Involving users in the design and usability evaluation of a clinical decision support system

Carmen Carroll a,*, Phil Marsden b, Pat Soden c, Emma Naylor b, John New a, Tim Dornan a

a Department of Diabetes/Endocrinology, Hope Hospital, Salford, Manchester M6 8HD, UK
b Department of Information Systems and Multimedia, School of Computing and Mathematics, University of Huddersfield, Huddersfield HD1 3DH, UK
c Department of Computing and Maths, University of Salford, Salford M5 4WT, UK

Received 6 February 2001; received in revised form 20 August 2001; accepted 20 September 2001

Abstract

Aim: To design and evaluate a clinical decision support system (CDSS) to support cardiovascular risk prevention in type 2 diabetes. Methods: A preliminary requirements specification and three prototype CDSS interface designs were developed. Seven patients and seven clinicians conducted ‘usability tests’ on five different task scenarios with the CDSS prototypes to test its effectiveness, efficiency and ‘user-friendliness’. Structured, qualitative questions explored their preferences for the different designs and overall impressions of clinical usefulness. Results: Patients and clinicians were enthusiastic about the CDSS and used it confidently after a short learning period. Some patients had difficulty interpreting clinical data, but most were keen to see the CDSS used to help them understand their diabetes, provided a clinician explained their results. Clinicians’ main concern was that the CDSS would increase consultation times. Changes suggested by users were incorporated into the final interface design. Conclusion: We have successfully incorporated patients’ and clinicians’ views into the design of a CDSS, but it was an arduous process. © 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Usability; Evaluation; Clinical decision support systems; Cardiovascular risk

1. Introduction

Using computers to support medical care has been an ongoing concern since at least the 1950s [1]. New disciplines, such as Artificial Intelligence in Medicine (AIM) and Medical Informatics, have emerged as scientists and physicians have come to appreciate the potential of computers to support information management within health organisations, and their potential to support health care itself as ‘intelligent agents’ [1,2]. The British Government is now committing itself to developing a ‘computational infrastructure’ within the National Health Service (NHS), through deployment of the NHS-Net and the development of electronic patient records [3,4].
Despite this investment, health-care institutions have been slow to incorporate information technology (IT) successfully into the work environment and more than half of medical information systems developed for clinical practice fail because of staff resistance [1,5,6]. Difficulties encountered with previous computer systems included their failure to fulfil a perceived clinical need, imposition of changes in the way clinicians worked, and their ‘unfriendliness’ in being difficult to use and providing no explanation of the reasoning behind proposed problem solutions [1,7,8]. It has been suggested that, in medicine, it is not enough for computers to perform well, but that they should also be compatible with the professional values and autonomous working ethics of physicians [9]. The lack of clinician input into the development of such systems has been cited as a major reason for their failure [10]. For these and other reasons, the IT community has increasingly acknowledged the role of human factors and ‘usability engineering’ in computer application design and evaluation [12,13].

Several studies have demonstrated the beneficial effects of improving cardiovascular risk factors on patients’ cardiovascular outcomes [13–16]. Patients with diabetes are an extreme case of susceptibility in that they have been shown to have as high a risk of suffering a cardiac event as non-diabetic patients with a previous history of myocardial infarction [17]. To address cardiovascular risk successfully with patients, clinicians need to estimate patients’ individual risks quickly, and provide information about the relative impact of individual factors on overall cardiovascular risk [18]. However, clinicians are not good at assessing risk without tables or other calculators [18]. Evidence from systematic reviews also shows that decision aids improve patients’ knowledge of diseases and that computerised prompting systems improve clinicians’ performance [19,20].

We have conducted qualitative interviews and focus groups with clinicians and type 2 diabetic patients to explore their perceptions about cardiovascular disease [21]. These showed that this group of diabetic patients was unaware of the risk diabetes per se conferred on them and had an incomplete understanding of cardiovascular risk. In addition, they were mostly computer illiterate and afraid of using computers themselves, although not averse to seeing computers in clinic. However, clinicians were positively disposed towards the increased and imaginative use of IT to support patient care. Throughout both focus groups and interviews, patients consistently expressed a wish to see their own data presented to them in clinics in a format they could easily understand [21].

