specific medical advice

3. Have you had problems in using PatCIS?

Answer (please describe any problems):

4. Is there any other information would you like made available to you that you cannot access through PatCIS? If so, what is it?

5. In general, are you able to find information you want using PatCIS?

PLEASE SELECT FROM THE FOLLOWING SCALES AND ENTER ANY COMMENTS YOU MAY HAVE

6. Text on the computer screen

Hard to read Easy to read
1 2 3 4 5 6 7

Comments? :

7. Graphs and tables

Easy to understand Hard to understand
1 2 3 4 5 6 7

Comments? :

8. Sequence of screens

Confusing Clear
1 2 3 4 5 6 7

Comments? :

9. Are the help/information buttons useful?

Not at all Very useful
1 2 3 4 5 6 7

Comments? :

Fig. 3. Excerpt from questionnaire about usability of PatCIS (Patient Clinical Information System).

cians–American Society for Internal Medicine (ACP–ASM). When physicians select ACP–ASM guidelines to view over the Web, a brief questionnaire immediately appears prompting them to select (from a menu) their location and context of system use (e.g. currently with a patient or not). After submitting the form, they are then redirected to the guideline. As will be described later, we are currently using this approach to collect detailed information about particular user interactions (correlated with other information such as records of browsing) that would be otherwise impossible to obtain or infer from usage logs.

3.2. *On-line commenting facility*

In our evaluation of systems such as PatCIS, end users (e.g. patients or physicians) are able to email any comments or concerns to the evaluators, and email is used by the evaluators to send messages to any patient subjects. Typically, subjects, at any time while using the system, can evoke a ‘mail-to’ function to directly send comments to the evaluators, as they occur to the subjects while using the system. Although subjects may be encouraged to send this type of comment, its use is typically infrequent (as is the filling out of voluntary ‘feedback forms’, which are available at most web sites). Thus, users also need to be probed for their comments (using questionnaires and interviews) as well as having their interactions tracked, as described below.

3.3. *Tracking system use*

In our studies of end-user interactions with health-care systems, we typically create log files capturing information about usage of system features that are automatically recorded for all subjects’ interactions with the

system under study. The information includes a record of functions accessed by the patient, buttons pressed, and time spent in each function. Previous evaluation of Web-based clinical information systems (designed for use by physicians) has shown that for the purposes of obtaining feedback for system improvement, such an analysis of log files can provide a rich source of data, particularly in determining which functions of a system may or may not be getting accessed by end users, as described by Cimino and Socratous [11]. In addition to constituting valuable data on how the system is being used, the log files are also used by the system for automatic selection and presentation of function-specific questionnaires and forms. For example, subjects whose log indicates that they are frequent users of a particular PatCIS function, such as its educational component, will be presented with on-line questionnaires for assessing use of that function in more detail.

In our recent work at McGill University, we have developed a system known as CHECKPOINT, which allows us to track users interactions with systems and Web sites that are located at remote locations. For example, using CHECKPOINT in Montreal, we are currently tracking use of on-line guidelines (located on a server in Philadelphia), which are being accessed by physicians in New York City. To accomplish this, users’ selection of web pages located at the remote (target site) are redirected to our server in Montreal, so that we can obtain a complete log of the users’ interactions with the site in Philadelphia (i.e. a record of pages accessed and length of browsing). CHECKPOINT acts as what is known as a gateway, which means that it works as an intermediary between the Web browser and the Web server. If every link that is followed by a user first goes through CHECKPOINT, it can keep a

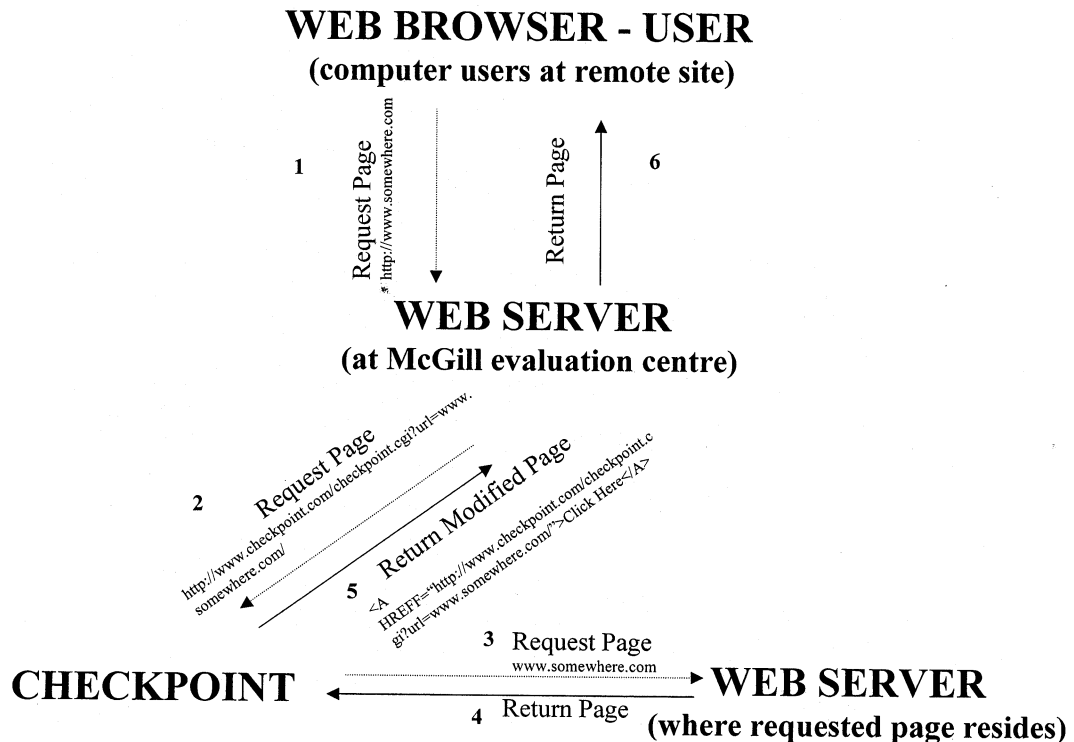


Fig. 4. Framework for tracking Web browsing remotely. Numbers designate the order of the processing steps in tracking a user's browsing.

log file of every site visited as well as a date and timestamp for each.

