RADAR

Research Archive and Digital Asset Repository

OXFORD
BROOKES
UNIVERSITY

WWW.BROOKES.AC.UK/GO/RADAR

Systematic analysis of mobile diabetes management applications on different platforms

Eva Garcia[1], Clare Martin[2], Antonio Garcia[1], Rachel Harrison[2], Derek Flood[2]

[1]University of Alcalá, Spain
[2]Oxford Brookes University, UK
eva.garcial@uah.es, cemartin@brookes.ac.uk, a.garciac@uah.es,
rachel.harrison@brookes.ac.uk, derek.flood@brookes.ac.uk

Abstract. There are a number of mobile applications available to help patients suffering from Type 1 diabetes to manage their condition, but the quality of these applications varies greatly. This paper details the findings from a systematic analysis of these applications on three mobile platforms (Android, iOS, and Blackberry) that was conducted to establish the state of the art in mobile applications for diabetes management. The findings from this analysis will help to inform the future development of more effective mobile applications to help patients suffering from Type 1 diabetes who wish to manage their condition with a mobile application.

Keywords: usability, keystroke level modelling, mobile, heuristic evaluation.

1 Introduction

The development of usability evaluation methods for mobile devices and their applications is a growing area of research which has been fuelled by the rapid growth in the use of smart phones in recent years. Healthcare services in particular stand to benefit from the huge potential offered by the combined technologies of smart phones and cloud computing [1], particularly those associated with chronic conditions such as diabetes that require a high degree of self-management [2], but such interventions are likely to fail unless sufficient attention is paid to usability [3, 4]. Standard evaluation techniques [5] can be broadly applicable to mobile phone applications, but they often need to be adapted to consider additional factors such as context, connectivity and security [6].

This paper contains the results of a systematic evaluation of mobile applications for diabetes management. The study was conducted partly to discover the usability issues of existing applications in order to avoid repeating them in the design of a new application for diabetes management, and also to elicit functional requirements for this new application. In total, over 400 apps were examined in order to give a broad overview of the current status of smart phone based diabetes applications from a developer's perspective. This evaluation was conducted by usability experts, and consisted of a methical survey based on efficiency, heuristics and functionality. A subsequent evaluation is planned to investigate the potential problems that are

experienced by patients while using some of these applications and to determine which interface components are most suitable for data input and output.

The methodology used here was devised in order to sift through the vast array of mobile applications that already exist for diabetes management, but it relies on general techniques that are equally applicable to other domains. This methodology has also been applied to spreadsheet packages [7]. The methodology comprises a series of steps that can be used to filter through a collection of applications by comparing functionality, efficiency and various other attributes that have been identified for use as mobile heuristics [6] such as personalization, ergonomics, flexibility, security and error management. In addition, the process was used to elicit functional requirements by collating a list of features offered by existing applications that will subsequently be ranked by potential users.

This paper is organized as follows. Section 2 introduces type 1 diabetes, Section 3 describes related work that has been conducted in this area. Section 4 explains the general steps of the protocol used to perform the analysis, and Section 5 presents the results obtained at each step of the protocol. Section 6 presents a discussion of these results and Section 7 discusses the threats to validity. The paper is then concluded in Section 8 with a summary of the work completed thus far and the work that will be performed in the future.

2   Type 1 Diabetes

Type 1 diabetes occurs when the insulin producing cells of the pancreas are destroyed leaving the body unable to control its blood glucose levels. People with this condition have to take insulin regularly to try to keep their glucose levels within a safe target range, and failure to do so can lead to short term complications such as heart palpitations or dizziness, and also longer term complications such as retinopathy and peripheral neuropathy.

Most medical professionals encourage patients to keep a diary containing their blood glucose levels before each meal, together with the corresponding carbohydrate intake and insulin dose, in order to help them stay within the target range. The insulin dose is usually calculated by the patient, according to the number of carbohydrates consumed and various other factors including exercise and hormones, but an increasing number of glucose monitors now offer a degree of decision support with such calculations. In addition, a vast number of mobile phone applications have been developed that help patients collect and maintain this information. These applications can now be distributed easily and quickly through locations such as the App store from Apple and the Android marketplace, making it easier for patients to experiment with them. Their usability varies greatly however, and it has been shown [8] that usability can be a determining factor in a users' choice of mobile application.