Salford has a comprehensive, population-based Diabetes Information System, which currently presents data to clinicians on paper in a tabular form. Therefore, we decided to develop a computerised clinical decision support system (CDSS) that would:

- Download patient-specific data from the Salford Diabetes Register to this novel IT application.
- Display these data in a graphical/illustrative format.
- Calculate patients’ cardiovascular risk using patients’ data and current epidemiological evidence.
- Introduce information technology into clinical consultations in a way that supports the patient/professional interaction, educates patients about their individual risk factors and motivates behaviour change (of both patients and clinicians).

A preliminary specification of the CDSS was drawn up using information derived from the interviews and focus groups. A specified goal of the CDSS was to encourage physicians to use this application to illustrate to patients their individual data and cardiovascular risk, thereby providing them with a visual image of their ‘medical condition’, as well as a verbal explanation of their results. As well as this the CDSS would draw clinicians’ attention to patients’ abnormal results and facilitate a more detailed discussion of them. It was felt that a combination of detailed explanations and visual images of their results, would help patients understand their condition better, thus motivating them to make appropriate behaviour changes. Thus the CDSS would act both as an educating and motivating tool for patients, as well as providing a means of bringing care up
to a minimum standard to patients. Clinicians were designated ‘primary stakeholders’, as they would actually use the application, and patients as ‘secondary stakeholders’ as they would benefit from seeing the application, although not expected to use it themselves.

The aim of this paper is to describe the formative evaluation of this application, using ‘usability tests’ with patients and clinicians, and in so doing demonstrate the applicability of this design philosophy to healthcare-related software development.

2. Methods

2.1. Usability testing

The term ‘Usability’ means the ease with which a system can be learned and has been defined by the International Standards Organisation (ISO) as:

[The usability of a computer product is] the extent to which the product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use (ISO 9241; 11: 1994).

Nielsen suggests IT applications should display five major usability attributes [22]. A system should be:

- Easy to learn—learnability
- Efficient to use, making it highly productive—efficiency.
- Easy to remember, so that the a casual user is able to use the system easily after a period of non-use—memorability.
- Relatively error-free, so that users make few errors and recover easily from those they do make—errors.
- Pleasant to use, so that users like it—satisfaction [22].

These attributes are not intuitive and have often been ignored by designers [23,24]. To address this, Gould et al. suggested that software should be designed around some fundamental usability principles: an early focus on users; integrated design; and early and continual user testing [24,25]. This focus on users as major ‘players’ in the design process is now an established principle, and certainly applies to the development of clinical decision support systems [26].

The term usability testing means getting ‘real’ users to test the usability aspects of a system’s design. Such testing is invaluable as it provides direct information about the exact problems people encounter with the particular interface under scrutiny [22,27,28].

The success and efficiency of human–computer interaction is determined by the efficiency of the system and the quality of the ‘man–machine interface’ [11]. Interface design is thus an important part of software development, and cannot be totally separated from it [29]. Similarly, design and evaluation are inter-linked, where evaluation is regarded as part of design [26,30,31]. Hence, usability testing becomes an integral part of the overall design process, and has been shown to result in likeable computer applications [24,32].

This paper describes a process of usability testing, involving ‘real’ end-users of the proposed application, conducted with limited resources in a health-care setting. For the purpose of this evaluation both clinicians and patients were designated ‘end users’ as, although clinicians would use the application, the material displayed on the computer screen was largely aimed at patients.

2.2. Description of the application

A colour-coded ‘risk thermometer’, labelled ‘low’, ‘moderate’, ‘high’ and ‘very high’, was used to denote patients’ absolute 10-year cardiovascular risk (Fig. 1). This representation of risk was chosen as patients in Salford were already familiar with seeing their glycaemic control depicted on a thermometer. The risk-calculating algorithm recommended by the Joint British Recommendations on Prevention of Coronary Heart Disease Guidelines, derived from the Framingham and Framingham Offspring studies, was chosen to calculate patients’ individual risk [18]. A written ‘numerical’ or ‘percentage’ risk was included next to the ‘risk thermometer’ to aid clinical decision-making.
A ‘Predict’ function allowed clinicians to alter biological results temporarily by means of slide bars so that both they and patients were able to assess the relative contribution of various risk factors to overall risk (Fig. 1). By using slide bars, clinicians could manually alter the numerical values of certain risk factors, such as cholesterol or blood pressure. These altered parameters were automatically re-entered into the risk-calculating algorithm, which recalculated the patient’s risk. Thus clinicians could illustrate the relative impact various factors had on overall cardiovascular risk to the patient. By pressing the ‘revert’ button on the ‘predict’ screen, the value of all parameters would return to the patient’s actual values. None of the manually altered values were stored in the database from which the patient’s actual values were retrieved, thus maintaining the integrity of the original data.

