Benchmarking User Interface Designs of a Patient Assistance Device for Peritoneal Dialysis: Quantitative Evaluation with Mobile Eye Tracking

Stephan Hess[1], Quentin Lohmeyer[1], Dimitri Wahlen[1], Sandra Neumann[2], Jean-Claude Groebli[2], Mirko Meboldt[1]

Keywords: home care, human factors engineering, usability, self-management, quantitative research, mobile eye tracking, benchmarking, quantitative evaluation

Introduction

Chronically ill patients cared for at home experience a higher health-related quality of life and a normalisation of everyday life, which is less dominated by the disease [1-3]. Therefore, 82% of end-stage renal disease patients and their families, if fully informed about their treatment options, would choose a home modality [4]. However, only 14% of dialysis patients in Europe are treated at home [5]. The main obstacle to home care is the availability of staff such as community nurses, neighbours or relatives [6, 7]. In order to give a broad range of people the opportunity to ensure and support home care, one approach is to design medical devices with greater ease of use. This allows non-experts to use medical devices that were originally too complicated to use. For a user-centric development of such medical devices, it is essential to understand which user interface (UI) design best supports patients, caregivers, and healthcare professionals [8, 9].

Human factors engineering is driving user-oriented design. Therefore, it has to take care of customized product UIs by testing them together with the intended users. Methods such as observations, questionnaires and interviews are used today to gain insight into the usability of a user interface. Here the focus is mainly on the graphical user interface on a screen [9-14]. However, users gain most information through visual perception [15] and the short-term memory has only a limited capacity [16, 17]. Therefore, it is difficult to understand the causes of use errors using traditional methods. Eye tracking provides a first-person perspective of the user and continuous localization of the gaze point. Eye tracking thus allows objective feedback to find perception problems [18, 19] and to gain valuable insights into hotspots in attention distribution on the user interface. This information can be used for both qualitative and quantitative evaluation of the usability of the UI.

As a result, in recent years eye tracking has increasingly become a method for testing perception and improving or evaluating the features of UIs. Examples are web and print advertisements [20, 21] and graphical representations like X-ray images of patients [22]. More complex subjects of the investigations were graphical UIs such as computer tomography interfaces [23], or spacecraft displays [24, 25]. Further, there are single studies where eye tracking is used to evaluate highly interactive UIs of tangible products like smart TVs [26], smart watches [11, 18] or medical devices [27, 28]. Most studies used a remote eye tracking system where the stimulus is presented on a screen and participants are asked to sit still in front of a desk. Aside this setup, mobile eye tracking with minimal invasive head mounted systems provides a degree of freedom in movability. This promises a natural user behaviour in the testing of tangible medical devices [28].

Using the benefits of mobile eye tracking, this work aims to gain a deeper understanding of the challenges in user cognition and thus of the obstacles to the user-friendliness of single UI features of a medical device. The medical device is a patient assistance device intended for home use in peritoneal dialysis (PD) therapy. Therefore, this article describes, to the authors' knowledge, the first benchmark tests of two different UI designs based on physiological measurements using mobile eye tracking.

Methods

The aspiration of the study was to present the intended use for the evaluated medical device. As a result, representatives of a diverse user group, which may be subdivided by age (young adults and seniors) were involved and the study was conducted in the intended environment. The medical device handling cycle consisted of seven main tasks (see Figure 2). The description of the stimuli, the characteristics of the participants, the data collection procedure and the subsequent data analysis are the focus of the following sections.

Stimuli

The stimuli of the study were prototypes with two different UI designs (D1 and D2, see Figure 1) of a medical device system. The system consists of the medical device itself, an inlet for guiding and manipulating a bag system with dialysis fluid and a catheter which is connected to the patient in the real therapy application. The most important interface features of the medical device are the buttons for manipulating the bag system and the lever for moving the inlet inside the device. The inlet has functions for fixing, clamping and breaking a fragile inside the bag lines. The medical device system supports peritoneal dialysis handling and is aimed at adults, including seniors (over 65 years of age). The stimuli provide acoustic (click sounds), haptic (positioning by stops) and visual (clear states and observation windows) feedback. Both prototypes support the same functionalities and require the same handling steps. At the top level, appearance of the UI designs was neutral in a monochrome design, as shown in Figure 1, to eliminate the effects of different colouring as an additional variable.