3 Related Work

Previous studies have been conducted to evaluate the effectiveness of mobile phones for assisting with diabetes self-management [2, 9], and some standards are emerging with regard to how to design mobile applications in general [10], as well as mobile medical applications in particular [11], but research about the usability evaluation of applications for mobile devices is still a relatively new area [12]. A comprehensive survey of the status and trends of 200 mobile-health applications was conducted in [13] but it differs from the work presented here in a number of ways: it considered a broader range of healthcare applications, it was based on the user perspective rather than that of experts and it was restricted solely to the iOS platform.

Some of the specific issues that affect the usability of mobile applications include the following [14]:

- *Mobile context*: this includes the location, identities of nearby people, objects and environmental elements that may distract the user's attention. In systems for diabetes management, the context is important as the need to measure blood glucose can arise at any time, in any place.
- *Connectivity*: network conditions (data transfer speed, reliability, strength of signals) may vary at different times and locations. Connectivity is significant for diabetes management systems, as the information collected by these apps may need to be synchronized with on-line repositories such as Microsoft© HealthVault.
- *Small screen size*: the small screen size can significantly affect the usability of mobile applications. More specifically in systems for diabetes management, screen size affect the ability to visualize the stored data and graphics, which could be particularly problematic for patients with retinopathy, a common complication of diabetes.
- *Different display resolution*: different levels of display resolution on different mobile devices may cause different usability test results. As in the previous case, display resolution affects the ability to visualize data in graphics, and this could be a serious problem for patients with retinopathy.
- *Limited processing capability and power*: computational power and memory capacity of mobile devices are much less than those of desktop computers. Because of this, developers may have to disable some functions (e.g., high-resolution images and dynamic frame movement). In systems for diabetes management, the limited processing capability could affect the choice of dosing algorithm.
- *Data entry methods*: small buttons and labels limit users' effectiveness and efficiency in entering data, which may reduce the input speed and increase errors. The data entry method is crucial in diabetes management applications, since people must log data frequently, and they could be hindered by peripheral neuropathy, a common complication of diabetes.

4 The Evaluation Method

The methodology used here was devised as a way of filtering the large number of mobile applications that already exist for diabetes management to obtain a set that was small enough to evaluate in greater depth using relatively traditional techniques. The protocol used has been applied to other studies of mobile applications [7] and consists of the following steps:

1. *Identify all potentially relevant applications.* This step consists of searching the applications related to a particular keyword. Current online stores facilitate this task, such as the App Store from Apple, the Android Market from Google and the App World from BlackBerry.
2. *Remove light or old versions of each application.* The trial versions (those that offer a subset of the functionality offered by the corresponding full application or access to the full application for a limited period of time) should be removed as the corresponding full version will also be evaluated.
3. *Identify the main functional requirements and exclude all applications that do not offer this functionality.* Only the applications that meet all main requirements are carried forward to subsequent steps of the protocol.
4. *Identify all secondary requirements.* This step consists of identifying the secondary requirements that each application offers, i.e., the additional functionality that is not required in the step 3.
5. *Construct tasks to test the main functional requirements using each of the methods below:*
   a. *Keystroke level modelling (KLM)* [15] is used to provide a measure of efficiency for each application. New interaction methods provided by mobile devices have not been incorporated into KLM. Therefore it was not possible to predict the efficiency in terms of time, however it was possible to use the number of keystrokes as an indication of the efficiency of these applications.
   b. *Heuristic evaluation* [16] is used to identify some usability problems. Nielsen recommends performing the heuristic evaluation with between three and five evaluators. The heuristics used during this evaluation, shown in Table I, were developed specifically for mobile applications in [6].

Table I: Mobile usability heuristics

Heuristic	Description
A	Visibility of system status and losability/findability of the device
B	Match between system and the real world
C	Consistency and mapping
D	Good ergonomics and minimalist design
E	Ease of input, screen readability and glancability
F	Flexibility, efficiency of use and personalization
G	Aesthetic, privacy and social conventions
H	Realistic error management

5 The Results

The following section details the results obtained at each step of the protocol.

Step 1: Identify all potentially relevant applications.

To identify all potentially relevant applications a search of the online store of each platform was conducted using the keyword *"Diabetes"*. The results of the search on each platform are presented in Table II. It can be seen that there are over 150 applications returned on both the iOS and Android platforms, but only 28 apps were found in the Blackberry App World.

Table II: Number of apps returned by platform

iOS	Android	Blackberry
231	168	28

Step 2: Remove light or old versions of each application.