Patients’ individual data were represented both as linear graphs over time and in a numerical, tabulated form. Linear graphs were chosen to represent patient data as previous experience indicated that patients have difficulty interpreting histograms [33]. The parameters displayed included: blood pressure, glycated haemoglobin (HbA1c), weight (denoted by Body Mass Index: weight in kilograms/(height in meters)²), and cholesterol. Total cholesterol, high density lipid (HDL) cholesterol, low density lipid (LDL) cholesterol and triglyceride concentrations were displayed as pie charts. Locally agreed targets for each parameter were shown as horizontal lines on the graphs, the target values equating with the ‘y-value’ at which the line is drawn (Fig. 3). Finally, a prompting system allowed clinicians to identify easily abnormal risk parameters, which could then be targeted in clinic consultations with patients.

Fig. 1. Screen print showing Version 1 of the CDSS prototypes: graph colours are grey; navigation uses ‘buttons’; abnormal risk parameters are shown in red. Separate window for ‘Predict’ function below, where slide bars can be moved to temporarily alter risk parameters.
Three user interface functional prototypes of the CDSS were designed using the Borland Delphi Development System (Figs. 1–3). All were in colour and in a ‘Windows’ format [34,35]. All contained the same trial data and performed the same functions, but differed in their navigation and design. These three prototypes were evaluated during formal ‘usability testing’, with the aim of deriving a ‘final’ or ‘composite version’ of the CDSS for use in consultations with patients.

2.3. Usability tests

Seven clinicians (three diabetologists and four diabetes nurse specialists), all working at Hope Hospital Diabetes Centre, performed the usability tests. Ten patients with type 2 DM, attending the diabetes centre in October 1999, were also recruited. After explanation of the study, three refused to take part; the remaining seven completed the usability tests. The study was approved by the Local Research and Ethics Committee.

Formative evaluation of the CDSS involved formal inspection of the computer system and its interface through:
- Informal heuristic evaluation.
- Cognitive walkthroughs.
- Test scenarios as part of the iterative design process.
- Thinking aloud protocols.

2.4. Heuristic evaluation

Heuristic evaluation and cognitive walkthroughs were conducted repetitively by the same researchers on each prototype as they were developed. Researchers CBC and PS examined the interfaces and judged their compliance with recognised usability principles (the heuristics) [36]. Nielsen’s ten basic usability heuristics were used to evaluate the different interface designs and were also incorporated into discussions with users at the time of prototype testing to assess their opinion of the interface designs [37,38].

2.5. Cognitive walkthroughs

Researchers stepped methodically through the different prototype systems, as they were developing, identifying possible user problems, goals and actions [28]. Cognitive walkthrough is a process that is grounded in exploratory learning and is
based on a model of human cognition that describes human–computer interaction in terms of four steps. The user:
1. Sets a goal to be accomplished with the system (e.g. identifying cardiovascular risk).
2. Searches the interface for currently available controls (menu items, file-tags etc.).
3. Selects the action that seems most likely to make progress towards the goal.
4. Performs the selected action and evaluates the system’s feedback for evidence that progress is being made towards the intended goal [39].

Both tests were conducted as part of an iterative design process, where the results of testing were fed back into the design cycle in order to refine each prototype’s design before they were presented to users.

2.6. Task scenarios

Task scenarios can be used to facilitate product design both as a means of evaluation (e.g. when performing cognitive walkthroughs), and as a means of assessing the potential uses of the system [40].

