The Role of Usability Engineering in the Development of an Intelligent Decision Support System

Clare Martin[1], Arantza Aldea[1], David Duce[1], Rachel Harrison[1], Marion Waite[2]

[1] Department of Computing and Communication Technologies,
Oxford Brookes University, UK.
[2] Department of Applied Health and Professional Development,
Faculty of Health and Life Sciences, Oxford Brookes University, UK.

Abstract. We describe the role of human factors in the development of a personalised clinical decision support system for type 1 diabetes self-management. The tool uses artificial intelligence (AI) techniques to provide insulin bolus dose advice and carbohydrate recommendations that adapt to the individual.

Keywords: Artificial intelligence, usability, safety, human factors.

1 Introduction

Most people with type 1 diabetes (T1D) have to perform multivariate dose calculations several times a day in a variety of contexts that might affect cognitive load. Many use mobile decision support tools to assist with the process, but these typically use simple formulae based on a limited set of parameters. This paper describes how human factors are being considered in the design of a more sophisticated system: PEPPER [4]. Its design involves users at all stages to ensure that it meets their needs.

Poor usability has long been identified as one of the barriers contributing to the lack of adoption of intelligent personal guidance systems for diabetes management [1]. There are related issues surrounding trust in applied artificial intelligence (AI). For example, concerns have been raised regarding loss of control [2] and there is scepticism towards systems that replace human decision-making [3].

The PEPPER (Patient Empowerment through Predictive PERsonalised decision support) system [4] uses artificial intelligence to offer insulin dose advice that is tailored to the individual. Most of the input data is transmitted to a mobile handset wirelessly from wearable devices. Additional information can be added manually, but only one such field is mandatory (carbohydrate), as people can find such interactions tedious [5]. The information gathered by the handset is processed by a Case-Based Reasoning (CBR) module to determine a personalised insulin recommendation that adapts over time. A second Model-Based Reasoning (MBR) module is used to maximise safety. Its computer model generates predictive glucose alarms, automatic insulin suspension, carbohydrate recommendations and fault diagnosis. The development of PEPPER uses an iterative methodology, integrated with clinical validation and formative usability evaluation. Methods for the latter are described in the next section.

1.1 User-centred Design

The usability engineering process for medical software is more rigorous than that for other domains because of the need to consider safety and hazards. This requirement is encapsulated by standards such as International Electrotechnical Commission (IEC) 62366 [6], which is recognised by, and similar to, the guidance offered by the U.S. Food and Drug Administration (FDA). Both protocols emphasise the importance of conducting a user-centred design to determine tasks and frequently used functions, as well as identifying risks and use-related errors, prioritised by severity of harm.

One of the shortcomings of the IEC standard is that it offers very little advice about how to evaluate technology in *context*, a crucial consideration for mobile devices. One way to gain an understanding of people's experiences and the real scenarios of use is to employ 'situated methods'. Such techniques have proved effective in a recent study of users of diabetes technology in diverse situations [7]. Factors such as the physical environment, as well as cultural or social context and lifestyle can have an enormous effect on how people interact with their personal equipment. There is an important trade-off between quality of life and health outcomes for instance. Similar methods will therefore be applied here.

2 Methods

The usability engineering process used for PEPPER follows the IEC 62366 standard [6]. As such, it comprises nine steps, which are grouped into four sections in Table 1. Formal definitions of the phrases in uppercase are given in the standard and therefore elided here. This structure follows an existing example [8], but our interpretation differs in the mixed methods it employs. Ethical approval has been obtained from the relevant authorities for all elements involving users.

Table 1. PEPPER Usability Engineering Process

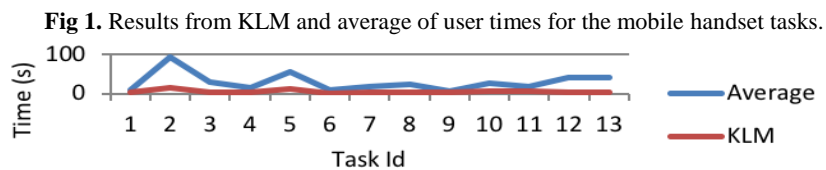|	
USER Research	1. Prepare USE SPECIFICATION: focus groups, interviews, surveys
Analysis	2. Identify characteristics for SAFETY: TASK analysis, document review
	3. Identify HAZARDS and HAZARDOUS SITUATIONS
	4. Identify and describe HAZARD-RELATED USE SCENARIOS
	5. Select USE SCENARIOS for SUMMATIVE EVALUATION
Iterative Design and FORMATIVE EVALUATION	7. Establish USER INTERFACE (UI) EVALUATION plan
	6. Establish USER INTERFACE SPECIFICATION
	8. Perform UI design, implementation & FORMATIVE EVALUATION
	Phases: 1. Analytical study 2. Empirical lab study 3. Contextual study
SUMMATIVE	9. Perform SUMMATIVE EVALUATION: repeat Phase 2

3 Results

We present limited preliminary results here, for reasons of space, focusing on Step 8. The methods were applied to both the mobile handset and web server.

