EG UK Computer Graphics & Visual Computing (2020)
P. D. Ritsos and K. Xu (Editors)

Visualizing Usage Data from a Diabetes Management System

D.A. Duce[1], C. Martin[1], A. Russell[1], D. Brown[1], A. Aldea[1], B. Alshaigy[1], R. Harrison[1], M. Waite[2], Y. Leal[3], M. Wos[3], M. Fernández-Balsells[3], J. M. Fernández-Real[3,5,8], L. Nita[4], B. López[5], J. Massana[5], P. Avari[6], P. Herrero[7], N. Jugnee[6], N. Oliver[6], and M. Reddy[6]
[1]School of Engineering, Computing and Mathematics, Oxford Brookes University, UK
[2]Oxford School of Nursing and Midwifery, Oxford Brookes University, Oxford, UK
[3]Diabetes, Endocrinology and Nutrition Unit, Hospital Universitari Dr. Josep Trueta, Institut d'Investigació Biomèdica de Girona, Girona, Spain
[4]RomSoft SRL and Technical University of Iasi, Romania
[5]University of Girona, Girona, Spain
[6]Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, UK
[7]Centre for Bio-Inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London, UK
[8] CIBEROBN Fisiopatología de la Obesidad y Nutrición, Instituto de Salud Carlos III, Spain.

Abstract

This article explores the role for visualization in interpreting data collected by a customised analytics framework within a healthcare technology project. It draws on the work of the EU-funded PEPPER project, which has created a personalised decision-support system for people with type 1 diabetes. Our approach was an exercise in exploratory visualization, as described by Bergeron's three category taxonomy. The charts revealed different patterns of interaction, including variability in insulin dosing schedule, and potential causes of rejected advice. These insights into user behaviour are of especial value to this field, as they may help clinicians and developers understand some of the obstacles that hinder the uptake of diabetes technology.

CCS Concepts

• ***Human-centered computing*** → ***Information visualization;*** • ***Applied computing*** → ***Health informatics;***

1. Introduction

This paper is a reflective account of the use of visualization techniques to explore usage data collected from a mobile medical application. The application was developed in the EU-funded PEPPER project [PEP20], and empowers individuals with Type 1 diabetes mellitus to self-manage their condition. It employs case-based reasoning (CBR) [AP94], to provide advice about insulin bolus doses, which are self-administered either by injection or insulin pump [TFL19]. The system also includes a safety layer to guard against faults and inappropriate recommendations [LAL*19].

Diabetes self-management apps need to be studied using a rigorous research methodology [HWC*16]. The usability, safety and feasibility of the PEPPER system was, therefore, assessed in a multi-phase clinical trial, conducted at the Institut d'Investigació Biomèdica de Girona, Dr. Josep Trueta, Girona, Spain and Imperial College London, UK. Examination of data usage was one of the tools employed in the assessment of the system's usability, which was carried out by various means [MAD*18]. The clinical study culminated in a randomised controlled, crossover trial, lasting for eight months. It is important to emphasise that this research is intended as a complement, not a replacement, for the clinical data analysis, which is reported elsewhere [ALH*]. We do, however, propose a number of ways in which our approach to analysing usage data may be of use to clinicians, thus potentially leading to improved care.

The PEPPER system includes a bespoke analytics library, which we refer to as a *remote usability framework*. The framework automatically captures data about user interactions, without the need for direct human observation. This relatively unusual approach for diabetes technology research projects [IH01] has given precise metrics about user behaviour over an extended period, which is of crucial importance in assessing the efficiency and redundancies of features. The premise that underlies this paper is that there is a role for visualization of these data in enabling a broader view of how the application was used.

Our approach was an exercise in exploratory visualization, as described by Bergeron's [Ber93] three category taxonomy of descriptive, analytical and exploratory visualization. We sought to explore, for example, which paths were taken through the mobile interface, and at what time of day. Research questions were developed along these, and related lines, as detailed in Section 4.4. We began to realise that the quantitative usage data we were collecting had potential be a useful adjunct to qualitative analysis, such as that gathered by diary studies [WAA*20] and interviews of the kind used in the study by Raj et al. [RTG*19] on how context affects self-care.

This paper first introduces the PEPPER system. It then discusses

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

DOI: 10.2312/cgvc.20201144

delivered by EUROGRAPHICS DIGITAL LIBRARY
www.eg.org diglib.eg.org

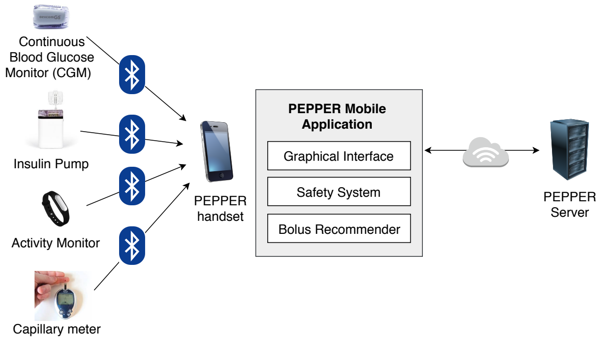


Figure 1: *PEPPER System Architecture*

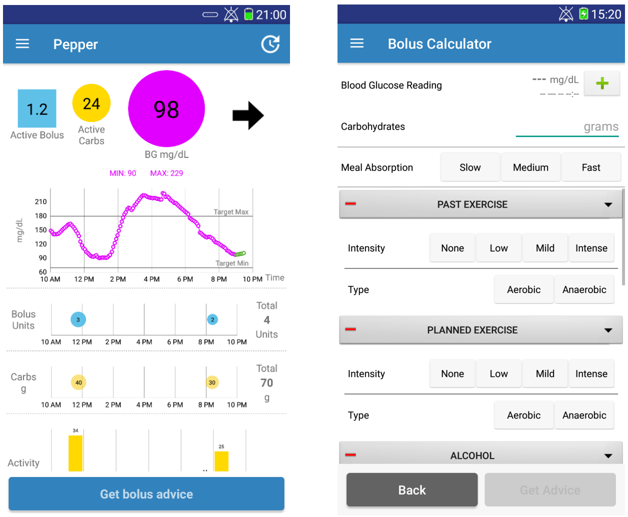


Figure 2: *PEPPER screens. Home Screen, showing blood glucose level, insulin doses, carbohydrates consumed and physical activity (left), Bolus Calculator (right)*

our rationale for creating a custom analytics library and sets out our approach in implementing the library within the system architecture. After exploring the potential for visualization to interpret the data, the paper then explains the significant benefits which this analysis of interaction data may offer healthcare technology projects.