Five task scenarios (Table 1), each constructed to reflect a clinical situation where the CDSS could be used, were used to assess the following aspects of the CDSS prototypes:
- Understanding and ability of users to interpret data presented.
- Issues related to screen designs.
- Issues related to navigation around the three different systems.
- Performance; i.e. the ease with which different tasks can be performed.
- Impressions of the tool’s overall usefulness in a clinical setting.

Table 1
Description of the five task scenarios used during the ‘formal usability tests’ on the clinical decision support system

<table>
<thead>
<tr>
<th>Task</th>
<th>Task description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Defining a person’s cardiovascular risk</td>
</tr>
<tr>
<td>2</td>
<td>Defining abnormal risk parameters; examining glycaemic control</td>
</tr>
<tr>
<td>3</td>
<td>Examining blood pressure results</td>
</tr>
<tr>
<td>4</td>
<td>Examining cholesterol results</td>
</tr>
<tr>
<td>5</td>
<td>Examining the impact of different risk factors on overall cardiovascular risk calculation</td>
</tr>
</tbody>
</table>
The number of successful responses to data interpretation questions and navigation events were recorded, as well as users’ screen design preferences. Users were asked short structured qualitative questions about navigation and performance issues, as well as their impressions of the tool’s overall usefulness. Problems with the CDSS were considered significant if more than one user experienced the same difficulty and if fewer than 90% of questions were answered correctly. Task scenarios were used both for the ‘cognitive walkthroughs’ conducted by developers and the ‘thinking aloud protocols’ conducted with users.

2.7. Thinking aloud protocol

A thinking aloud protocol was followed, whereby users were asked to vocalise their thought processes as they worked through the tasks, while being observed [38]. This has the dual benefit of allowing the assessor to judge the ease with which the system can be learned (‘learnability’), as well as providing first hand information about design problems [38]. Prompts were only given if the user was unable to proceed. Users were asked to record answers to specific data interpretation questions on a sheet provided and answer structured questions designed to give some indication of their opinion of the usefulness of the system and its ease of use. These questions formed the basis of a structured interview, so that answers given could be explored in more depth.

Users were presented with written questions relating to the five tasks they were expected to perform. Tasks were given in the same sequential order, becoming progressively more complex, with ‘users’ having to perform more ‘steps’ at each stage. The three different prototypes were randomly allocated to each task throughout all tests involving clinicians. In this way, only one prototype was tested with each task, but clinicians were exposed to all three during testing.

Patients were initially shown all three CDSS prototypes to assess their visualisation preferences, then one prototype was chosen with which to complete the five task scenarios. ‘Performance’ and ‘navigation’ were not assessed with patients since they were not expected to use the computer directly.

All tests were performed by the same two researchers, CC and EN. CC kept a record of errors made or difficulties encountered when users performed the computing steps required, as well as their screen visualisation preferences. EN kept notes of the users’ comments, any difficulties encountered, and their overall reaction to the system. At the end of each session, users were asked structured questions relating to heuristics of screen design and the tests were discussed. After each test, the two researchers conducted a debriefing with each other, in order to discuss any major problems which had occurred. The whole process lasted approximately 1 h.

3. Results

From the written information collected during all tests, including ‘cognitive walkthroughs’, statements relating to each user, task and prototype were collated. From these, specific task-related problems were identified, as well as problems with the system as a whole and alterations made to the ‘final version’.

3.1. Data interpretation

Both patients and clinicians readily identified absolute cardiovascular risk values from the thermometer. However, both groups had difficulty deciding at what level of cardiovascular risk treatment is required. Doctors tried to answer this question using their knowledge of the risk literature, whereas nurses experienced much more difficulty. Some patients used the colours on the thermometer as a guide to treatment, one saying that she would be happy “as long as it was always in the blue [low]”. Another said he would not treat risk until it “was in the high range”. Others could not answer this question. As a solution a clear ‘treatment line’ was incorporated onto the ‘risk thermometer’ in the ‘final’ version (Fig. 4).

Graphical data were easily interpreted, especially by clinicians, who answered all data interpretation questions correctly. Patients described trends in data easily, but experienced difficulty interpreting ‘target lines’ for specific biological
parameters. Those relating to glycaemic control (HbA1c) caused particular difficulty as patients were not sure whether the target value referred to people without diabetes or whether it was a value they were meant to strive for, in order to minimise their risk of diabetic complications.