The next step of the protocol is to remove any light or old versions of applications already included in the results. As these applications are already represented, to evaluate both would require additional, unnecessary resources. The number of applications that were removed on each platform is presented in Table III. It can be seen that only a small number, (less than 10%) of applications were removed, indicating that most applications are unique.

Table III: Number of apps remaining after removing old and light versions

	iOS	Android	Blackberry
Number of Applications	231	168	28
Old or light versions	9	6	1
Apps remaining	222	162	27

Step 3: Identify the main functional requirements and exclude all applications that do not offer this functionality.

Recent studies have shown that users of healthcare apps prefer those that facilitate everyday tasks such as tracking blood glucose to those which are designed as reference tools [13]. The primary functionality was therefore chosen to model a daily diabetes management diary. The diary is only useful if the results can be visualized and transmitted to a backup computer however, and so tasks were developed to test these features as well. The functionality that was considered to be the minimum for any such application was mapped to the tasks shown in Table IV:

Table IV: Tasks to evaluate main functional requirements.

Task	Functional requirements
1	Set measurement units
2	Log blood glucose level
3	Log carbohydrate intake
4	Log insulin dose
5	Display data graphically
6	Export data via email or similar

The reasons for choosing this functionality are outlined below:

- *Set measurement units*: some applications only permit units to be set to those used in a particular country, such as Mmol/L in Europe or Mg/dl in the United States. To be used internationally the app must allow the measurement units to be changed according to the users' preference.
- *Log blood glucose, insulin and carbohydrate*: Maintaining the blood glucose, insulin and carbohydrate levels over time can help patients with diabetes to better predict the onset of symptoms and take corrective action when necessary.
- *Display data graphically:* data in graphics can be easily understood and shows trends and patterns that can be used to indicate when patients are required to adjust the amount of insulin they are required to inject.
- *Export data via email or similar:* mobile devices have a very limited storing capacity, so there needs to be a facility to export data to an external source.

The descriptions and the developers' websites for each application were examined to exclude those that clearly did not offer the primary functionality shown in Table IV. After this, the remaining applications were downloaded and tested in order to determine if all six tasks could be completed and the results are summarized in Table V. It can be seen that only a very small number of applications on each platform provided all of the necessary functionality.

Table V: Number of apps on which all tasks can be performed

	iOS	Android	Blackberry
Number of apps	8	6	1

The applications that were excluded at this step were varied and included general medical information references to cookbooks offering recipes suitable for patients suffering from diabetes. Some of the applications did offer some of the desired functionality, but these applications were removed since the completion of all of the tasks was deemed to be essential for the effective management of diabetes.

Step 4: Identify all secondary functional requirements.

The purpose of this step was to elicit potential functional requirements that go beyond the minimum level identified in Step 3, which could subsequently be ranked by

patients to obtain a measure of their usefulness. The requirements common to all the three platforms were as follows:

- *Log physical activity and other medication*: these factors, among others, can affect the insulin dose and are therefore logged by some applications.
- *Allow personal settings*: nearly all of the applications include a number of personal settings, such as the target blood glucose level. Some applications which featured a dosage calculator, for calculating the correct dosage, also allowed additional personal settings required for this calculation.

A number of additional features were found on only one or two platforms:

- *Decision support*: some applications offer basic support in calculating the insulin dose, but none offers any intelligent decision support.
- *Alcohol intake, illness, weight, blood pressure, allergies and hypoglycemia*: these factors, as well as physical activity and other medication, can affect the insulin dose and are therefore considered by some applications.
- *Carbohydrate database*: some applications include a database of nutritional information of certain types of food to facilitate calculations.
- *Discussion forums*: anonymous discussion forums can be a source of support for people with diabetes.
- *Export to online healthcare systems*: there is a growing demand for applications that store information securely in a way that can be shared with other selected users [17], including clinicians.
- *Ability to set reminders*: a number of applications have this facility, which is useful for long-acting insulin.
- *Backup and restore:* some applications have automatic backup and restore facilities over a wireless network. These facilities allow data to be recovered if the app is lost or damaged.
- *Add notes:* some applications allow users to enter a note with the logged data to record additional information about the data entered.
- *Insulin dose calculator*: some applications include a calculator for recommending insulin dosages.
- *Log pharmacies*: some applications have an option to log the information (phone, address, etc.) of pharmacies.
- *Log lab results*: there is an option to log the results of the laboratory analysis like sugar, A1C, HDL, LDL, cholesterol, triglycerides and creatinine.