2. The PEPPER System

2.1. Architecture

The PEPPER system architecture (see Figure 1) includes the following components:

1. PEPPER handsets and bluetooth-connected devices
2. Secure PEPPER web server
3. Meal-insulin bolus recommender
4. Safety System

PEPPER makes use of two types of handsets. One is designed for participants using multiple daily injections (MDI) of insulin, and the other for participants on continuous subcutaneous insulin infusion (CSII, insulin pump therapy). The pump-participant handset is a portable touch-screen device which communicates directly with the insulin pump. Its primary function is to allow the user to precisely manage insulin therapy by accepting or rejecting bolus insulin dose recommendations, calculated by the bolus recommender. Each rejected recommendation may be overridden by a user-supplied value. In addition, it measures and automatically records glucose levels via a continuous blood glucose monitor (CGM). A capillary blood glucose meter is used to calibrate the CGM device at least twice per day. The handset allows logging of food intake and other parameters and is connected to a physical activity monitor whose inputs are recorded.

The MDI handset is a commercially available smartphone with the insulin recommendation application running locally on the Android operating system. It has the same functionality as the pump participants' handset, except that it does not communicate with an insulin pump.

The bolus recommender algorithm uses case-based reasoning to supply users with advice about insulin dosage at meal times [TFL19]. The handsets wirelessly report the user's case history to the clinicians' portal on the PEPPER web server. Through this portal clinicians may add new users and review cases. A case comprises multiple parameters such as carbohydrate intake, blood glucose reading, meal composition, physical activity and hormone cycle. The Safety System addresses various safety considerations, providing predictive glucose alerts and alarms, dynamic safety constraints of the recommended insulin, carbohydrate recommendations to address hypoglycaemia, and a low-glucose insulin suspension feature for pump participants [LAL*19]. All the individual system components are available as part of an API library [PEP20].

2.2. PEPPER Screens

Two screens are principally used to communicate diabetes self-management information and advice to the participants: the Home Screen and the Bolus Calculator (see Figure 2). There is also a Statistics Screen that allows users to explore longer term trends in their blood glucose levels and the threshold for the percentage time in glycaemic range.

The Home Screen provides visualizations of blood glucose levels, physical activity, carbohydrate intake and insulin doses to enable at-a-glance self-monitoring. Users can change the time period shown on the graphs through pinch and spread gestures. A trend arrow quickly imparts recent changes to their blood glucose levels and a 30-minute glucose prediction trend is displayed.

The most complex feature of the PEPPER system from the user interface (UI) perspective is undoubtedly the bolus recommendation (BR) feature. The Bolus Calculator Screen enables users to enter a number of inputs which enable the CBR system to personalise insulin advice. Three inputs are mandatory: 1) a blood glucose reading which is automatically taken from the continuous glucose monitor, 2) the quantity of carbohydrates, 3) the absorption rate of the meal. It is also possible to enter a variety of optional inputs

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

which could affect insulin dosage, including hours of sleep, alcohol intake, tiredness, stress levels and so on. These are not pictured on the screenshot in Figure 2, but can be accessed by scrolling down on the same screen.

The bolus recommender allows the user to accept, reject or cancel the recommended bolus. If users reject a recommendation, they can enter their own chosen dose instead. Records of these decisions are saved to the PEPPER web server, except in the case of cancellations. It is, however, possible to retrieve evidence of cancellations from the usability data stream.

3. Motivations for Remote Usability Analysis

There were a number of reasons why we decided to use a remote usability framework during the trial. In clinical studies user behaviour must sometimes be measured "invisibly" to ensure that the outcomes are measured solely on the technological intervention. This requirement harmonised with our research goals, because we wished to observe the participants' interactions with the system in day-to-day use, rather than within a laboratory setting. Two major factors led us to create a customised analytics platform. Not all of the handsets were equipped to use popular analytics frameworks such as Google Analytics or Firebase. The handsets for the pump participants, in particular, only enabled a restricted set of Android operations in order to prevent events, such as notifications, from interfering with pump delivery. There were also regulatory restrictions on the pump handset as a medical device, which had to be respected. Aside from these technical considerations, we wished to transcend the limitations of generic analytics frameworks by honing in on interactions with specific modules of the PEPPER system, such as its CBR bolus recommendation feature. The constraints imposed by the trial protocol had many positive implications for our analysis. The automated capture of user events provided enormous savings in cost and gains in efficiency by comparison with human observation [IH01]. More importantly, automated logging revealed patterns of user behaviour over longer time periods which limited human observation could not hope to gather [LH12]. The further benefits of a customised remote usability framework to a healthcare technology project will be discussed in greater depth below, once our methodology and findings have been explained. A great advantage of our design was that we could coordinate the clinical data recorded on the PEPPER server with the usability data in order to provide a contextualised analysis of user behaviour.

4. Methodology

When constructing the usability evaluation framework we drew on the four-phase model devised by Balagtas-Fernandez and Hussmann, and revised by Kluth, Krempels and Samsel [BFH09, KKS14]. The data flow within our analytics framework is pictured in Figure 3. The distinct phases of the evaluation framework can be summarised thus:

1. **Preparation Phase.** The analytics library is integrated into the PEPPER system to capture user events.
2. **Capture Phase.** User interactions on registered UI components are logged by session. The events are sent to the Oxford Brookes University (OBU) server.