For example, as shown in Fig. 4, to start CHECKPOINT, an initial link must point to it. For example, as shown in the figure, when the user accesses a page with the address: `http://www.somewhere.com/`, CHECKPOINT is called using the syntax: `http://www.checkpoint.com/checkpoint.cgi?url=www.somewhere.com` (steps 1 and 2 in Fig. 4) where `www.checkpoint.com` is the location of CHECKPOINT. At this point, CHECKPOINT proceeds to request the page the user wants from `www.somewhere.com` (step 3 in Fig. 4). It then takes the HTML source code from this page and locates all the hyperlinks within the page. Once located, it changes all the links to the above form (i.e. redirected

through CHECKPOINT) to ensure that the next link followed also passes through CHECKPOINT. At this point, the user's Web browser is still waiting for the page to be returned and CHECKPOINT returns the modified HTML source code of `www.somewhere.com` (steps 5 and 6 in Fig. 4), which should be indistinguishable from the actual page, except that the addresses of the links are different.

CHECKPOINT, which is written in Perl, has been designed to be extensible, and we are currently using it in conjunction with other methods, described below, including presentation and scripting of on-line forms (also scripted in Perl), which appear at the point of use (e.g. at the time physicians actually choose to access guidelines over the

Web). Our use of a remote tracking system can be distinguished from other tracking systems (e.g. Lamprey, developed at Stanford [14]) by its integration and merging of output logs (from the tracking system) with results from other data collection methods in order to obtain a more complete and integrated picture of user interactions, which goes beyond recording of only user mouse clicks and pages browsed. For example, the logs from tracking users interacting with on-line information resources can be timestamped and written to a file interleaved with the questionnaire results from the presentation of Web-based forms at point of use (designed to probe the user about their interactions with the resource). Thus, a more complete record of the context of a user's comments can be obtained in terms of the interactions and browsing pattern that took place around the time the user fills out a Web-based form. We have found that the server that runs CHECKPOINT can allow for a large number of concurrent users and that the system is reliable (with little down-time).

3.4. Telephone interviews

In a number of our studies we have also conducted periodic telephone interviews with subjects regarding their use of the system, usability issues, cognitive and lifestyle issues and suggestions for improvements or refinements of the system. The approach taken is that of a semi-structured interview (with interview probes being designed prior to phoning end users), which may be conducted at several points in the evaluation study. We typically audio record the interviews (which last from 15 to 30 min) and transcribe them for analysis, which is based on methods described by Patel and colleagues, for a principled analysis of discourse from telephone interviews [15]. For example, in our study of

PatCIS, we asked patient users of the system to elaborate on their experience with the system, including discussing if they had encountered problems using the system or had suggestions for improvements. During the interviews, the subjects were encouraged to elaborate on their thoughts regarding their experiences with the system and the interviews were transcribed and coded to obtain details from the users' perspective on problems or concerns. These interviews have been extremely useful in shedding light on usage patterns that were obtained from pure tracking of user interactions, as will be described in a subsequent section.

3.5. Video-based usability testing

For a subset of subjects who volunteer to be recorded and are representative of typical end users of the system, we also typically employ in-depth cognitive usability testing methods to assess user interactions with systems (taking place in a 'laboratory' setting where patients are asked to interact with a system). This involves asking subjects to interact with the system under study (e.g. PatCIS) to perform typical tasks (e.g. data entry or review, access of information). Subjects are instructed to 'think aloud' while doing so, and complete video recordings are made of computer screens, as well as audio recording of their verbalizations. This allows researchers to obtain a complete record of the *process* of the use of the computer system under study. In our original work, the recordings were made using a PC-Video converter, with the video input to the VCR feeding in from the converter (i.e. the computer screens) and the audio input to the VCR feeding in from a microphone for picking up the subjects' verbalizations and comments as they interact with the system under study [6].

Using computer-aided methods for video analysis that we have developed and refined, problems encountered by patients while they interact with the system can be identified from such data [6] [10]. The approach taken to analyzing the video data involves the following: (1) transcribing all audio recorded verbalizations into a word-processing format; (2) linking the transcriptions in the word-processing file to the corresponding video sequences using a video annotation software package (and cable to connect the computer to the VCR) known as CVideo, which allows transcripts on a computer (opened in a word-processing file) to be linked to the corresponding frames on the video of the interaction; and (3) coding the corresponding sections of the transcriptions by ‘timestamping’ events on the video tape (e.g. the user opening of a window, or selecting from a menu, or a specific user problem, such as an inability to enter data in a field) with the corresponding transcriptions of subjects’ verbalizations. The annotated transcripts can be used to characterize user interactions in detail, including tabulation of the type and frequency of user problems coded for, and identified from, watching the video recording of the user’s interactions [16].