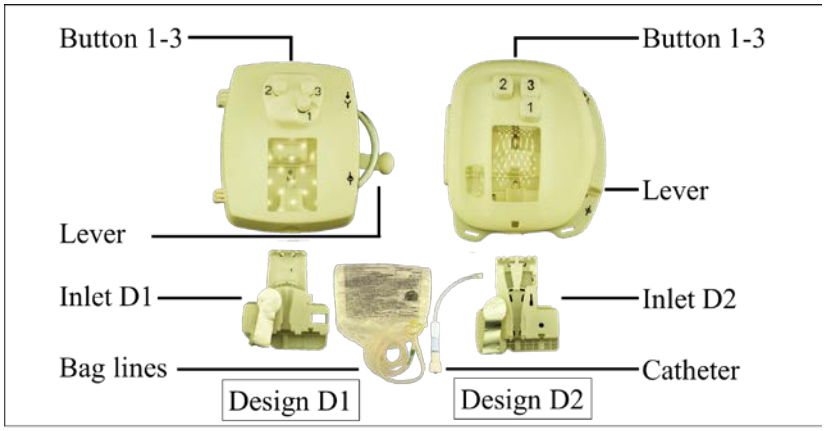


Figure 1. Illustration of the two UI designs D1 (left) and D2 (right) including the features *lever* and *button 1-3*. Additional parts for the therapy handling with the medical device are *bag lines*, *catheter* (standard parts that are used in the standard therapy) and *inlet* (respectively compatible to the UI designs D1 or D2).

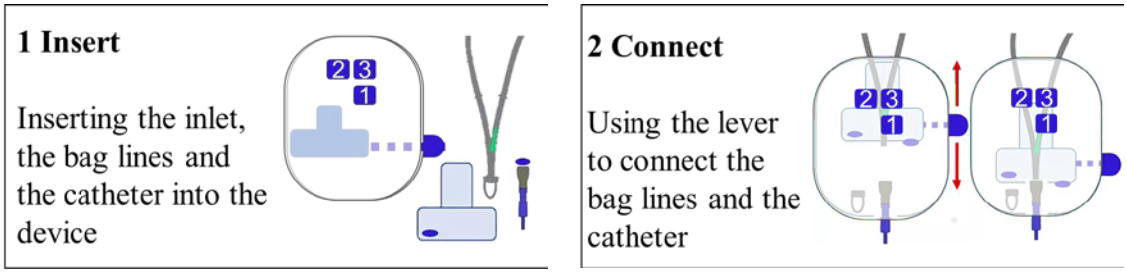
Recruitment and data exclusion

A total of 25 participants consisting of 15 seniors (10 men and 5 women, average 74.0 years, range 67-90 years) and 10 young adults (8 men and 2 women, average 25.1 years, range 24-26 years) were recruited and participated in this study. All participants were in good physical and mental condition and assessed the suitability of a study participation themselves. No participant was familiar with peritoneal dialysis therapy or mobile eye tracking. All participants had normal or corrected vision with contact lenses or corrective lenses that could be connected to the mobile eye tracking system.

One senior left the study prematurely after the first handling cycle and was therefore excluded from the analysis. For five seniors and one young adult, the data quality was insufficient due to measure errors by the eye tracker caused by factors of drooping eyelids, watery eyes or long eyelashes. Therefore, this data was excluded from the data set. In order to create a complete counterbalance, the data sets of a randomly selected senior and a young adult were not included in the data analysis. Thus, a total of 16 data sets with eight data sets from each group of young adults and seniors could be analysed. Four participants in each group started with D1 and four with D2, thus achieving a complete counterbalance. As a result, this within subject design mitigated the effect of individuality. Consequently, measures that naturally differ from participant to participant, such as fixation durations, could be compared with this balanced design of the study.