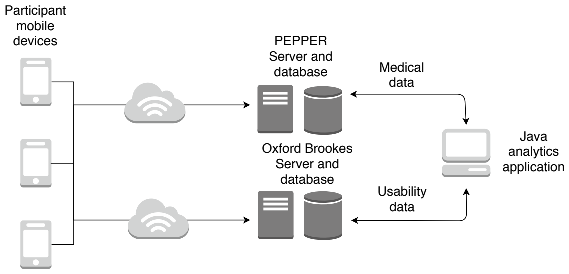

Participant mobile devices
PEPPER Server and database
Medical data
Java analytics application
Oxford Brookes Server and database
Usability data

Figure 3: *Data flow diagram*

3. **Analysis Phase.** User events are extracted from the OBU server by the analytics application. Pattern recognition is used to categorise sessions and isolate bolus recommendation interactions. Data is coordinated with clinical data from the PEPPER server.
4. **Critique Phase.** Evaluation of user interactions on the sorted data now takes place to answer research questions.

4.1. Preparation Phase

The data flow from the Java analytics library to the OBU server was separated from the flow of clinical data to the PEPPER server to protect the latter from failures.

Method calls were inserted in the application to notify the analytics library of interactions with specified UI components, such as buttons or menu items, and the analytics library logs these events, periodically sending its data to the OBU server. Each user event (a touch on the screen or the selection of a button) is assigned a session identification. A session is defined as a set of user events with an interval of no more than 5 minutes between any two events.

4.2. Capture Phase

During the capture phase, the analytics library recorded the following fields for each user event:

- **Timestamp**. An epoch timestamp in milliseconds recording the time at which the event occurred.
- **Event Type.** The type of UI component that generated the event (e.g. Menu Item, Motion Event, etc.)
- **Activity.** The name of the Android Activity in which the event was triggered.
- **Event Source.** The specific name of the UI component.

Timestamps were essential in producing statistics about the duration of sessions and specific event sequences, as well as the total time spent on each screen. The *Activity* classification provided a framework by which events could be easily sorted according to various analytical goals. Each Android Activity denotes a separate screen, so this distinction was useful for breaking down dwell-time by screen type. The classification could also be used to decompose a sequence of events into a screen flow. The *Event Source* classification was useful in fine-grained analysis where we isolated particular button presses and the selection of menu options. The analytics library also logged values denoting an event's absolute x- and y-coordinates on the screen. A limitation of our library was its inability to record particular motion types, such as tap, double tap, pinch,

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

pan, swipe, rotate etc. This prevented us from automatically isolating patterns of user gestures indicative of faulty UI design [PSC17]. Nevertheless, the fields listed above supplied us with a rich source of data on which to base our analysis.

4.3. Analysis Phase

The Java analytics application pictured in Figure 3 used regular expression pattern recognition to arrange the events into sequences of particular kinds, drawing on pioneering work on sequence mining in automated usability analysis [LGH14,KKS14]. In contrast to these studies, which aimed to create a generic resource for usability analysis, we targeted our regular expression pattern recognition to mine particular sequence types, especially participants' interactions with the bolus recommender.

The scheme classified events as letters from `A` to `Y` and interaction sequences as strings over this alphabet, such as `IQPPPKXPLR`. The letter `I` represented a touch event on the Get Bolus Advice button, `Q` a screen change to the Bolus Calculator, `P` a motion event, `K` a touch on the Get Advice button on the Bolus Calculator screen, and `X`, a screen change to the Bolus Advice screen. `L` was a touch event on the Accept button and `R` a screen change away from the Bolus Calculator and Bolus Advice screens. `M` and `N` represented touches on the Reject and the Cancel buttons respectively. Regular expressions were then used, for example, to identify sequences resulting in acceptance, rejection, or cancellation of advice.

After processing the data and executing the pattern recognition algorithms, the Java application created lists of sequences and sessions, as well as aggregated summaries, and entered these results in a series of spreadsheets. UI sessions were correlated with clinical data in the PEPPER server through time-stamp analysis. Visualization and statistical analysis techniques were then used on this data.

4.4. Critique Phase

Exploratory data analysis was used to make sense of the data. We began by formulating questions and then proceeded to visualize, transform and remodel the data to arrive at answers and to refine the following research questions [WG16]:

1. RQ1: [Features] Which features did participants use? What paths were taken?
2. RQ2: [Time] When did participants use the system, and for how long? Was time spent correlated with glycaemic condition?
3. RQ3: [Additional Inputs] What range of additional information did participants provide? Is this affected by the interface design?
4. RQ4: [Advice] Is there any (non-clinical) insight into when the advice provided by the system was not acceptable?

Visualizations were generated with *gnuplot* for graphics and *dot* for transition diagrams. Other visualizations not described here were generated with *R*. Bespoke *awk* scripts were used to generate the input files for these packages. These choices were mediated by familiarity and flexibility.

5. Results

Fifty-eight participants meeting appropriate clinical criteria were enrolled into the PEPPER trial across both sites. Fifty-four of the participants completed the run-in period, making up the intention to treat (ITT) study population. Twenty-six participants (48.1 %) were male and 28 (51.9 %) were female. The median (interquartile range) age was 41.5 (32.3-49.8) years with diabetes duration of 21.0 (11.5-26.0) years and HbA1c of 61.0 (58.0-66.1) mmol/mol. All were adults with well-controlled type 1 diabetes for over one year and previous structured diabetes education. The trial period comprised 4 stages: an initial 4 week run-in period, followed by a 3-month intervention phase, then a washout period of 3-4 weeks and finally a further 3-month intervention period. Full approval was obtained from the relevant ethics committees and regulatory bodies.

The participants were not all able to complete the trial, for various reasons. The results presented below use data from 33 individuals (24 MDI and 9 CSII participants) who fully completed the crossover and a further 10 (CSII participants) who completed one arm of the study only. Two eligible participants were omitted because of issues collecting full sets of usability data.