A summary of the fulfillment of these requirements by applications on the three platforms is shown in Table VI.

Table VI: Additional requirements met by applications on each platform

Requirement	iPhone	Android	BlackBerry
Ability to set reminders	X	X	
Add notes		X	
Allow personal settings	X	X	X
Backup and restore		X	
Carbohydrate database	X		
Decision support	X		
Discussion forums	X		
Export to online health system	X		
Insulin dose calculator		X	
Log alcohol intake	X		
Log allergies			X
Log blood pressure		X	
Log hypoglycaemia	X		
Log illness	X		
Log lab results			X
Log other medication	X	X	X
Log pharmacies			X
Log physical activity	X	X	X
Log weight		X	

Step 5a: Perform keystroke level modelling analysis.

Each of the 15 apps that resulted from Step 3 were now subjected to a keystroke level modelling analysis, using the tasks listed in Table IV. The KLM was performed by counting the number of interactions required to complete each task.

Results for iOS

The results of the KLM analysis performed on the final 8 applications on the iOS platform are shown in Table VII.

Table VII: Results of KLM for iOS applications

Application Task:	1	2	3	4	5	6	Total
RapidCalc Insulin Dose Manager	10	2	1	3	2	1	19
GluCoMo	5	5	4	4	1	3	22
Diabetes Diary	5	3	3	2	3	6	22
Diabetes Personal Manager	3	7	4	3	3	6	26
DiabetesPlus	2	7	6	6	1	6	28
LogFrog DB	6	5	6	5	1	7	30
Diabetes Buddy Control your Blood Sugar	3	9	10	7	1	8	38
Diabetes Pilot	5	5	7	3	3	27	50

In Table VII it can be seen that task 6 (export via email or similar) is far less efficient on the Diabetes Pilot app than any of the others. This is because it requires the user to type the entire address which can be difficult on a small keyboard rather than using default settings or the contact list.

Also task 1 (set measurement units) is less efficient on the RapidCalc app than the other applications because it is necessary to exit the application and enter the iPhone settings to carry out this task, but it is arguable that efficiency is less crucial here since the task is usually only carried out once.

The data logging occurs so frequently that it is essential for it to be as efficient as possible. The tasks that are concerned with data logging are 2 (log blood glucose), 3 (log carbohydrate) and 4 (log insulin), and the figures in Table 4 suggest the RapidCalc app and the Diabetes Diary app use the most efficient methods for this. Unfortunately this does not tell us anything conclusive about data entry methods, since RapidCalc uses sliders, whereas Diabetes Diary uses either a picker or a keyboard. The input methods used for each application are shown in Table VIII.

Table VIII: Data entry methods for iPhone applications

Application	Data Entry Method
RapidCalc Insulin Dose Manager	Sliders on same screen
GluCoMo	Pickers on separate screens
Diabetes Diary	Pickers or keyboard on same screen
Diabetes Personal Manager	Keyboard on same screen
DiabetesPlus	Pickers on separate screens
LogFrog DB	Dials on separate screens
Diabetes Buddy	Keyboard on separate screens
Diabetes Pilot	Keyboard on separate screens

Results for Android

The results of the KLM analysis performed on the final 6 applications on the Android platform are shown in Table IX. It can be seen that tasks 3 (log carbohydrate) and 4 (log insulin dose) are less efficient on the GlucoJournal app than on any of the others.

Focusing on data logging (tasks 2, 3 and 4), the most efficient application is Glucometro, and the least efficient is GlucoJournal. These data are consistent with the total scores, because the GlucoJournal and DiabetesBox apps score the worst in the KLM analysis, as they need 38 touches of the screen to fulfill all tasks, and the Glucometro app is the best, as it only needs only 29 interactions.

Table IX: Results of the KLM analysis for Android applications

Application Task:	1	2	3	4	5	6	Total
OnTrack Diabetes	4	7	8	9	2	5	35
Glucool	3	9	8	9	2	3	34
DiabetesBox	5	9	8	10	3	3	38
Glucometro	4	7	7	6	1	4	29
Track3 Diabetes Planner	5	9	9	6	2	4	35
GlucoJournal	3	5	10	13	2	5	38

The data entry methods of these applications shown in Table X do not really explain the differences in efficiency since all of the applications used a keyboard, and in general each number had to be input on a separate screen.