Study procedure

When the participants arrived in the test environment, they were welcomed and thanked for their participation. Before the study began, participants were invited to read the information for participants, containing information of the goal of the study, data safety and data management. If they agreed to participate in the study, they were asked to sign the consent form. Subsequently, the participants put on the mobile eye tracking system and the moderator conducted a three-point calibration. Since all participants were beginners in PD therapy and in the use of the device, the moderator briefly described the disease and the associated PD therapy. Next, the moderator showed the handling procedure with a low-level representation of the UI, which was designed and built for this purpose in addition to D1 and D2. After the introduction, the participants started the handling cycle of tasks 1-7 (see Figure 2) with either D1 or D2, guided by instructions. Each instruction was printed in a neutral design on an individual sheet to test the usability of the medical device and not the instruction. Subsequent to the first completed handling cycle, participants were asked to give their feedback on usability in a semi-structured interview with predefined high-level questions guiding to the root-causes of handling difficulties and use errors. Starting with the handling cycle, this process was repeated for the remaining prototype of the UI design. Once the participants had completed the handling study, they were again thanked for their participation and subsequently said goodbye.



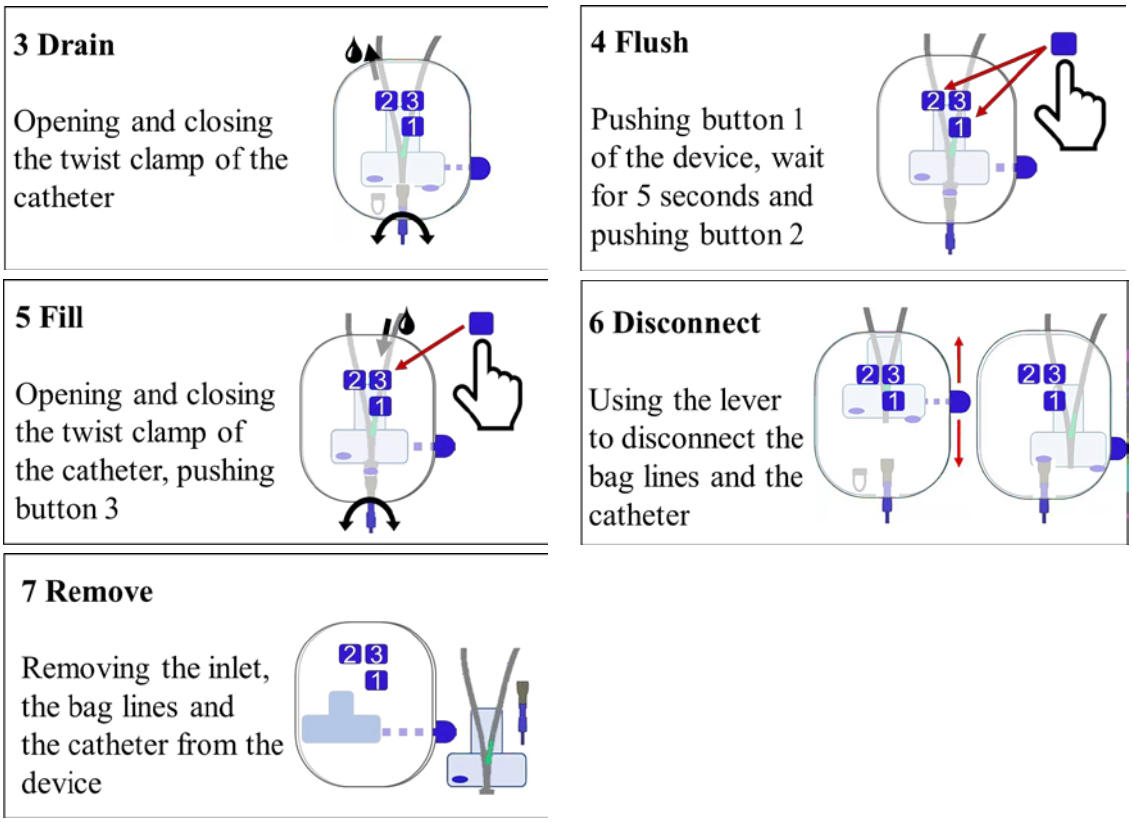


Figure 2. Seven tasks with the medical device. The user interacts manually with the inlet, the bag lines, the catheter, and the UI features of the device button 1-3 and lever.