Phase 1: Analytical Study. Two procedures were chosen for this phase: heuristic evaluation and the keystroke-level model (KLM) [9]. The evaluators were dual-domain experts: computer scientists familiar with both standard human interface evaluation methods and T1D. Three evaluators conducted the heuristic evaluation independently, with 11 heuristics, scored on Nielsen's SRS scale. The results were analysed and presented to the evaluators at a debriefing session to develop recommendations for redevelopment of the prototype prior to user testing. For the KLM, a single expert conducted a series of tasks (see Table 2) to provide a baseline of ideal timings for comparison with the results of Phase 2, see Fig. 1 for the results of the handset evaluation.

Table 2. Sample Tasks from the PEPPER Mobile Device Task Model

2. Calibrate the continuous glucose meter using the Bluetooth capillary meter.
3. Locate and state the most recent bolus in the last 12 hours.
5. Request bolus advice, for the following situation: 45g carb, Medium absorption…etc.

Phase 2: Empirical Laboratory Study. The purpose of this study was to measure the performance of the system with regard to the usability goals of simplicity, effectiveness, efficiency, and satisfaction. The systems were tested in one-to-one sessions, each lasting two hours. During each session the investigators gave participants the same series of tasks to perform as in the KLM evaluation. Their behaviour was audio recorded, and the interaction of their hands on the device was video recorded. Fifteen patients were enrolled in the handset study: 7 in Spain and 8 in the UK. Four clinicians participated in the server study. The SUS questionnaire was used to determine satisfaction [10]. The SUS scores were excellent for the handset (87%) but showed that the web server needed redesign (66%). Video data analysis showed there were few errors and most tasks were completed, showing high simplicity and effectiveness respectively. Inefficient tasks were identified from the average times, see Fig. 1. Think-aloud comments contributed to the recommendations for redesign.

Fig 1. Results from KLM and average of user times for the mobile handset tasks.

Phase 3: Empirical Contextual Study. The PEPPER situated study was closely based on an existing method [7]. It lasted 4 weeks, involved 15 participants and included auto-ethnography, an initial interview, a diary study and a contextual group meeting. The diary study formed the heart of this usability study. Its purpose was used to understand the day-to-day user experience with PEPPER over a period of several weeks. The goal was to see how context affects the use of the technology and also to understand which features affect motivation, either positively or negatively. Participants were asked to make diary entries each time they used the PEPPER bolus advisor and they were also phoned at weekly intervals. The final step is an observational study of

the group in a social setting at an informal location such as a cafe. Its purpose is to validate findings from the prior steps and to observe discourse about the experience.

This part of our usability study is not yet fully completed. Early findings show that the participants like the system and trust the recommendations and alarms. However, it makes them constantly aware of their condition, and concerned about glucose targets. They also thought that there were too many parameters required for the CBR model. These preliminary results have important implications for developers of AI self-management systems for diabetes and other conditions.

4   Conclusion

In this paper we have given a very brief overview of how one project is conducting a process that adheres to international standards for consideration of human factors in the design of medical software. We have also proposed a method in which a situated study can be incorporated into the standards, to fill a perceived gap around evaluation of systems in context. An additional aim of this 'in the wild' study is to answer important research questions about adherence, health beliefs and trust in the AI.

References

1. Bellazzi R. Telemedicine and diabetes management: current challenges and future research directions. Journal of diabetes science and technology. 2008;2(1):98-104.
2. Fast, E., & Horvitz, E. 2017. Long-Term Trends in the Public Perception of Artificial Intelligence. In Proc.of the Thirty-First AAAI Conference on Artificial Intelligence 963-969.
3. Hengstler M, Enkel E, Duelli S. Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. Technological Forecasting and Social Change. 2016;105:105-20.
4. Herrero P, López B, Martin C. PEPPER: Patient Empowerment Through Predictive Personalised Decision Support. Proceedings of the 1st Workshop on Artificial Intelligence for Diabetes, The Hague. 2016.8-9.
5. Årsand, Eirik, et al. Mobile health applications to assist patients with diabetes: lessons learned and design implications. Journal of diabetes science and technology. 2012; 6(5): 1197-1206.
6. International Electrotechnical Commission. IEC 62366-1:2015: Medical devices -- Part 1: Application of usability engineering to medical devices. Geneva, Switzerland: International Organization for Standardization. 2015.
7. O'Kane AA. Individual differences and contextual factors influence the experience and practice of self-care with type 1 diabetes technologies (Doctoral dissertation, UCL). 2016.
8. International Electrotechnical Commission. IEC/TR 62366-2:2016: Medical devices -- Part 2: Guidance on the application of usability engineering to medical devices. 2016.
9. Nielsen J. Usability Engineering Morgan Kaufmann San Francisco. CA, USA. 1994.
10. Brooke, John. SUS-A quick and dirty usability scale. Usability evaluation in industry 189.194 (1996): 4-7.