Table X: Data entry methods for Android applications

Application	Data Entry Method
OnTrack Diabetes	Keyboard (on same or separate screen)
Glucool	Keyboard or "+" and "-" buttons (on separate screens)
DiabetesBox	Keyboard on separate screens
Glucometro	Keyboard on separate screens
Track3 Diabetes Planner	Keyboard on separate screens
GlucoJournal	Keyboard (on same or separate screen)

Results for Blackberry

The results of the KLM for the BlackBerry application are shown in Table XI.

Table XI: Results of KLM for the BlackBerry application

Application Task:	1	2	3	4	5	6	Total
iRecordit Diabetes Sugar Glucose and Health Tracker	20	57	16	13	2	11	119

As there is only one relevant application on the BlackBerry platform there is no comparison we can perform, but it is clear that task 5 is the easiest to accomplish (only 2 interactions) and task 2 is the most difficult to perform (57 interactions).

The data entry methods of this application is a picker for glucose and keyboard for carbohydrate and insulin, all on the same screen, as it is shown in Table XII. The picker for glucose is very small, and it contains only default values from which the user has to select one. The carbohydrates and the insulin can be entered through the keyboard.

Table XII: Data entry methods for the BlackBerry application

Application	Data Entry Method
iRecordit Diabetes Sugar Glucose and Health Tracker	Little picker (for glucose) and keyboard (for carbohydrate and insulin) on same screen

Summary of KLM Results

The process of conducting the KLM led to the following observations, all of which can affect the usability of mobile applications:
- Neither of the standard numeric keyboards offered by the iOS platform is ideal for the efficient input of decimal point numbers, since one is intended purely for integers and the other includes unnecessary symbols such as currency delimiters
- Pickers and sliders need to be designed carefully, since they are very difficult to use if they are too fine grained.
- If several numbers need to be input together then the screens should be designed so that they can all be input from the same screen.
- The haptic feedback offered by the keyboard on the Blackberry does not appear to offer increased efficiency but it may increase satisfaction.

Step 5b: Perform heuristic evaluation.

The applications that scored the best in the KLM analysis were then the subject of heuristic evaluation by three to five expert evaluators. For the purposes of this study, each mobile heuristic, defined in Table I, was broken down into sub-divisions to be assessed by each evaluator using Nielsen's severity Ranking Scale as detailed in Table XIII. Some of these sub-divisions were specific to the domain of diabetes management and are given in Table XIV.

Table XIII: *Nielsen's Severity Ranking Scale (SRS)*

Rating	Description
0	I don't agree that this is a usability problem at all
1	Cosmetic problem only. Need not be fixed unless extra time is available on project
2	Minor usability problem. Fixing this should be given low priority
3	Major usability problem. Important to fix, so should be given high Priority
4	Usability catastrophe. Imperative to fix this before product can be released

Table XIV: Heuristic sub-divisions.

A	A1 The battery status is visible
	A2 The network status is visible when transmitting data
	A3 The time is visible when entering data
	A4 The previously logged data and personal settings can be recovered if the device is lost
B	B1 The information appears in a natural and logical order
	B2 The information is presented clearly
	B3 You can see where everything is that you might need
C	C1 It is easy to see how to do tasks like entering blood glucose and carbohydrates
	C2 There are no objects on the interface that you would not expect to see
D	D1 The screens are well-designed and clear
	D2 The dialogues do not contain information that is irrelevant or rarely used
E	E1 It is easy to input the numbers
	E2 It is easy to see what the information on each screen means
	E3 You can easily navigate around the app
	E4 The screens have a 'back' button
	E5 The user can get crucial information 'at a glance'
F	F1 The user can personalise the system sufficiently
	F2 The system allows efficient input of data
G	G1 The design looks good
	G2 There are suitable provisions for security and privacy (eg Transmission of data is encrypted.)
H	H1 Users can recover from errors easily (If something goes wrong you can get back to where you were easily.)
	H2 If data is input incorrectly, it can be edited
	H3 There an Undo button, where appropriate

Each evaluator ranked each sub-division and provided evidence and comments for each ranking. This evaluation highlighted some problems that had not been identified by the KLM. For example, there are some applications that include unnecessary options or irrelevant information, which may confuse the user. Security was also identified as a problem of the evaluated applications, because none of these encrypt the data transfer, which would be appropriate for sending personal healthcare information over the Internet. Some applications show errors that are unrecoverable and cause the application to shutdown, which is disconcerting and annoying to the user. It is beyond the scope of this paper to give a full qualitative summary of the evaluation, which is presented in [18], but the summaries of the numerical rankings are included here.