Two patients experienced difficulty interpreting the graphs and tended to answer questions according to their common sense view of illness rather than by using the computer directly. For example, when asked whether the ‘dummy patient’s’ risk needed treating, a patient answered that it did, not because the graph showed the risk to be increasing with time, but because he believed the risk was “liable to rise” without treatment.

Patients answered 102 (91%) data interpretation questions correctly indicating that they understood the graphical representations of clinical data.

3.2. Visualisation preferences

Numerical data, relating to the colour and design preferences of both patients and clinicians, were collated. Preference was given to patients’ opinions as the designs were primarily aimed at them. Patients were very concerned about the clarity of information portrayed, recognising that many diabetic people are visually impaired. As a result they often chose plain backgrounds to graphs and liked the pie-graph representation of cholesterol, which they felt provided them with detailed information about their condition. This contrasted with the views of clinicians, who preferred the more interesting, coloured backgrounds to graphs and felt the pie chart provided patients with more information than they needed to know. Patients’ and clinicians’ visualisation preferences were incorporated into the ‘final version’ of the CDSS (Fig. 4).

3.3. Navigation

184 (90.6%) navigation events were successfully performed by clinicians. Problems were encountered when clinicians had to switch from one prototype to another during the usability tests. However, the system as a whole was felt to be easy to learn and most improved with increasing
exposure to the prototypes. Pockets of difficulty were encountered; for example several clinicians experienced difficulty navigating around the cholesterol screens, but these were simplified in the ‘final version’.

3.4. Reaction to the CDSS

General reactions towards the CDSS were positive. Clinicians, even those wary of IT, said they enjoyed doing the usability tests once they had learned to navigate the system. In particular, they liked the risk prediction facility of the CDSS, as they thought this would be useful for both patients and clinicians, to illustrate the impact of individual risk parameters on overall cardiovascular risk. Thus, they felt it could modify physicians’ prescribing behaviour by helping them target risk factors that had the greatest impact. Patients also thought this was a useful function of the CDSS, as it emphasised to them the importance of treatment.

Physicians viewed the ‘prompting system’ (highlighting of abnormal parameters) as useful, whereas nurses did not. Nurses explained this discrepancy as being due to their working ethic, which is to treat the ‘whole patient’ when they came to see them. They liked to show patients all their results, regardless of how ‘good’ or ‘bad’ they were. However, physicians thought the prompting system would help increase consultation efficiency by allowing them to target areas that need addressing with patients. As a result, nurses appeared more concerned with the potential for the CDSS to increase consultation time than doctors.

Neither clinicians nor patients thought the CDSS was intuitive. Both groups needed a period of learning before they became familiar with it. However, the general consensus was that this was easily achieved and that, once learned it was an easy system to use. This view is reflected in the results of the tests, during which most people improved with increasing exposure to the application.

Most patients (6/7) and all clinicians thought it was useful and had definite potential for use in clinic consultations. Some patients were enthusiastic about its educational value. In particular, they liked seeing individual results and trends as they felt this would help them understand what the doctor was saying to them, gain a deeper insight into their specific problems, and prompt them to ask more pertinent questions in clinic:

P: Because you’ve just it put in front of me in black and white and colours, and it’s just showing me a picture... so it’s straightforward isn’t it that? It’s so straightforward it jumps out really. That’s good.

P: That seems a sort of instant interpretation. The black line is in ‘high’ so you’ve got problems... This is just saying what the situation is now.

It’s a frightener... And this one is a more long term. I can pat myself on the back because I’ve done the right thing.

In contrast, one patient did not think the CDSS would confer any extra benefit to his consultation and preferred the more traditional approach of placing complete trust in doctors and doing what he was told by them. Despite the fact that he was very adept at interpreting the information presented to him, he doubted others’ ability to do the same: “They don’t want to see pretty pictures... [they would] find it difficult to understand”.

Throughout his test he made comments suggesting that it was the doctor’s rather than his responsibility to deal with his illness. However, this view was not echoed by others who actively wanted to know about their illness and the way it affected them, and most importantly what they could do to improve it.