5.1. RQ1: Features

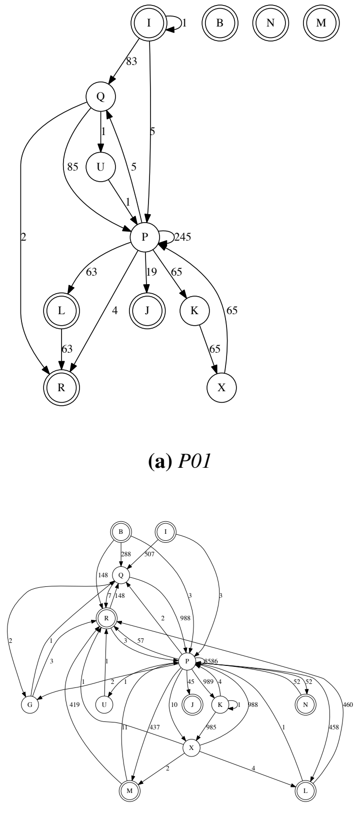
The usage of different features was explored by capturing the UI events and mapping them to sequences of strings, as explained in Section 4.3. A certain amount of analysis was possible by identifying the most commonly occurring sequences, and from frequency counts of letters in the sequences. Frequency counts showed that some features were barely used, and other configuration settings were never used at all.

The most common sequences for participants in the CSII and MDI groups were `IQPPPKXPLR` and `BQPPPKXPLR` respectively. These represent the most efficient way to request and accept bolus advice. The only difference is that CSII participants always entered the bolus advice screen from the home screen (`I`) whereas MDI participants entered from the navigation menu item (`B`).

Visualizations of the interaction sequences were generated as state transition diagrams, one per participant. Events are denoted by states (circles), and each transition (arc) is labelled with a number showing the occurrence count for that ordered pair of events, among all of the interaction sequences. Figure 4(a) shows a participant, in the MDI group, who used the system in a very regular way. The majority of sequences were `IQPPPKXPLR`. For comparison, (b) shows a participant in the same group with more complex usage patterns. Noticable here are the number of transitions to the `L` node (accept button) and a similar number to `M` node. Such information could potentially be used summatively, to frame questions about individual usage in an exit interview, or formatively, to improve efficiency of the interface.

5.2. RQ2: Time

Explorative visualizations of timing information were created for all participants, in tandem with evolving research questions. The charts were based on simple ways of visualizing time series data. See Guo *et al.* [GGJ*20] for a survey of techniques for visual analysis of event sequence data. In the first approach, each session was represented as a mark on a chart of time against day of trial. This basic representation was then enriched using different colours to indicate various types of session outcome, as shown in Figure 5. The



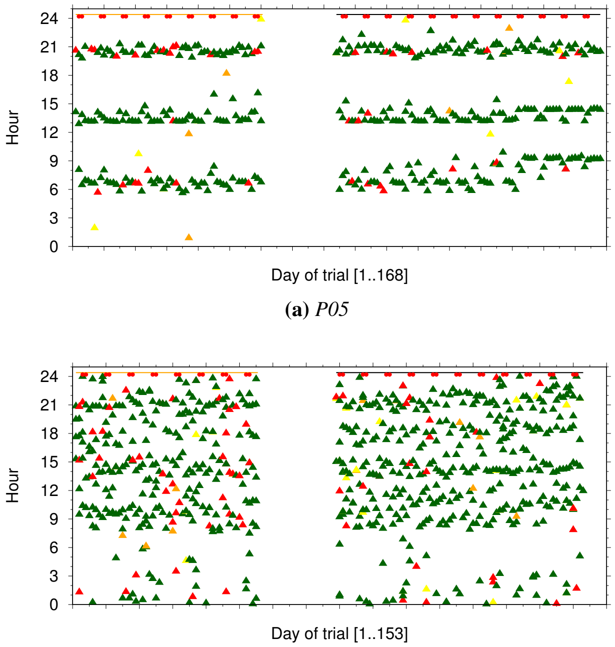
(a) *P01*

(b) *P02*

Figure 4: *Example sequence graphs, shown as state transition diagrams, for two participants (a) P01 and (b) P02.*

horizontal axis runs from the first day on which data were available to the last. Gaps correspond to days for which no data was collected, because of the washout period or a communication failure for example. The space at the top of this and subsequent charts is used to show the intervention periods as horizontal lines, coloured black for the control arm and orange for the other arm, which we refer to as the PEPPER arm. Weekend days are indicated as red dots.

Participant P05 (a) provides a good example of the regular schedule adopted by many MDI participants. The bolus recommendation interactions tend to cluster strongly around the 6 am line, presumably representing a session at breakfast, usually with an accompanying bolus dose. Other regular session times include just after midday and between 6 pm and 9 pm, presumably coinciding with meal times. It is not altogether surprising that for an MDI participant the system use should coincide with routines of insulin injection. The pump participants, on the other hand, tended to be more flexible about the times of day at which they requested bolus advice. P06's data (b) are typical of this group. Some faint clustering can still be seen around 9 am, for example, but it is obvious that the times of day at which interactions occurred were much more scattered than P05, with no strict routines. Because the pump infusion of insulin is less disruptive than a manual injection, it may be easier for the pump participants to be more flexible with their schedule of daily usage.



(a) *P05*

(b) *P06*

Figure 5: *Example time charts from participants (a) P05 and (b) P06. Colours indicate different outcomes: green - advice accepted, red - rejected, yellow - cancelled and orange - navigation error.*

Because we had the enormous and relatively unusual advantage of being able to coordinate clinical data (from the PEPPER server) and usability data (from the OBU server), we were able to situate user behaviour in the context of diabetes self-management. In particular, we wanted to account for the fluctuation in interaction time. Did participants spend longer on the PEPPER system during periods of hypoglycaemia (low blood glucose levels, less than 3.9 mmol/l) or hyperglycaemia (high blood glucose levels, above 10 mmol/l)? At these times, one would expect users to monitor their blood glucose levels with greater care. For illustration, a statistical analysis was carried out on a subset of the data collected at an early stage in the trial. We used approximately two months' worth of data from 22 participants. Using the matched pair t test (significance level $\alpha = 0.05$), we confirmed that for this small dataset the mean durations of sessions coinciding with hyperglycaemia were significantly longer than the mean durations of sessions within glycaemic target range ($t = 20.35s$, $df = 21, P < 0.001$), where df denotes the degrees of freedom. The target range was defined as 3.9–10 mmol/l. The mean session durations during hypoglycaemia were also significantly longer than mean sessions within glycaemic target range ($t = 6.61s$, $df = 4, P < 0.001$). It would seem, therefore, provisionally, that the system was useful to participants at times when establishing glycaemic control was a pressing priority.