We have recently extended our video-based usability testing from the laboratory setting to application over the WWW from real settings. This has involved remotely video recording users interacting with medical applications at distant sites, through use of NetMeeting, a collaborative Web tool that allows for remote login and meeting capabilities. By logging on to user applications at remote locations using NetMeeting, at our evaluation site in Montreal, we can observe the computer screens of subjects interacting with information systems at a distance (the bandwidth requirements are low, since we do not employ all features of NetMeeting —

only the capability of recording screens remotely). By concurrently converting the computer screens to a movie format, using software such as Lotus ScreenCam, we are able to obtain a complete on-line recording of the users’ interactions at the remote site in real time. The corresponding audio recordings of subjects’ verbalizations (e.g. subjects ‘thinking aloud’ while interacting with a system, or discourse involved in collaborative projects) can be obtained using a microphone plugged into the user’s computer (or alternatively by using a speaker phone to hear the subject’s verbalizations and a microphone at the evaluation site to merge the audio with the video recording of the computer screens). Instructions may also be given to subjects using the phone system, or on an audio channel using the computer, allowing complete usability testing to be conducted remotely.

In order to capture not only computer screens at the remote site, but also behavioral actions (e.g. the end user shifting in his chair or visibly concentrating on the screen), a digital camera mounted on the computer at the end user’s site can also be employed to capture these actions over the Web. To illustrate our use of these methods, an example is provided in the next section of a usability study that we conducted to examine the processes involved in use of a computer-supported tools for encoding clinical guidelines into a computer-based representation (the recording was conducted at McGill, while subjects interacted with the system at the Harvard site in Boston).

4. Examples of teleevaluation

In this section of the paper, we provide examples of the methods and approaches described above. The first study involves the distance evaluation of a patient clinical infor-

mation system, known as PatCIS. This system, which was developed at Columbia University, provides patients with customized views of their own medical records via the WWW. PatCIS allows patients with chronic illnesses, such as diabetes and asthma, to enter their own health data and receive advice about the management of their illness from home [12]. The second example focuses on the evaluation of the use and development of computer-based clinical guidelines that physicians can access over the WWW. Both examples illustrate a number of the methods and techniques described above in the previous section on methodology.

4.1. Example 1: remote evaluation of a patient clinical information system (PatCIS)

For our initial study of PatCIS, recruitment letters were mailed to over 200 physicians. Permission forms were returned by 11 physicians, who suggested 11 patients that might be interested in participating over the year. Letters were sent to these patients, telling them how to register via the World Wide Web. Eight of these patients responded and were enrolled. The interactions of these physicians were studied in depth over a period of a year, using several methods described below.

4.1.1. System usage

In order to assess how often and what components of PatCIS were accessed by the patient subjects, monthly logs of all patient interactions were automatically generated, as described in Section 3.3. The patients who participated in using PatCIS logged on a total of 243 times during the course of the study. The first patient session was in April 1999; the pilot phase was continued through the end of February 2000 with a total of 31 patient months of use.

Of the 243 total login attempts, 33 initially failed due to incorrect passwords or SecurID codes. There were no attempts to login with an illegal ID, and all initially unsuccessful attempts were immediately followed by a successful login. Fourteen sessions involved a successful login, but no subsequent activity. Over the period of the study, interactions with the system remained relatively distributed among the eight subjects.

For security reasons, a 'Logout' function is included to allow patients to use PatCIS in public places (such as a library) and end their session without closing the browser application. They used this function 122 times and did not use it 74 times.

4.1.2. Function usage

The most frequently used function was the review of laboratory data, which was done at least once in 140 (71%) of the sessions. The 'Laboratory' sub-button shows a list of panels (CBC, Chem7, etc.) and allows the user to select a panel for detailed display. Users selected this function 270 times and examined details 340 times. The 'Laboratory Detail' sub-button shows all the panel details as a single list, rather than requiring the intermediate panel list. This option was selected 69 times. Selecting a specific test produces a summary of results for that test. Patients used this function 129 times (114 from Laboratory and 15 from Laboratory Detail). 'Reports' was the next most often used function (40 times). Patients selected a variety report headers to obtain details, including radiology (24 times), cardiology (17 times), and pathology (10 times). 'PFT' (Pulmonary Function Test) was selected 22 times and microbiology results eight times.

The other functions were used more sparingly than the review of data function. For example, vital signs were entered 31 times, and diabetes information (blood sugars and/

or insulin doses) was entered 14 times. These data were also rarely reviewed (26 and 18 times, respectively). Educational functions, providing passive links to other Web sites, were used 25 times (Diabetes 11 times, Geriatrics eight times, Home Medical Guide four times and Aging twice). Advice functions, which use data taken from the patient's record as input to active guideline programs, were used 16 times (Cholesterol 15 times and Mammography once).

4.1.3. On-line questionnaires and interviews

In order to investigate further the issues underlying these usage patterns and to assess patients' subjective experience in using Pat-

CIS, subjects received a questionnaire, which was emailed to them 9 months after their initial logins. The subjects were also asked to respond to questionnaire items (on a five-point scale) related to aspects of human-computer interaction, to assess the following:

- Willingness to enter data into PatCIS
- Willingness to review their own health data over the WWW
- Perceptions of their interactions with health-care providers.