Data analysis

In the data analysis, a two-step approach was used. It started with the analysis of task effectivity searching for use error related tasks. Subsequently, the in-depth gaze data analysis focused on these identified critical tasks. Gaze data was recorded with the mobile eye tracking system SMI ETG 2 with a scene resolution of 1280 x 960 px (viewing angle: 60° horizontal, 46° vertical) of the front camera offering a sampling frequency of 24 Hz with the gaze point measurement having an accuracy of 0.5° over all distances. The raw gaze point data was classified into three events fixation (nearly no eye movement), saccade (fast eye movement) and blink (closed eye) by an algorithm provided by the manufacturer of the mobile eye tracking system. This information was used in the gaze data analysis.

For the analysis of the task effectivity, the handling process of the participants was observed via a live recording from the first person's perspective from the eye tracking system. The performance in each task was evaluated by an observer. In the evaluation, two categories were distinguished according to the international standard IEC 62366-1 (2015). The first category safe use is defined as “normal use without use error” [29]. The second category use error is defined as “user action or lack of user action while using the medical device that leads to a different result than that intended by the manufacturer or expected by the user” [29].

In order to understand the challenges of user cognition in the critical tasks, gaze data was analysed with regard to single UI features of the medical device system (Figure 1). Therefore, it focused on the two dimensions dwell time (DT) and fixation duration (FD) of

the gaze. DT is the cumulative total time that the gaze is directed to a particular evaluated feature for a particular task. It is associated with the length of the information extraction [25, 35]. FD is the time, in which the gaze approximately stands still. It is associated with the processing depth, which leads to longer fixations [30-33] and with the rate of information extraction [20, 33, 34].

For a better understanding of the two analyzed dimensions length and depth of user perception, Figure 3 combines the information of the two measured parameters. Evaluating the user perception of all participants as a whole, it shows the relationship between mean fixation duration (FD) and mean dwell time (DT) for individual UI features of D1 compared to D2. Based on the two analyzed dimensions assigned to D1 in the coordinate origin, the mean FD and mean DT of D2 can be longer or shorter. Consequently, four different categories or patterns can be distinguished. The boxes in Figure 3 represent dwell times and the black bars represent single fixations in order to visualize the changes in the four categories from D1 to D2. A suggested interpretation of these patterns in terms of workload or gaze behavior is given in Figure 3. The equations for calculating the values of the shift in both dimensions (ΔDT and ΔFD) from D1 to D2 for the diagram are as follows.

$$\Delta DT = \frac{DT_{D2} - DT_{D1}}{DT_{D1}} \qquad (1)$$

$$\Delta FD = \frac{FD_{D2} - FD_{D1}}{FD_{D1}} \qquad (2)$$

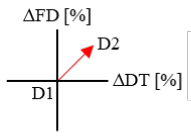
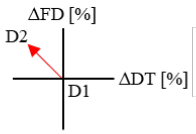
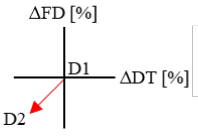
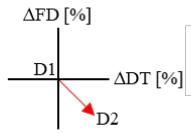
	Visual representation	Schematics of dwell time and fixations	Explanation
Category 1			Combination of *longer dwell time* and *longer fixation duration*. This indicates a *more scrutinizing* eg in handling tasks.
Category 2			Combination of *shorter dwell time* and *longer fixation duration*. This indicates *less skimming* eg while searching or checking.
Category 3			Combination of *shorter dwell time* and *shorter fixation duration*. This indicates a *less scrutinizing* eg in handling tasks.
Category 4			Combination of *longer dwell time* and *shorter fixation duration*. This indicates *more skimming* eg while searching or checking.

Figure 3. Visualization of shifts in two dimensions of the physiological gaze data measurements fixation duration (FD) and dwell time (DT). The displayed shifts are from a Design 1 in the coordinate origin to a Design 2, presented in the left column. In the middle

column, the boxes represent dwell times and the black bars represent single fixations. In total, a distinction is made between four categories. The right column explains the four patterns.