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

5.3. RQ3: Additional Inputs

When requesting a bolus dose, users' blood glucose levels were automatically obtained from the CGM. In addition, the users were required to enter their blood glucose level, the value of the carbohydrates they were about to ingest and their meal absorption rate (slow/medium/fast). They could additionally enter any number of the following optional inputs relating to: Meal Absorption, Alcohol Type, Alcohol Quality, Hours of Sleep, Stress, Happiness, Tiredness, Hormone Cycle, Fever, Digestive Illness, Medication, Ambient Temperature, Past Activity Levels and Future Activity Levels.

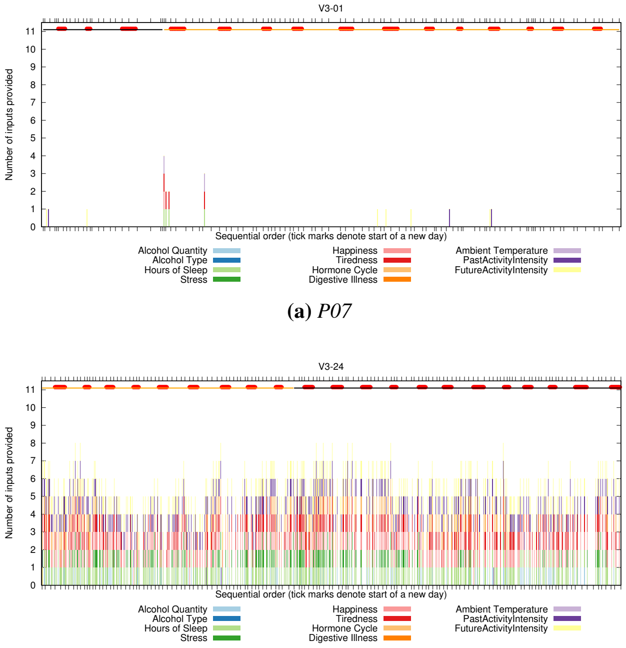
There was a large amount of variation between the participants in terms of the number of inputs entered, as can be seen from Figure 6, comparing P07 and P08's results. Each stacked bar contains coloured sections of equal length representing a distinct optional input entered. The bars are plotted against sequential days on the two arms of the trial, so there is no gap for the washout period since this is not to a time scale. Interactions in which no inputs were entered are represented by blank spaces.

Some users, such as P07, hardly entered any optional inputs. P08, on the other hand, entered a total of 2848 inputs over 1114 sessions, making use of a large spectrum of the available input fields. P08's average number of inputs per session was 2.5, whereas the average number of inputs per session for all the CSII participants at this site was 1.02 (the standard deviation was 0.80). The distribution of input counts for the participants shows, however, that the most common number of inputs in a Bolus Recommender session for many participants was 0. This applies even to P08 (588 sessions out of 1114 had no inputs, though the figure needs to be viewed at higher magnification to see the gaps) and P09 who entered the largest cumulative number of inputs. For comparision, the average number of inputs per session for the MDI participants at this site was 1.57 and standard deviation 1.01. There is evidence that the number of additional inputs reduced significantly as the trial progressed. It is obvious, therefore, that the optional inputs were generally underused. What is more, a small number of the fields were not relevant at all for these participants, e.g. Medication. Some individuals consistently entered the same type of information, such as hours of sleep and ambient temperature. These were non-trivial parameters to input, involving scrolling and sub-menus, which raises questions about perception of cause and effect, as well as interface design.

Exploratory visualization suggests further questions, such as what were the reasons behind these behaviour patterns, but answers cannot be given by exploration of this dataset alone. We return to this point in Section 6.

5.4. RQ4: Advice

We used data from the PEPPER server to determine the relative proportion of accepted and rejected bolus recommendations. The information was filtered to the date ranges of intervention phases of the trial, but included some cases that were removed from the clinical analysis [ALH*]. The clinicians calculated that the majority (88 %) of bolus recommendations were accepted by each participant. There was considerable variation between the acceptance rates of individuals however. We set out to understand why this was.



(a) *P07*

(b) *P08*

Figure 6: *Types of additional inputs provided for participants (a) P07 and (b) P08. Coloured sections in stacked bars represent distinct optional inputs.*

We used a variety of chart types to capture differences between accepted and recommended bolus doses, both for individuals and groups of participants. These included scatter plots of accepted against recommended doses, sequence plots of dosage difference, and plots of dosage difference against recommendation. Boxplots were useful in summarising the differences over all participants. Medians for positive and negative values of *accepted - recommended* doses were in the range 1.0 to 2.5 and 1.0 to 1.5 units respectively.

Drilling down into the data of participants with particularly low acceptance rates is informative. Here we explore data from the PEPPER arm for CSII participant P10, who had an acceptance rate of 29.4% during that period. Figure 7 plots this user's responses to the bolus recommendations. Those that were accepted are shown as green points. The majority are on the solid red line, though due to rounding differences in the interface and recorded data a few are offset by a small margin. Orange points indicate the chosen dose in cases where the user rejected the recommendation and administered a different dose. The dotted red lines are offset by the medians of the positive and negative differences between accepted and recommended dosage. It is noticeable that the median differences for rejected advice are much lower for this participant than those for the group as a whole at +0.75 and -0.5 units. The cluster of values for recommendations of 0 is noticeable. Dosage of 0 was recommended for 259 out of the 430 rejected recommendations and for 232 of these the accepted dosage was 1 or less. There is a tendency to increase the dose when the recommendation is low, and to re-