Table 1 provides examples of several of the questions used, along with responses from the first four patients who participated in the study. The subjects were also asked to respond to questionnaire items (on a five-point

Table 1

Examples of questionnaire items and responses from four current PatCIS users (deployed 9 months after the user's first login)

Question	User 1	User 2	User 3	User 4
How often do you use PatCIS?	Once a month or less	Several times a week	Never	Once a month or less
For what purposes do you use PatCIS?	Review data Education	Review data	Review data Education Advice	Review data Advice
I find PatCIS useful	Definitely agree	Unsure	Unsure	Definitely agree
I am willing to enter my own data into my record using the WWW	Definitely agree	Agree	Unsure	Agree
I am willing to review my own health information using the WWW	Definitely agree	Agree	Agree	Definitely agree
PatCIS has improved my interactions with health professionals	Definitely agree	Agree	Unsure	Definitely agree
PatCIS has improved my understanding of health and illness	Definitely agree	Disagree	Agree	Definitely agree
PatCIS has changed how my health care is managed	Definitely agree	Unsure	Disagree	Definitely agree
Have you had problems in using PatCIS?	'Occasionally, it is very difficult to access, typically in evening'	None		'Sending e-mail from the site'

scale) related to aspects of human–computer, to assess the following:

- Ability to understand graphs and tables
- Clarity of screen sequences
- Usefulness of help and information buttons
- Learnability of the system
- Usefulness of linkages to other sites
- System reliability and speed.

Results to date indicate that all eight users found that information presented was presented on the screen in a way that was easy to read, graphs and tables were comprehensible and that overall the system was reliable. However, from our preliminary analysis, areas where user responses indicated that the system might use improvement included improving error and system messages, as well as streamlining the sequence of screens.

Table 2 provides the overall impressions of PatCIS from four patients who have used the system for at least 9 months. The most positive ratings came from subject 4, who, from examination of the baseline questionnaire, had the most extensive computer background. In addition, analysis of log files indicated that this subject used the various features of the system most extensively. In contrast, user 3 had the least computer experience and reported considerable trouble in attempting to login and access PatCIS (this subject stated that he had trouble in using the Secure ID card for gaining access to PatCIS).

In addition to logging user accesses and emailing usability questionnaires, interviews were conducted with users to assess for example, why they find PatCIS useful or not useful (these interviews were conducted over the telephone, and user's responses to questions and their 'thinking aloud' were audio-taped and transcribed). We also contacted patients who enrolled in the study who did not use the system (according to our log files) in order to determine why. Analysis of interviews indi-

cates that users may have not been fully aware of all of the more advanced features of PatCIS (such as the advice function), which may have been the reason for the low usage of some of these functions, as described above (despite the questionnaire results indicating a high degree of willingness to use the system). In addition, patients indicated that the review of data function was very useful in allowing them to follow their own results over time and that this improved communication with their physicians (consistent with the logging data indicating that this was the most frequently used function).

Using the multi-method approach described above, we are working on identifying problems and issues of those patients who have adopted use of the system. We are now following up with more detailed targeted usability testing of potentially problematic areas based on this pilot data (e.g. sequencing of screens, usefulness of error messages, need for user training and relatively low usage of data entry and advice facilities).

4.2. Example 2: evaluation of the use and generation of Web-based clinical guidelines

We have employed a number of our methods described above to assess the use of clinical guidelines by physicians (in a clinical practice setting) using the WWW. The approach taken involved the tracking of use of guidelines (from ACPOnline) as physicians use a Web-based CPR system. The purpose of this work is to characterize how guidelines are actually being used by physicians in clinical practice, as well as their effectiveness. Questions to be answered included the following: (a) Why do physicians access Web-based clinical guidelines? (b) Are patients present when they are accessed? (c) Do physicians find the guidelines useful and applicable, and, if not, what are the reasons? In the

Table 2
Overall impression of PatCIS of four users after a 9 month period

User 1	User 2	User 3	User 4
<p>‘PatCIS makes it much easier to actively participate in my wife’s medical care by allowing me to closely monitor her lab results. It also allows me to more intelligently interact with her doctors.’</p>	<p>‘I find it very good for my purposes. By and large, I have no idea to improve it. However, I remain confused about getting it from a remote computer.’</p>	<p>‘I feel as though I am in a very unique position, I know what my problem(s) are, and it’s just a matter of time.’</p>	<p>‘Overall program is excellent. Would be better if you could access it over Internet Explorer as that is what most people have. Could present a problem with people who have old machines with limited memory.’</p>

area of evidence-based medicine, a considerable number of papers have appeared describing both the advantages [17] and limitations of clinical guidelines, particularly in the context of actual clinical practice. However, few studies have reported hard empirical data on how guidelines are actually used at the point of care, due to difficulties in obtaining this type of information. In the example described below, we employed our methods in an attempt to identify the following: (1) the purpose of use of guidelines by physicians in a clinical practice setting; (2) the patient setting in which the guidelines were applied; (3) the quality of information provided by the guidelines as assessed by the physician; and (4) the specific reasons for reduced effectiveness or applicability of guidelines. Ideally, this information will allow us to determine how guidelines (and their presentation methods) may be enhanced in order to improve their usability at the point of care.

4.2.1. Data collection using triggered on-line forms

Usage tracking and data collection were based on methods described in Section 3. Specifically, user interactions with ACPOne were recorded during accesses to evidence-based guidelines by physicians from Columbia University. Physicians were able to access a reference page directly from a Web-based computerized patient record system. From the reference page, they could select for viewing on-line guidelines for specific diseases from ACPOne via the WWW (illustrated in Fig. 5a). The user's selections were captured using the CHECKPOINT tracking system (described in detail above) developed for remotely tracking use of Web-based information systems.