Results

Each of the 16 participants performed 30 handling steps in the seven tasks with both UI designs, resulting in a total of 480 evaluated handling steps for each UI design. The results of the task performance are shown in Figure 4. Overall, 97% of the handling steps were performed correctly for D1 and 96% for D2.

According to the results, the main challenges were in task 1 *insert*, task 2 *connect* and task 6 *disconnect* for both UI designs. The remaining four handling tasks were performed without errors, except for one missing catheter closure in task 5 *fill* with D2. The observed use errors in the first task were mainly incorrectly inserted bag lines in the inlet. Further use errors were forgetting to attach the cap of the bag lines to a safety feature of the device and folding the protective film of the inlet outwards. All use errors were corrected at a later stage. The safety function for the bag line cap blocks the system, when it is not activated. In task 2 *connect*, use errors occurred when the lever should have been used to connect the bag lines and the catheter. In task 6 *disconnect*, some participants forgot to operate the lever for disconnecting the catheter from the bag lines and for placing a new cap onto the catheter.

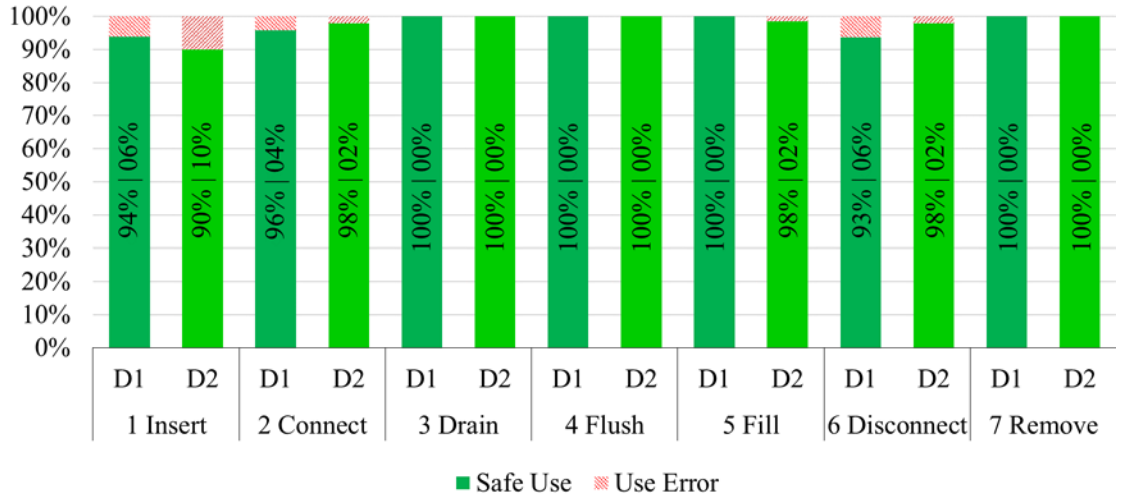


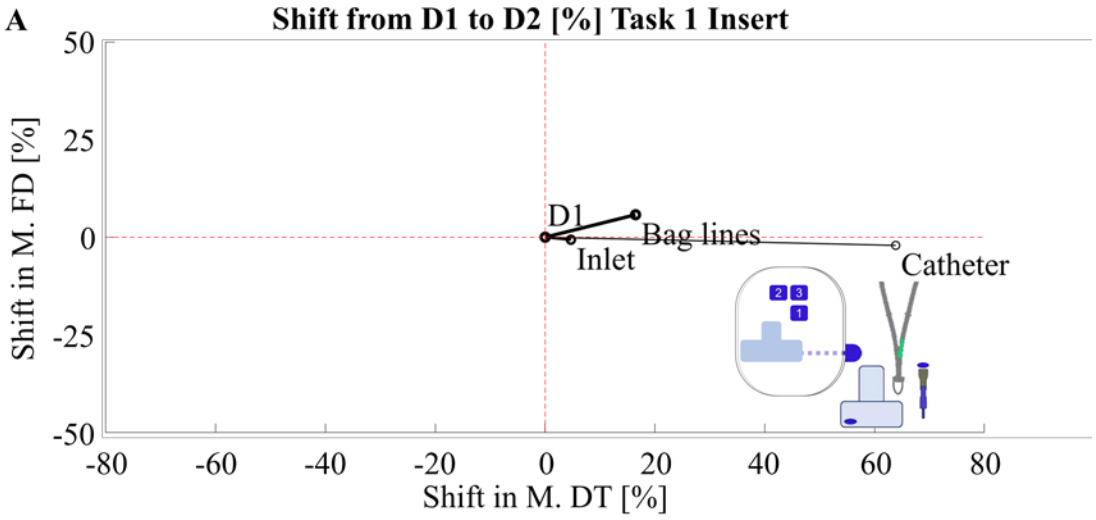
Figure 4. Comparison of task performance between D1 and D2 for all seven tasks. Evaluation in two categories safe use and use error according to IEC 62366-1 (2015).

Table 1 focuses on the handling tasks with observed use errors. It shows the results of the data analysis of the physiological gaze data in both dimensions. The mean values for FD are given in milliseconds and for DT in seconds. The mean FD for single UI features was between 149 and 405 milliseconds. The mean DT for single UI features was between less than 1 second and 28 seconds. At task 1 *insert*, there were large shifts for the UI features bag lines, inlet and catheter, and the UI features levers and buttons. While the first group is