Results for iOS

Five evaluators ranked the iOS applications, and a summary of the results is shown in Table XV.

Table XV: Average heuristic ranking for iOS applications

Heuristic:	A	B	C	D	E	F	G	H	Total
RapidCalc	1.1	0.9	0.4	0.4	0.8	1.3	1.3	2.0	8.2
GluCoMo	0.6	1.4	1.2	1.7	0.8	1.3	1.3	1.7	10
Diabetes Diary	0.5	0.8	0.9	0.5	0.9	0.7	1.4	0.8	6.5
Diabetes Personal Manager	0.6	0.8	0.9	1.1	1.1	0.9	1.6	2.3	9.3
Total	2.8	3.9	3.4	3.7	3.6	4.2	5.6	6.8	34

The application which produced the lowest score, and hence is deemed the most usable, was Diabetes Diary; and the least usable was GluCoMo, because it scored the most severe usability ranking. The heuristics that GluCoMo scored least well on were D and H, which were those concerned with minimalist design and error management, and this was probably because this application includes a lot of additional functionality beyond that specified by the tasks in Table IV.

The heuristics that had the highest total scores and were therefore the worst overall were G (Aesthetic, privacy and social conventions) and H (Realistic error management). This is because most applications did not offer any form of security when transmitting data, and because the iPhone does not include a 'Back' or 'Undo' button, unlike the other platforms considered here.

Results for Android

The four applications that were determined to be most efficient through the KLM analysis for the Android platform were also subject to heuristic evaluation, but this time by four expert evaluators. The average rankings for the Android applications are shown in Table XVI.

Table XVI: Average heuristic ranking for Android applications

Heuristic:	A	B	C	D	E	F	G	H	Total
OnTrack Diabetes	0.1	0.8	0.9	0.5	0.8	0.6	0.6	1.0	5.3
Glucool	0.3	0.3	0.4	0.4	0.3	0.6	0.6	0.3	3.2
Glucometro	0.5	1.0	1.4	0.8	0.9	1.1	0.6	1.7	8.0
Track3 Diabetes Planner	0.8	1.3	1.1	0.9	1.00	1.1	1.3	2.0	9.5
Total	1.7	3.4	3.8	2.6	3.0	3.4	3.1	5.0	26

The application which scored the lowest, and hence was most usable, was Glucool; and the least usable was Track3 Diabetes Planner. Some of the features that caused Glucool to score well were its good backup and restore options, its ability to log all data on a single screen and the clarity of its graphs. Track3 on the other hand was over-complicated with graphics that were difficult to interpret and no option to edit data. The heuristics that had the highest total scores and were therefore the worst overall were C (Consistency and mapping) and H (Realistic error management).

Results for Blackberry

Four expert evaluators performed the heuristic evaluation on the Blackberry application, and the averages of their evaluations are shown in Table XVII.

Table XVII: Average heuristic ranking for the BlackBerry application

Heuristic:	A	B	C	D	E	F	G	H	Total
iRecordit Diabetes Sugar Glucose and Health Tracker	2.44	1.58	1.38	1.88	1.5	1.5	1.75	0.42	12.4

The heuristic that was scored worst was A (Visibility of system status) because the battery status and the network status are never visible while using the application; the time is visible at the top of the screen but it is necessary to scroll down to add data which loses visibility of the status bar, including the time; and there is no option to backup up and restore the data leading to recovery problems. Heuristics D (Good ergonomics and minimalist design) and G (Aesthetic, privacy and social conventions) were found to be the next worst source of problems because some screens are too long and difficult to use; and the application does not have provisions for security and privacy as data is sent by unencrypted email.

6 Discussion of the Results

Perhaps the most striking observation is the sheer number of applications that are associated with diabetes management on each of the three platforms considered here. There were far more applications available for iOS and Android than Blackberry perhaps because the latter is aimed predominately at the business market. The quality of the applications was highly variable, with those written for the Android platform being particularly prone to crashing.