After clicking on a particular guideline, the users were presented with a Web-based form

(illustrated in Fig. 5b) asking them to indicate their intended purpose for accessing the guideline (e.g. 'To help me care for a specific patient', 'For general knowledge/education', 'Other'), as well as the clinical context in which the guideline was accessed (e.g. 'I am seeing the patient now in an out-patient setting', 'I am seeing the patient now in an in-patient setting', 'I saw the patient previously', 'Other'). Once this information was entered, the users submitted the form by selecting a button at the bottom of the form. The guideline was then presented to the users, and their interaction with the guideline (i.e. browsing at the ACP site where the guideline resides) was tracked.

Upon exiting the guideline, a second Web-based form was presented to users (illustrated in Fig. 5c) asking them to indicate the quality of information they received from the guideline (e.g. 'I got all of the information I wanted from the guideline', 'I got some of the information I wanted from the guideline', 'I did not get the information I wanted from the guideline'). Furthermore, users who did not find the information they wanted from the guideline were asked to indicate the reason for this (e.g. 'The guideline did not apply to the specific patient I was considering', 'I could not find the information relevant to what I needed to know', 'The guideline may be incorrect', 'The guideline is not current', 'Other'). The users exited the system upon submission of this form.

4.2.2. Results: the use of on-line clinical guidelines at the point of care

The responses of physicians indicating the purpose of accessing guidelines is shown in Table 3. The purpose of use was characterized as being for (1) general knowledge/education, (2) care for a specific patient or (3) other reasons. The percentage of uses falling into each of these categories is shown in

The screenshot shows a Netscape browser window displaying a website titled "Clinical References - Netscape". The address bar shows the URL: <http://cpmnet.columbia.edu/clinical/ref.html>. The page content includes several sections:

- Clinical References**
 - [Columbia University College of P & S Complete Home Medical Guide](#)
 - [Guide to Clinical Preventive Services](#)
 - [Tuberculosis Information for Providers and Patients](#)
 - [Medicare Part B: Documentation Guidelines](#)
- Guidelines from the American College of Physicians website:**
 - [Lyme Disease: Guidelines for Laboratory Evaluation in the Diagnosis of Lyme Disease](#)
 - [Lyme Disease: Laboratory Evaluation in the Diagnosis of Lyme Disease](#)
 - [Coronary Artery Disease: Perioperative Assessment and Management of Risk from Coronary Artery Disease with Major Noncardiac Surgery](#)
 - [Coronary Artery Disease: Guidelines for Assessing and Managing the Perioperative Risk from Coronary Artery Disease Associated with Major Noncardiac Surgery](#)
 - [Thyroid Disease: Screening for Thyroid Diseases](#)
 - [Thyroid Disease: Screening for Thyroid Disease: An Update](#)
- The Following Resources are Restricted to CPMC**
 - [MEDLINE](#) and other bibliographic databases of the medical literature

At the bottom of the page, there is a message: "Medline -- Access for CU employees, students, etc. Must have a valid CUMX id to connect from an outside Internet provider like AOL or IPM Global Network. http://71.32.206.83:20069/bv/columbia_project/feedback.cgi?

The browser's status bar at the bottom shows the time as 4:32 PM and the URL: http://71.32.206.83:20069/bv/columbia_project/feedback.cgi?

Fig. 5. (a) Resource page showing links to clinical guidelines available from within a computer-based patient record system.

Please Take A Moment To Complete The Following:

Please Indicate the Purpose of Accessing the Guideline:

To help me care for a specific patient

For general knowledge/education

Other

If the answer to the above question was "for a specific patient", please indicate the patient-care setting:

I am seeing the patient now in an out-patient setting (e.g. clinic, emergency room, private office, etc.)

I am seeing the patient now in an in-patient setting (e.g. hospital admission)

I saw the patient previously

Other

Fig. 5. (b) Form to assess a clinician's reason for accessing a guideline (which appears when the user selects a guideline from the resource page).