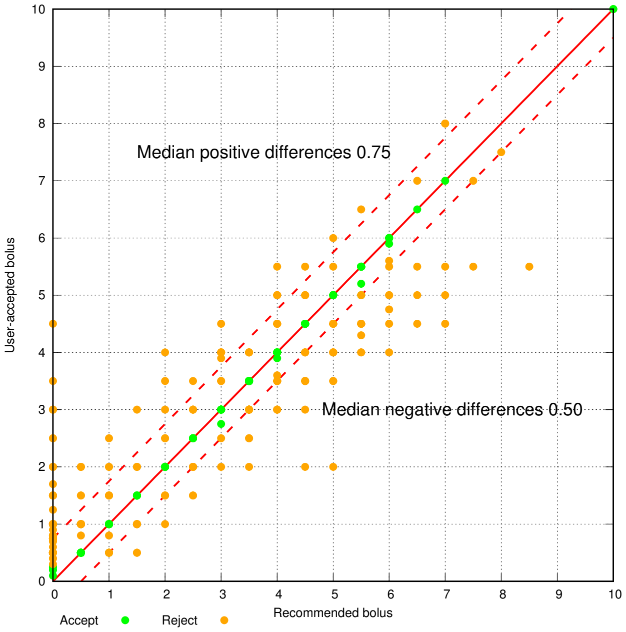


Figure 7: *Accepted vs recommended doses for P10*

duce it when high. This participant also cancelled a large number of Bolus Calculator interaction sequences.

These observations suggest questions that one might have liked to ask this participant in a follow up interview. What concerns did he/she have about low recommendations? Are there associated user interface issues (perhaps caused by rounding or the mechanics of the insulin pump)? Why were so many visits to the Bolus Calculator cancelled? Once again, this points to the need to use these analysis techniques in conjunction with other methods.

6. Discussion

Our experience is that remote usability analysis is a very useful supplement to qualitative data, such as that from interviews and questionnaires, that was collected during an earlier phase of the mixed-methods usability evaluation of the PEPPER system [MAD*18].

A remote analytics framework generates precise measurements about which aspects of an application are used, thus permitting a fuller consideration of a UI's design implications. But remote usability analysis may also have implications for diabetes treatment. As far as we know, these implications have not been discussed in depth before. Diabetes specialists recognise that the precise analysis of user engagement is an essential preliminary step to the improvement of type 1 diabetes self-care behaviours [DBW*18, LB19]. This being the case, we advocate for the increased use of remote usability analysis in the evaluation of diabetes self-management devices and in healthcare technologies more generally. Such frameworks have the potential to provide sophisticated and precise measurements which can refine the process of design, and can provide valuable clues about how user adherence can be maintained [Klo19].

6.1. Contextualised reading of user metrics

Remote usability analysis of the preliminary data set has shown that users spend more time on the system in sessions which coincide with periods of hyperglycaemia or hypoglycaemia than at other times. Other studies have shown that there is a demonstrable relationship between the frequent viewing of CGM screens and improved time in glycaemic targets [Wel18]. Our usability findings further nuance such conclusions, by showing, provisionally, that users spend greater time on the system at decisive moments of self-care intervention, monitoring how their glycaemic condition is changing.

We have also been able to isolate differences in daily usage patterns between CSII participants and MDI participants. This provides further evidence that insulin pumps may provide greater flexibility in diabetes self-management, and may free their users from a rigid daily schedule of blood glucose monitoring and injections.

Metrics about the usage time need to be handled with care in an application such as this. Longer than average usage may stem from user difficulties with the UI. But it could, more positively, stem from a higher than average level of engagement. Less usage time need not, however, imply a lack of engagement. The stabilisation of a user's glycaemic condition, for example, might lead to a change in patterns of behaviour, including decreased time on a self-management device [KKPB15]. The consideration of context is extremely important when evaluating time metrics [RTG*19]. Figures about usage time should be coordinated with user feedback to assess how well a tool fits with various personal self-management strategies. Does the investment of time match the rewards for the user in the form of increased glycaemic control, improved quality of life and mental ease [dCA*17]? The findings of remote usability analysis will help to frame these discussions, and provide one source of empirical evidence of user satisfaction or dissatisfaction.

6.2. Assessment of cognitive overload

Remote usability analysis allows us to determine which features of an application are used and which are not, more precisely than through user feedback alone. This is especially important when dealing with a set of features which may place excessive demands on users' time or attention. There are, of course, sound clinical reasons for a completist approach to system design. But such an approach has the potential to clash with usability considerations.

The users only sporadically entered the optional inputs on the bolus recommendation screen. The designers of the bolus recommender understandably wished to have a complete record of the user's condition when a dose was requested. Given the reluctance of users to enter the optional inputs, a reconsideration of our strategy may, however, be in order. Should users receive more training about the importance of optional inputs in personalising dosage recommendations? Or should the number of inputs be reduced in order to keep users engaged? The latter path may be preferable if it emerges that some of the optional inputs had only a minimal impact on the size of bolus recommendations. The point of this discussion is not, however, to advance general recommendations about design, but to highlight the kinds of analysis and questions that remote usability metrics can promote.

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

7. Related Work

Hilbert and Redmiles [HR00], in their paper published in 2000, survey methods for extracting usability information from user interface events and provide a framework for categorisation. Their highest level categories are: synchronisation and searching (enabling interface events to be linked with other data sources), transformation, analysis, visualization and integrated support. The work described in this paper falls into the first three categories, enabling us to combine data from two different sources, to analyse temporal event sequences and to visualize results to support analysis.

Zapata et al. [ZFAAT15] published a systematic review of the literature (up to 2014) on usability in mHealth Apps in 2015. Their conclusions identified the need to use automated evaluation tools, with 73% of the papers selected using only interviews or questionnaires. Some of the authors of this present paper carried out a systematic review of the literature in human factors for advanced technology for patients with type 1 diabetes, including the use of data logging [WMF*18].

van Dortmont et al. [vDvdEvW19] present a visual analytics method for the combined exploration of time series and event series data. One of their case studies is intensive care unit data where blood pressure and other monitors provide time series data and interventions such as administration of medication are event series. Their approach to exploring correlations combines algorithmic and interactive techniques. A recent paper by Nguyen *et al.* [NHC*19] describes an elegant application of visual analytics to the analysis of user behaviour, at individual and group levels. They discuss generalizability to contexts that can be categorized by *actions*, *segments* and *actors*. Groups are identified through common tasks, though since all PEPPER users perform similar tasks this might not be the best key to exploring PEPPER data.