The keystroke level modelling highlighted some differences in the usability and efficiency of the apps on the various platforms. The number of keystrokes for finishing the tasks on average was 29 on the iPhone, 34 on the Android and 119 on the BlackBerry. The difference is evident especially between BlackBerry and the other platforms. This is due to the hard buttons that are on the BlackBerry devices, as opposed to the soft buttons and touch screens of the Android and iPhone devices. So with a Blackberry a key must be pressed several times to move the mouse or the

cursor across the screen, but the other devices only need one click. The efficiency analysis has led us to suggest the following guidelines:

- Use default settings where possible (eg email addresses, dose calculation parameters)
- When using pickers and sliders, do not make them too fine grained, otherwise data entry is very inefficient
- When entering numeric data, display a numeric keyboard instead of an alphabetical keyboard, and only use the integer keyboard when appropriate.
- Do not use acronyms without explaining what they mean [19].
- Personalisation of settings should be adjusted within the app, rather than within the device settings

The heuristic evaluation exposed a different set of problems from the KLM, because it was guided by the mobile heuristics of [6], which are not related to efficiency. The most commonly breached heuristics were A, C, D, G and H, which lead us to the following associated guidelines

- The battery status and time should be visible while using the application. (A)
- The network status should be visible while sending data. (A)
- Provide options for backing up and restoring data. (A)
- Do not include unnecessary options or irrelevant information. (C)
- Do not overload a screen with too many elements [20]. (D)
- All data transmission should be encrypted.(G)
- Do not allow unrecoverable errors.(H)

The heuristic evaluation also showed that the most usable application was found on the Android platform. Glucool on the Android was found to contain the least usability issues across all platforms. The leading application on either of the other platforms was Diabetes Diary which was found to be less than half as usable as Glucool. The following shows how the best rated apps on each platform compare.

- The best application for the iPhone for the heuristic evaluation is the Diabetes Diary (rated 6.5), which was the second best rated (with 22 keystrokes) application in the KLM analysis of the iOS applications.
- The best application for the Android platform for the heuristic evaluation is Glucool (rated 3.2), which was also the second best rated (with 34 keystrokes) application in the KLM analysis.
- Finally, the best (and unique) application for the BlackBerry platform for the heuristic evaluation is iRecordIt (rated 12.4), which had 119 keystrokes in the KLM analysis.

The Diabetes Diary application has now been tested further on potential users in a subsequent experiment, which was restricted to the iOS platform for practical reasons.

The gathering of secondary functional requirements generated a wider range for the iOS platform than in the Android or Blackberry platforms, the latter of which was

the most restricted, with only four additional functions. The full list of functionality gathered in Step 4 has also been presented to all of the participants in the user study for utility ranking. The most common functions are logging physical activity, logging other medication and allowing personal settings as they are included in applications of each of the three platforms, and so it is expected that these features will be rated highly by users.

7 Threats to validity

This study had a number of limitations. First, it was restricted to applications running on the iOS, Android and Blackberry smart phone platforms. These were chosen because they have the highest market share in the UK [21], and jointly occupy 75.5% of the total market. It was also restricted to native applications, since such applications can take advantage of device-specific features which are particularly interesting to developers. The user perception of applications, as demonstrated by characteristics such as the popularity rating, was ignored because the goal was to assess each one solely according to the attributes listed above. Finally, recent studies have shown that users overwhelmingly prefer apps that use the unique features of mobile phones to facilitate tedious tasks such as tracking blood glucose to those which offer purely reference information [13]. Therefore, this survey was focused on apps that were designed for such purposes, and tasks were devised specifically to evaluate data entry methods, data visualisation and personalisation.

8 Conclusions and Future Work

This paper contains a summary of the results of a systematic survey of mobile applications for diabetes management across three smart phone platforms. The platforms operated on devices with different characteristics, with the iPhone being entirely touch screen, the Blackberry having no touch screen capability and the Android phone with touch screen capability, as well as the three standard keys (Home, Menu and Back).

This study concentrated on apps that were designed for frequent logging of data, but which also offered the capability for data visualisation, transmission and personalisation. The research has achieved its goal in terms of exposing problems with efficiency and various other attributes corresponding to mobile heuristics, as well as helping to elicit functional requirements, and this has resulted in a list of guidelines for development. It is hoped that the protocol followed here will therefore also prove useful to developers in other application domains, as it is a lightweight protocol which is not too time-consuming.

The study was restricted both in the choice of platforms, and in the choice of diabetes as a healthcare domain. Future work will include a survey of similar apps on other platforms, and research into self-management applications for other branches of healthcare. The work described here was restricted to expert evaluators rather than potential users, and utilised only two of the many existing techniques for evaluating

user interfaces, namely KLM and heuristics. It might therefore be useful to consider incorporating other techniques in the future, such as cognitive walkthrough. Users are also being involved in a complementary user study, using the iOS app that scored best in this heuristic evaluation (Diabetes Diary). The purpose of the user study is to determine the nature and frequency of user errors, and to establish a comprehensive list of functional and non-functional requirements for the apps.