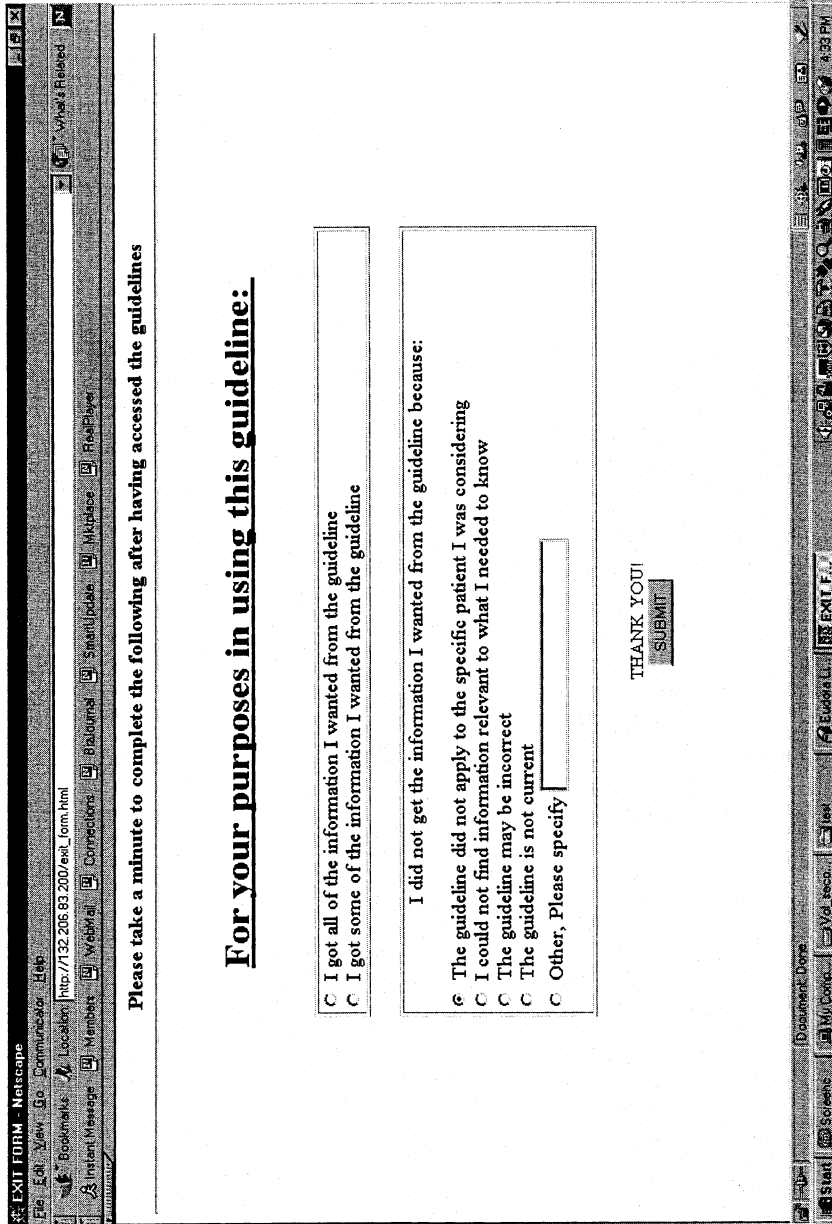


Fig. 5. (c) Form to assess success in obtaining desired information from the clinical guideline accessed (which appears on exit from the guideline).

Table 3
Results from an analysis of use of on-line guidelines by physicians at the point of care

Purpose of use	Frequency	Percentage
General knowledge/education	128	81
Care for a specific patient	22	14
Other	8	5
Total	158	100
<i>Patient setting</i>		
Seeing patient now as an out-patient	9	41
Seeing patient now as an in-patient	6	27
Saw patient previously	7	32
Total	22	100
<i>Quality of information</i>		
Got all information wanted	19	46
Got some of the the information wanted	10	25
Did not get the information wanted	12	29
Total	41	100
<i>Reason for not getting desired information</i>		
Guideline did not apply	6	50
Could not find relevant information	5	42
Guideline may be incorrect	1	8
Total	12	100

Table 3. Specifically, it appears that the majority of physicians accessing these guidelines on-line are doing so for general knowledge or educational purposes. Thus, physicians were using the guidelines for the care of a specific patient only a fraction of the time.

Physicians who indicated that they were accessing the guideline for the care of a specific patient were asked to indicate the patient-care setting in which this was done. The responses indicating the patient-care setting are shown in Table 3. The patient-care setting

was characterized as being (1) in the presence of the patient, in an out-patient setting, (2) in the presence of the patient, in an in-patient setting, or (3) for a patient seen previously. The percentage of uses in each of the patient-care settings is shown in Table 3. Specifically, the use of guidelines in each of the three main patient-care settings was fairly even. Furthermore, roughly 68% of physicians using the guidelines for the care of a specific patient did so in the presence of that patient. Therefore, it appears that when physicians do access guidelines for the care of a specific patient, it is done most frequently at the time of care.

Upon exiting ACPOne, physicians were asked to characterize the quality of the information that they received from the guidelines. Their responses are shown in Table 3. Specifically, physicians indicated whether they (1) obtained all of the information they wanted, (2) obtained some of the information they wanted, or (3) did not obtain the information they wanted. The percentages of responses falling into each of these categories is shown in Table 3. Roughly half of the physicians did find the information they were seeking in the guidelines, while the other half obtained some or none of the information they required. However, this does seem to imply that there are indeed problems or barriers that may be reducing the effectiveness with which physicians are able to apply guidelines in clinical practice.

Physicians who indicated they did not receive the information they wanted from the guidelines were asked to indicate the reason for this. Their responses are shown in Table 3. Specifically, the reason for not obtaining the desired information was characterized as being because (1) the guideline did not apply to the specific patient they were considering, (2) they were unable to find the relevant information, or (3) the guideline may have

been incorrect. The percentage of responses in each of the categories is shown in Table 3. The results indicate that half of the physicians did not find the information they were seeking because the guideline did not apply to the specific patient they were considering. Furthermore, 42% of the physicians were unable to find information relevant to what they needed to know, while 8% indicated that the guideline may be incorrect. Therefore, these findings support the implication that the current applicability and effectiveness of guidelines in clinical practice may be issues that need to be addressed.

To investigate further the patterns emerging from the data we are collecting, as described above, we plan to extend our current analyses to more clinical settings and a greater number of physicians. We will continue to track user interactions using CHECKPOINT and present users with Web-based forms that are timed to appear upon invocation of guidelines. It should be noted that baseline tracking of user interactions with the guidelines (for a period prior to the additional presentation of the forms) is important in order to assess any affect that completing the forms may have on the users' normal interactions. For example, in the study described above, we found no difference in the frequency of accesses to guidelines prior to and after the forms were added. In future work, we will complement the collection of data using tracking and forms to include telephone interviews with physicians using the guidelines who indicate that they are willing to discuss their interactions. We have found in our other studies that such interviews can provide information that can help to explain the patterns of behavior recorded during tracking, as well as the results obtained from presentation of Web-based forms at point of use.