Our work can be situated in the field of Computational Ethnography, defined by Zheng et al. [ZHWA15] as 'a family of computational methods that leverages computer or sensor-based techniques to unobtrusively or nearly unobtrusively record end users' routine, *in situ* activities in health or healthcare related domains for studies of interest to human-computer interaction.' The authors argue that the benefits of recording activity in this way include higher objectivity, less intrusion and better scalability for aggregation and analysis. They also emphasise the importance of combining analytics data with that from other sources, as we do here, using clinical data recorded in a different part of the PEPPER system.

Sukumar et al. [SMG*20] describe how interaction logs have been used in the analysis of a visualization interface for exploring personal data such as activity. Little detail is provided about the types of logging data gathered and the method of collection. Interactions were characterised by type and metrics such as time spent were computed. There is the potential for integrating personal medical data in interfaces of this kind.

8. Conclusions

The PEPPER UI is unique and informative, enabling evaluation on an individual basis, through familiar visualization techniques. We have been able to identify trends that can aid with targeted user education, and well as ongoing system/product development.

We recommend this framework approach in the development of healthcare technology; the findings offer a set of indicators which could prove useful to clinicians. If a certain user is consistently rejecting a large number of bolus recommendations, and there is a pattern to these rejections, clinicians may consider modifying this user's treatment. This could be accomplished by changing the insulin to carbohydrate ratio (ICR) and insulin sensitivity factor (ISF) which are important factors in the decision support calculations. On the other hand, in some circumstances further user training and clinical advice might be a better alternative. Low acceptability of a diabetes self-management tool may adversely affect user adherence in the long run, which will have implications for the treatment provided. There are no hard and fast solutions to this. Responses to users who lack confidence in the system need to be personalised on a case by case basis. Our contention is that data provided by remote analysis are a valuable, and indeed essential, tool to enable this.

This exploration of visual presentations of the information collected by a remote usability framework has brought a new dimension to a mixed-method usability study. The framework has been implemented seamlessly within a clinical study, placing no burden on users or clinicians, whilst generating new insights and research directions. The techniques have potential to be used formatively, to improve the efficiency of a user interface. The visualizations could also enrich a summative evaluation methodology, in combination with qualitative data analysis of the kind used in an separate phase of the PEPPER project, to gain a more holistic understanding of the reasons for observed patterns of user behaviour, including deviations and outliers.

In circumstances where experimental activity is restricted (e.g. the current Covid-19 pandemic), remote usability analysis may well be an attractive alternative. From our experience there are a few factors to bear in mind. (1) The ability to obtain logs of appropriate attributes and level of granularity from applications, through existing facilities or enhancement; (2) appropriate shared identifiers to combine logs from different sources and (3) filters to transform log data to the inputs required for an analytics tool chain. We found that using familiar visual representations lessened potential barriers to communication in the multi-disciplinary project team.

9. Acknowledgement

This work has received funding from the EU Horizon 2020 research and innovation programme, grant agreement No. 689810. The authors wish to the study participants for their valuable time, and the anonymous referees for their helpful suggestions.

References

[ALH*] Avari P., Leal Y., Herrero P., Wos M., Jugnee N., Thomas M., Massana Q., Lopez B., Nita L., Martin C., Fernández-Real J. M., Oliver N., Fernández-Balsells M., Reddy M.: Safety and feasibility of the pepper adaptive bolus advisor and safety system; a randomized control study. In preparation. 1, 6

[AP94] Aamodt A., Plaza E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Commun. 7*, 1 (Mar. 1994), 39–59. doi:10.5555/196108.196115. 1

[Ber93] Bergeron D.: Visualization reference models (panel session position statement). In *Proceedings of IEEE Visualization '93* (1993), Nielson G., Bergeron D., (Eds.), IEEE Computer Society Press. 1

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.

[BFH09] Balagtas-Fernandez F., Hussmann H.: A methodology and framework to simplify usability analysis of mobile applications. In *Proceedings of the 2009 IEEE/ACM International Conference on Automated Software Engineering* (Washington, DC, USA, 2009), ASE '09, IEEE Computer Society, pp. 520–524. doi:10.1109/ASE.2009.12. 3

[DBW*18] Duke D. C., Barry S., Wagner D. V., Speight J., Choudhary P., Harris M. A.: Distal technologies and type 1 diabetes management. *Lancet Diabetes & Endocrinology 6* (2018), 143 – 156. doi:10.1016/S2213-8587(17)30260-7. 7

[dCA*17] del Rosario Vallejo Mora M., Carreira M., Anarte M. T., Linares F., Olveira G., Romero S. G.: Bolus calculator reduces hypoglycemia in the short term and fear of hypoglycemia in the long term in subjects with type 1 diabetes (CBMDI study). *Diabetes Technology & Therapeutics 19* (2017), 402 – 409. doi:10.1089/dia.2017.0019. 7

[GGJ*20] Guo Y., Guo S., Jin Z., Kaul S., Gotz D., Cao N.: Survey on visual analysis of event sequence data, 2020. arXiv:2006.14291. 4

[HR00] Hilbert D. M., Redmiles D. F.: Extracting usability information from user interface events. *ACM Computing Surveys 32*, 4 (December 2000), 384–421. doi:10.1145/371578.371593. 8

[HWC*16] Hood M., Wilson R., Corsica J., Bradley L., Chirinos D., Vivo A.: What do we know about mobile applications for diabetes self-management? a review of reviews. *Journal of Behavioral Medicine 39* (2016), 981–994. doi:10.1007/s10865-016-9765-3. 1

[IH01] Ivory M. Y., Hearst M. A.: The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv. 33*, 4 (Dec. 2001), 470–516. doi:10.1145/503112.503114. 1, 3

[KKPB15] Klasnja P., Kendall L., Pratt W., Blondon K.: Long-term engagement with health-management technology: A dynamic process in diabetes. In *AMIA Annual Symposium Proceedings. AMIA Symposium* (2015). URL: https://europepmc.org/abstract/med/26958211. 7