One of the shortcomings of this work was the KLM analysis, which involved counting the number of keystrokes to accomplish a task. This can be rather error-prone, since (for example) the number of finger swipes required to move a picker to a desired location can vary considerably, depending on the user. This is just one type of interaction introduced by mobile technologies that will be investigated in the future to determine how the KLM analysis technique can be adapted to provide a more accurate measurement of efficiency.

Acknowledgments

References

1. Alagöz, F., Calero Valdez, A., Wilkowska, W., Ziefle, M., Dorner, S. & Holzinger, A. (2010). From Cloud Computing to Mobile Internet, From User Focus to Culture and Hedonism: The Crucible of Mobile Health Care and Wellness Applications. IEEE 5th International Conference on Pervasive Computing and Applications (pp.38-45).
2. Preuveneers, D. & Berbers Y. Mobile phones assisting with health self-care: a diabetes case study. Proceedings MobileHCI '08 Proceedings of the 10th international conference on Human computer interaction with mobile devices and services (2008).
3. Bellazzi, R., Telemedicine and Diabetes Management: Current Challenges and Future Resarch Directions. Journal of Diabetes Science and Technology, 2(1), pp.98-104 (2008).
4. Calero-Valdez, A., Ziefle, M., Firat, A., Holzinger, A. (2010). Mental Models of Menu Structures in Diabetes Assistants. In: Lecture Notes in Computer Science (LNCS 6180), (pp. 584–591). Berlin, Heidelberg, New York: Springer. (ISBN: 978-3-642-14099-0)
5. Nielsen, J. Usability Engineering. Morgan Kaufmann Publishers. (1994).
6. Bertini, E., Catarci, T., Dix, A., Gabrielli, S., Kimani, S. & Santucci, G. Appropriating Heuristic Evaluation Methods for Mobile. International Journal of Mobile Human Computer Interaction, Vol. 1, Issue 1, pp. 20-41 (2009).
7. Martin, C., Flood, D. & Harrison, R. A Protocol for Evaluating Mobile Applications. Proceedings of the IADIS (2011).
8. Flood, D., Harrison, R., & Duce, D., "Using Mobile Apps: Investigating the usability of mobile apps from the users perspective", International Journal of Mobile HCI [Submitted]
9. Poropatich, R., Pavliscsak, H.H., Rasche, J., Barrigan, C., Vigersky, R.A, Fonda, S.J et al Mobile Healthcare in the US army. Wireless Health 2010, San Diego, California. p184-187.
10. W3C. Mobile Web Application Best Practices. http://www.w3.org/TR/mwabp/

11. International Organization for Standardization. Medical devices – Application of usability engineering to medical devices. International standard ISO 62366: 2007. Switzerland.

12. Kjeldskov, J. & Stage, J New techniques for usability evaluation of mobile systems. International Journal of Human Computer Studies, 60(5-6), pp.599-620. (2004).

13. Liu, C., Zhu, Q., Holroyd, K. A. and Seng, E. K. Status and Trends of Mobile-Health Applications for iOS Devices: a Developer's Perspective Journal of Systems and Software (July 2011)

14. Zhang, D. & Adipat, B. Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. International Journal of Human-Computer Interaction, 18(3), pp.293-308 (2005).

15. Card, S. K., Thomas, T. P. & Newall, A. The Psychology of Human-Computer Interaction., London: Lawrence Erbaum Associates. (1983).

16. Nielsen, J. & Molich, R. Heuristic evaluation of user interfaces. The Conference on Human Factors in Computing Systems: CHI90 (1990).

17. Greenhalgh, T., Hinder, S. et al. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ (Clinical Research Ed.), 341, p.c5814. (2010).

18. Martin, C. , Flood, D., Sutton, D., Aldea, A., Waite, M., Garcia, E., Garcia. A. A Systematic Evaluation of Mobile Applications for Diabetes Management using Heuristics. [Submitted].

19. Nielsen, J. & Tahir, M. Homepage Usability: 50 Websites Deconstructed. New Riders Publishing. (2001).

20. Rosenholtz, R., Li, Y., Mansfield, J. & Jin, Z.. Feature Congestion: A Measure of Display Clutter. The Conference on Human Factors in Computing Systems: CHI05 (2005).

21. http://www.onlinemarketing-trends.com/2011/04/smartphone-marketshare-2011-italyus-aus.html