4.2.3. Distance usability testing of the process of guideline encoding

In a related study, we are also examining the processes involved in the *generation* of on-line guidelines using computer tools. We are using the method described earlier in this paper for conducting distance usability testing, using remote recording of computer screens via NetMeeting and collection of audio recordings of subjects 'thinking aloud'. Specifically, we are studying physicians (at a site at Harvard) who are using computer tools to encode paper-based clinical guidelines into a computer-based knowledge representation known as GLIF [18]. By encoding guidelines in this format, the guidelines can be deployed over the WWW and integrated into health-care systems, such as computerized patient records. The evaluation procedure has involved video-recording the subjects as they take paper-based guidelines (e.g. for treating thyroid problems) and convert them into GLIF, using a special graphical editing program developed at Harvard. All computer screens are video-recorded at the McGill site (using NetMeeting to view the subject's interactions with the system at the remote Harvard site), with the audio input coming from the subject's verbalizations (recorded over a speaker phone at the McGill site). The data have been coded to identify categories of problems (e.g. inability of the subject to encode a step in the paper guidelines into computer format, or a need for further information about the clinical context of the guidelines).

Through our preliminary work in this area, described above, we have been able to characterize the steps taken in modeling guidelines using GLIF and identify potential difficulties encountered in the task of encoding guidelines. For example, difficulties were encountered by the subjects in deciding which knowledge representation was most appropri-

ate for (1) modeling the procedural knowledge required for the GLIF developers to encode the domain knowledge, and (2) modeling the general informational statements to be placed in the body of the guidelines. The approach taken has led to identification of both generic aspects of the process of guideline encoding as well as specific problems in doing this task. Regarding the guideline resource material used by the subject in carrying out this task, key background information (e.g. regarding normal ranges) was found to be needed in order to encode the guideline. The information from our evaluation of the physicians' interaction with this system is currently being used for feedback into the re-design and improvement of computer tools used for encoding guidelines, as well as for iteratively improving the underlying guideline GLIF representation and guideline content.

5. Discussion

In this paper, we have described our approach to evaluation, including examples of an assessment of a patient clinical information system (PatCIS), as well as an evaluation of physicians' use of clinical guidelines over the WWW. In designing such evaluations, we adopt a multi-method approach in order to assess a variety of related questions (ranging from assessment of user expectations to analysis of specific interface problems). 'Teleevaluation' methods, involving evaluation from a distance, have a number of distinct advantages, including the ease in creating and administering evaluation instruments via the web. However, previous work in assessing the use of systems by physicians and patients [4,6] has indicated that questionnaires alone may need to be supplemented by other techniques, such as interviews with patients

(which can be conducted via telephone communication), and the use of in-depth usability testing methods, which can now be conducted over the WWW. In this way, the limitations of any one method can be offset by the advantages of another. For example, questionnaires and on-line forms may tell us what users think they may be doing while interacting with a system, but automatically logged data of user interactions may provide more detail on what they actually do (which is often not the same [4]). In addition, in-depth usability studies (involving video recording) can provide some insight into the reasoning and decision-making of subjects as they interact with systems to solve complex tasks.

An important aspect of the distance evaluation of Web-based information systems has been the development of effective on-line questionnaires. The use of HTML and a variety of editors that generate HTML code greatly facilitates this process. Prototype mock-ups of questionnaires can be rapidly tested as they would appear to end users, by using web browsers to display the HTML-coded questionnaires. In the context of our current work, we have extended a number of our methods for iterative testing and refinement of computer systems [6] to the testing of questionnaire content and display. This has involved observing pilot subjects interacting with prototype versions of the questionnaires, while 'thinking aloud' (verbalizing their thoughts) regarding the questionnaire's layout and content. This information can be used to improve the design of the questionnaires prior to deployment over the Web. One of the most critical stages in our evaluation studies is selecting from the multiple potential methods an approach that will meet the overall objectives of the evaluation. Given the multiple objectives of our studies, several complementary methods have been

included in the design of the evaluation to achieve a broad assessment of factors such as system usability. An important consideration in administering particular components of our evaluations is whether the evaluation instrument (e.g. a particular questionnaire) is to be administered once, at the beginning of the study (to obtain baseline data) or at several points throughout the entire period of the study to assess change. Analysis of on-line data collected from patients' interactions with systems on the World Wide Web is greatly facilitated by the fact that electronic data collection can be automated. For example, CGI scripts can be written to process data automatically (e.g. perform summary statistics) as they are entered (e.g. via on-line questionnaires) and trigger the presentation of on-line forms automatically to assess aspects of usability at the actual point of care.

At McGill University, we are currently involved in developing a 'virtual' teleevaluation laboratory. At the core of this facility is software described in this paper for remotely tracking end users of systems. Our system at the McGill site essentially acts as a server for conducting evaluation of system use at distant sites. In addition, all data collected from our studies goes to the McGill site, and multiple sources of information (e.g. buttons clicked and results from questionnaires and on-line forms) are automatically merged at the McGill site. At the remote site, our approach involves simply linking applications to our virtual laboratory (e.g. rerouting end users' inputs to McGill). At the McGill site, all the script programs that process information (e.g. that present on-line questionnaires, log end user browsing, etc.) reside on our server. This architecture is transparent to the users interacting with the system and has the advantage of being able to interface with systems being evaluated at remote sites by simply linking to those sites through our

server. It has been our experience that the initial investment in the development of a virtual evaluation laboratory (including the purchase of server hardware and development of tracking programs) can be offset by using the laboratory in the evaluation of a number of health-care information systems, involving only a minor modification of the laboratory's software to accommodate a diverse range of assessments. Our experience in successfully conducting distance evaluations suggests that major development efforts involving health-care information systems could, and perhaps should, be subjected to some form of evaluation even if the facilities are not readily available at the actual site of system deployment. We plan to extend our current laboratory to include equipment used for conducting high-bandwidth teleconferencing methods in order to conduct remote studies of aspects of system use involving collaborative team decision making.