[KKS14] Kluth W., Krempels K.-H., Samsel C.: Automated usability testing for mobile applications. In *Proceedings of the 10th International Conference on Web Information Systems and Technologies - Volume 2: WEBIST,* (2014), INSTICC, ScitePress, pp. 149–156. doi:10.5220/0004985101490156. 3, 4

[Klo19] Klonoff D. C.: Behavioral theory: The missing ingredient for digital health tools to change behavior and increase adherence. *Journal of Diabetes Science and Technology 13*, 2 (2019), 276–281. doi:10.1177/1932296818820303. 7

[LAL*19] Liu C., Avari P., Leal Y., Wos M., Sivasithamparam K., Georgiou P., Reddy M., Fernández-Real J. M., Martin C., Fernández-Balsells M., Oliver N., Herrero P.: A modular safety system for an insulin dose recommender: A feasibility study. *Journal of Diabetes Science and Technology* (May 2019), 1–10. doi:10.1177/1932296819851135. 1, 2

[LB19] Liberman A., Barnard K.: Diabetes technology and the human factor. *Diabetes Technology & Therapeutics 21*, 138-147 (2019). doi:10.1089/dia.2019.2511. 7

[LGH14] Lettner F., Grossauer C., Holzmann C.: Mobile interaction analysis: Towards a novel concept for interaction sequence mining. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices* (New York, NY, USA, 2014), MobileHCI '14, ACM, pp. 359–368. doi:10.1145/2628363.2628384. 4

[LH12] Lettner F., Holzmann C.: Automated and unsupervised user interaction logging as basis for usability evaluation of mobile applications. In *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia* (New York, NY, USA, 2012), MoMM '12, ACM, pp. 118–127. doi:10.1145/2428955.2428983. 3

[MAD*18] Martin C., Aldea A., Duce D., Harrison R., Alshaigy B.: The role of usability engineering in the development of an intelligent decision support system. In *Artificial Intelligence in Health: First International Workshop, AIH 2018, Stockholm, Sweden, July 13-14, 2018, Revised Selected Papers* (2018), vol. LNAI 11326, Springer, pp. 142–161. doi:10.1007/978-3-030-12738-1_11. 1, 7

[NHC*19] Nguyen P. H., Henkin R., Chen S., Andrienko N., Andrienko G., Thonnard O., Turkay C.: Vasabi: Hierarchical user profiles for interactive visual user behaviour analytics. *IEEE transactions on visualization and computer graphics 26*, 1 (2019), 77–86. doi:110.1109/TVCG.2019.2934609. 8

[PEP20] PEPPER Project Website. http://www.pepper.eu.com, 2020. Accessed: 2020-02-10. 1, 2

[PSC17] Paternò F., Schiavone A. G., Conti A.: Customizable automatic detection of bad usability smells in mobile accessed web applications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2017), MobileHCI '17, ACM, pp. 42:1–42:11. doi:10.1145/3098279.3098558. 4

[RTG*19] Raj S., Toporski K., Garrity A., Lee J. M., Newman M. W.: 'My blood sugar is higher on the weekends': Finding a role for context and context-awareness in the design of health self-management technology. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), CHI '19, ACM, pp. 119:1–13. doi:10.1145/3290605.3300349. 1, 7

[SMG*20] Sukumar P. T., Martinez G. J., Grover T., Mark G., D'Mello S. K., Chawla N. V., Mattingly S. M., Striegel A. D.: Characterizing Exploratory Behaviors on a Personal Visualization Interface Using Interaction Logs. In *EuroVis 2020 - Short Papers* (2020). doi:10.2312/evs.20201052. 8

[TFL19] Torrent-Fontbona F., López B.: Personalized adaptive CBR bolus recommender system for Type 1 Diabetes. *IEEE Journal of Biomedical and Health Informatics 23*, 1 (2019), 387–394. doi:10.1109/JBHI.2018.2813424. 1, 2

[vDvdEvW19] van Dortmont M., van den Elzen S., van Wijk J.: Chronocorrelator: Enriching events with time series. *Computer Graphics Forum 38*, 3 (2019), 387–399. doi:10.1111/cgf.13697. 8

[WAA*20] Waite M., Aldea A., Avari P., Duce D., Herrero P., Jugnee N., Leal Y., LÃşpez B., Martin C., Oliver N., Reddy M.: Trust and contextual engagement with the pepper system: The qualitative findings of a clinical feasibility study. In *Proc. ATTD* (2020), Mary Ann Liebert, pp. A18–A19. doi:10.1089/dia.2020.2525.abstracts. 1

[Wel18] Welsh J. B.: Role of continuous glucose monitoring in insulin-requiring patients with diabetes. *Diabetes Technology & Therapeutics 20* (2018), 42–49. doi:10.1089/dia.2018.0100. 7

[WG16] Wickham H., Grolemund G.: *R for Data Science: Import, tidy, transform, visualize and model data.* O'Reilly Media, Sebastopol, CA, 2016. 4

[WMF*18] Waite M., Martin C., Franklin R., Duce D., Harrison R.: Human factors and data logging processes with the use of advanced technology for adults with type 1 diabetes: systematic integrative review. *JMIR human factors 5*, 1 (2018), e11. doi:10.2196/humanfactors.9049. 8

[ZFAAT15] Zapata B. C., Fernández-Alemán J., Ali I., Toval A.: Empirical studies on usability of mHealth Apps: A systematic literature review. *J Med Syst 39*, 1 (2015). doi:10.1007/s10916-014-0182-2. 8

[ZHWA15] Zheng K., Hanauer D. A., Weibel N., Agha Z.: Computational ethnography: Automated and unobtrusive means for collecting data *in situ* for human-computer interaction evaluation studies. In *Cognitive Informatics for Biomedicine*, Vimla L. Patel T. G. K., Kaufman D. R., (Eds.). Springer International Publishing Switzerland, 2015. doi:10.1007/978-3-319-17272-9_6. 8

© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association.