3.5. Final design specification

As a result of the findings from the usability tests, a second design specification for the CDSS was drawn up. This incorporated the ‘users’ views, giving preference to patients, and sought to rectify highlighted deficiencies in the system. Unlike the preliminary specification, this was concerned with the ‘look’ and ‘feel’ of the product rather than setting the context of its use.
Some design changes included:
- Marking a clear treatment line for cardiovascular risk on the ‘risk thermometer’.
- Adding a ‘Print screen’ function, so that individual screens could be printed out for patients to take home.
- Changing the cholesterol pie charts so they displayed total cholesterol:HDL ratio rather than HDL:LDL:triglyceride ratios.
- Altering graphs so that they all have a white background.
- Placing the ‘patient data’ window opposite the graphs rather than having it on the same side as the graphs.

4. Discussion

Evaluation is a broad term encompassing activities such as the qualitative description of a system’s development process and quantitative assessments of its performance, as well as assessments of its impact on users, patients and the organisations within which they work [41]. A fundamental principle of evaluation is that it occurs throughout the development cycle of a product and involves users early in the design process [42,43]. However, any evaluation procedure should include: a user’s needs assessment, which is used to develop a functional and software specification; iterative usability testing throughout the design cycle; and an evaluation of the system’s performance and impact in controlled and ‘real’ environments [30,43]. This paper describes the process of formative evaluation, involving iterative usability testing, occurring during the design stage of a CDSS in a ‘controlled’ or ‘laboratory’ environment.

By conducting a number of different types of test at different stages in the design cycle, we fulfilled important principles of usability testing [44]. However, as well as being determined by the type of evaluation we wanted to do, our choice of tests was also influenced by resource and time constraints, and the fact that we had little access to usability engineering expertise. In his work, Nielsen recognised that these are common problems for many software designers, and as a result developed the idea of ‘discount usability engineering’ [45]. This process suggests using the same tests as we had empirically chosen, but stresses that they can be conducted and evaluated by non-usability engineering experts [45]. The advantage of these methods is that they rely on observation, and are thus relatively easy to perform [46].

Several studies have compared the efficacy of using different usability methods and expert versus non-expert evaluations, in identifying usability problems [45,47,48]. These studies have shown that approximately one third of serious usability problems are identified by all methods regardless of evaluator expertise, and that different tests are complimentary to each other [47]. This suggests that relatively well-planned usability testing, even when performed by a non-expert, is likely to result in the discovery of major usability flaws.

This point is re-iterated by Nielsen, who in relation to his ‘discount usability process’ said: “that even fairly methodologically primitive experiments will succeed in finding many usability problems” [45]. None of the researchers involved in the evaluation of this product were usability experts, which limited the scope of our evaluation. Nonetheless, many usability problems were identified in the prototype versions of the CDSS.

In conducting formal usability tests with ‘end-users’ we collected a combination of numerical and qualitative data. Numerical data were used to help determine the system’s functionality i.e. how easily could users navigate the various prototypes? How many mistakes did they make? How easily interpretable were the clinical data presented? However, this did not provide us with any insight into issues relating to the learnability of the system, or how acceptable or useful it was judged to be. The thinking aloud protocol was particularly useful in this respect, as qualitative data regarding
these issues were collected concurrently with task performance. In some ways this information could be suspect, as users are under a reasonable amount of pressure to perform and are well aware of the investment designers have put into the system [22,27]. As well as this, the principle researcher was well known to most users, a fact which is likely to have influenced their responses. Nonetheless many valuable insights were gained.

As we had three prototypes to test and only limited time in which to test them, we conducted the tests using a ‘within-subject design’ where all users are exposed to all prototypes [27]. This process has the advantage of controlling for individual variability in test performance, as someone who is used to computers will be equally superior in all tests [27]. This was particularly relevant for our study, where there was a reasonable amount of disparity in computing abilities amongst clinicians. However, unfortunately it also means that users cannot be considered as novice users once they have been exposed to the first prototype, as some transfer of skill will take place between systems [27]. We were interested in the ease with which the system as a whole could be learned. The test prototypes were sufficiently similar in design to facilitate this, as most users, including patients, improved in their performance with increasing exposure to the three different prototypes.