In conclusion, we feel that an approach to evaluation in medical informatics that takes into account both technological advances in the systems that we are evaluating, as well as methods and theoretical frameworks from areas such as cognitive science, promises to provide us with detailed information on system use. This information will prove to be invaluable in the iterative process of design, evaluation and re-design that our complex systems will inevitably require.

Acknowledgements

The studies reported in this paper were conducted at McGill University under direct contractual and sub-contractual agreement from the funding agencies. The authors would like to thank Dr. E.H. Shortliffe for his suggestions and comments in the development of our web-tracking component

CHECKPOINT. This work was supported in part by a National Information Infrastructure contract N01-LM-6-3542 from the National Library of Medicine, a grant from the Canadian Centres of Excellence network HEALNet and a grant from the American College of Physicians–American Society for Internal Medicine (ACP–ASM).

References

- [1] J. Nielsen, *Usability Engineering*, Academic Press, New York, 1993.
- [2] R. Kittredge, U. Rabbani, F. Melanson, G.O. Barnett, Experiences in deployment of a Web-based CIS for referring physicians, in: D.R. Masys (Ed.), *Proceedings of the 1997 Fall AMIA*, Hanley & Belfus, Philadelphia, PA, 1997, pp. 320–324.
- [3] R.D. Cork, W.M. Detmer, C.P. Friedman, Development and initial validation of an instrument to measure physicians' use of, knowledge about and attitudes toward computers, *JAMIA* 5 (1998) 164–176.
- [4] A.W. Kushniruk, V.L. Patel, Cognitive evaluation of decision making processes and assesment of information technology in medicine, *Int. J. Med. Inform.* 51 (1998) 83–90.
- [5] Patel V.L., Kushniruk A.K., Yang S., Yale J.F. Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning, *JAMIA* 7 (6) (2000) 569–585.
- [6] A.W. Kushniruk, V.L. Patel, J.J. Cimino, Usability testing in medical informatics, in: D.R. Masys (Ed.), *Proceedings of the 1997 Fall AMIA*, Hanley & Belfus, Philadelphia, PA, 1997, pp. 218–222.
- [7] V.L. Patel, D.R. Kaufman, Science and practice: A case for medical informatics as a local science of design, *JAMIA* 6 (1998) 489–492.
- [8] V.L. Patel, J.F. Arocha, D.R. Kaufman, Diagnostic reasoning and medical expertise, *The Psychology of Learning and Motivation* 31 (1994) 187–252.
- [9] Patel, VL, Arocha, JF, Kushniruk, AW.. EMR: Re-engineering the organization of health information. *Methods of Information in Medicine* (in press).
- [10] K.A. Ericsson, H.A. Simon, *Protocol Analysis: Verbal Reports as Data*, MIT Press, London, 1993.
- [11] J.J. Cimino, S.A. Socratous, Just tell me what you want!: The promise and perils of rapid prototyping with the World Wide Web, in: J.J. Cimino (Ed.), *Proceedings of the 1996 Fall AMIA*, Hanley & Belfus, Philadelphia, PA, 1996, pp. 719–723.
- [12] J.J. Cimino, S. Sengupta, P.D. Clayton, V.L. Patel, A.W. Kushniruk, X. Huang, Architecture for a Web-based clinical information system that keeps the design open and the access closed, in: C. Chute (Ed.), *Proceedings of the 1998 Fall AMIA*, Hanley & Belfus, Philadelphia, PA, 1998, pp. 121–125.
- [13] B. Schneiderman, *Designing the User Interface*, Addison-Wesley, New York, 1992.
- [14] R.M. Felciano, R.B. Altman, Lamprey: Tracking users on the World Wide Web, in: J.J. Cimino (Ed.), *Proceedings of the 1996 Fall AMIA*, Hanley & Belfus, Philadelphia, PA, 1996, pp. 757–761.
- [15] E.H. Shortliffe, V.L. Patel, J.J. Cimino, G.O. Barnett, R.A. Greenes, A study of collaboration among medical informatics research laboratories, *AI Med.* 12 (2) (1998) 97–123.
- [16] A.W. Kushniruk, V.L. Patel, J.J. Cimino, R.A. Barrows, Cognitive evaluation of the user interface and vocabulary of an outpatient information system, in: J.J. Cimino (Ed.), *Proceedings of the 1996 Fall AMIA*, Hanley & Belfus Inc, Philadelphia, PA, 1996, pp. 22–26.
- [17] C.J. McDonald, Medical heuristics: The silent adjudicators of clinical practice, *Ann. Intern. Med.* 124 (1) (1996) 56–62.
- [18] L. Ohno-Machado, J.H. Gennari, S.N. Murphy, N.L. Jain, S.W. Tu, D.E. Oliver, E. Pattison-Gordon, R.A. Greenes, E.H. Shortliffe, G.O. Barnett, The guideline interchange format: A model for representing guidelines, *JAMIA* 5 (4) (1998) 357–372.