looked at for an average of 7 to 28 seconds, the second group is looked at for an average of less than 1 to 3 seconds. At task 2 *connect* and task 6 *disconnect*, the DT varies between less than 1 second for the buttons and 3 seconds for the bag lines and catheter. For the mean FD, clustering was not found in any of the three tasks.

Table 1. Analysis of eye tracking metrices for the user interface features bag lines, inlet, catheter, lever and buttons. Mean dwell time of the gaze in seconds and mean fixation duration in milliseconds for D1 and D2. The evaluated tasks are insert[a], connect[b] and disconnect[c].

	UI Features									
	Bag lines		Inlet		Catheter		Lever		Buttons	
	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2
M. Dwell Time [s]										
	24.3[a]	28.3[a]	19.2[a]	20.1[a]	7.4[a]	12.1[a]	1.0[a]	2.8[a]	0.5[a]	0.6[a]
	2.7[b]	1.8[b]	1. 5[b]	1.5[b]	2.5[b]	2.8[b]	1.7[b]	1.7[b]	0.3[b]	1.0[b]
	1.0[c]	0. 8[c]	1. 5[c]	1.6[c]	2.2[c]	1.7[c]	1.0[c]	1.6[c]	0.3[c]	0.4[c]
M. Fixation Duration [ms]										
	383[a]	405[a]	309[a]	307[a]	334[a]	327[a]	239[a]	337[a]	268[a]	223[a]
	356[b]	360[b]	349[b]	286[b]	380[b]	371[b]	309[b]	298[b]	236[b]	195[b]
	333[c]	364[c]	342[c]	373[c]	299[c]	343[c]	259[c]	289[c]	199[c]	149[c]

For better understanding, Figure 5 visualizes the data presented in Table 1. As shown in Figure 3, this visualization combines FD and DT as two dimensions of the gaze data. In task 1 *insert* (Figure 5 A), the mean DT for all task-relevant UI features are longer for D2. The bag lines show a strong category 1 pattern, while the other two UI features show little to no shift for the mean FD. For task 2 *connect* (Figure 5 B), three UI features show a strong category 4 pattern, while the bag lines show mainly shorter mean DTs, but only slightly longer mean FD, thus showing a weak category 2 pattern. For task 6 *disconnect* (Figure 5 C), the UI elements located inside the device in this task show a strong category 2 pattern, while the lever on the outside of the device shows a strong category 1 pattern.