Although we were able to test specifically some usability characteristics of the CDSS, such as learnability, errors and functionality, we did not specifically test memorability or its efficiency. Memorability refers to the ease with which a user remembers how to use the system after a period of disuse. It is particularly relevant to users who are likely to use the system on an intermittent basis, as are clinicians and patients [22]. Ideally we should have retested users some time after they had performed the formal usability tests, but this was difficult given our time constraints and clinicians’ commitments [22]. However, given that improvements in learnability often make interfaces easy to remember, it is unlikely that this would be a significant problem for our user population [22].

Efficiency testing requires the system to be used by ‘expert’ users once a steady state of performance has been reached. None of our users fulfilled this category. Also most efficiency tests require users to perform timed tasks [22]. We could not perform these as we used a ‘thinking aloud protocol’. As well as this, the CDSS was designed to be used during 15–20 min consultations with patients. Hence, the rate or speed at which it will be used will be determined by patients’ abilities to understand what is being shown to them, rather than the interface design itself. Therefore, efficiency in its classical sense becomes relatively meaningless for this application.

It is the nature of a clinical consultation that clinicians and patients share information with each other. Up to the time of this study, the majority of patient-specific information was presented to clinicians in Salford diabetes clinics as numerical tables, and explained in a qualitative manner to patients, without them ever directly seeing their results. The primary aim of the CDSS was to display patient-specific information and cardiovascular risk to patients in a format that they could understand. The CDSS was envisaged to be a specific tool, used by clinicians in diabetes clinics to facilitate patients’ understanding of their condition. Hence, the CDSS interface had to satisfy different user groups, such as doctors, nurses and patients. In order to achieve this, we decided the CDSS had to display certain pre-defined usability criteria. These were that:

- All information should be clearly displayed.
- All information processing should be ‘invisible’ to the user.
- There should be no need for clinicians to input data themselves, when using the system.
- The system should be easy to learn and navigate around.
- All clinical data should be represented clearly in a format familiar to clinicians and easily understood by patients.

By including these criteria within the design of the CDSS, we showed it could be easily learned by different groups of professionals and hoped it would fit into the natural flow of the consultation. Difficulties were encountered in designing the interface displays as they were essentially designed by clinicians, for use with patients. However, we felt it was unrealistic to expect patients to understand all the concepts and functions the CDSS
was expected to perform. Therefore they were not included in the preliminary design phase, but were provided with concrete examples of design options, which elicited detailed responses. We relied on these to derive the final specification of the CDSS, where patients’ design preferences were given priority. However, clinicians’ needs were also recognised and compromises made with the design, such as the inclusion of numerical tables as well as graphs.

Despite our lack of usability engineering expertise, the data we collected were rich both in identifying usability problems and the reasons why these occurred. As a result, we were able to derive a final design and usability specification of the CDSS, which was used to design the final version. Although the CDSS was positively received by most, the usability tests showed that different professional groups are likely to use different aspects of the system, and that it may not be suitable for all patients. Some patients may not be able to understand it and others may not want to know about their illness in such detail. However, the fact that many patients were enthusiastic about it shows that it has definite potential to help further their understanding, and provide them with the necessary impetus to adhere to the behavioural changes necessary to reduce their individual cardiovascular risk.

We were limited by resource constraints and therefore had to choose methods that were easy to perform and evaluate. Our situation is likely to be a realistic reflection of the resources available to many software developers within the NHS. As such the methodology we employed, based on Nielsen’s ‘discount usability method’, has been validated by our ‘real life’ study, as they were successful in highlighting major usability problems. However, the system we designed was relatively simple both in its functions and appearance. The methodology we employed was time consuming, because of the qualitative nature of the work, and used small numbers of participants. Larger more complex systems would require larger more frequent tests and as such benefit from tests requiring the collection of large sets of numerical data and statistical analysis, as well as more formal input of usability expertise.

Finally, this study only describes iterative usability testing of product prototypes in a ‘controlled’ environment. Further studies need to be conducted to evaluate the ‘final version’ of the product in real consultations with patients, in order to evaluate its impact on the consultation process and the cardiovascular risk beliefs and knowledge of this population of diabetic patients.

References