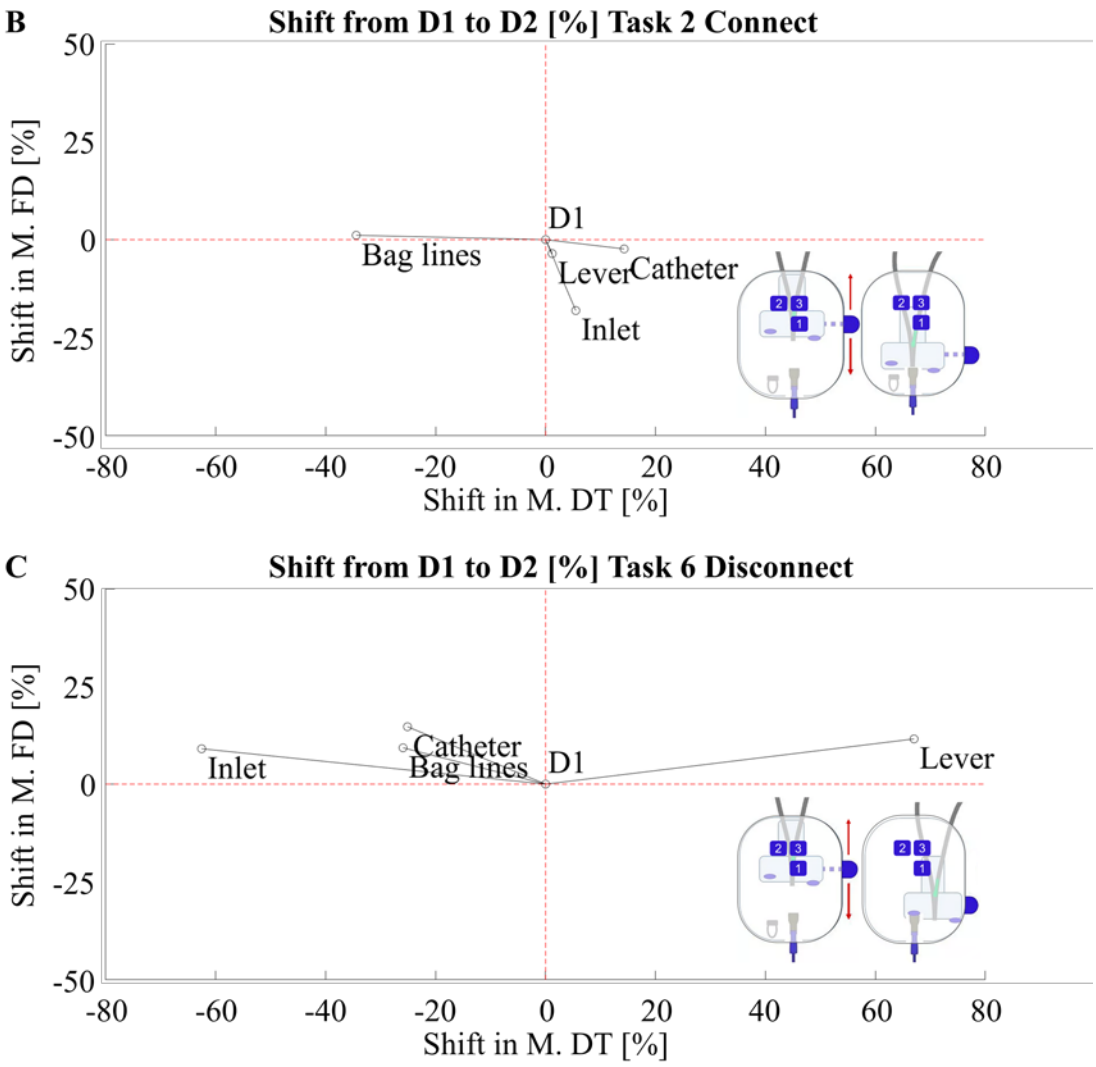


Figure 5. Shifts from D1 (in the coordinate origin) to D2 in terms of mean fixation duration (FD) (ordinate) and mean dwell time (DT) (abscissa) for the task 1 insert (A), task 2 connect (B) and task 6 disconnect (C). The relevant UI features in these three tasks are *bag lines*, *inlet*, *catheter* and *lever*.

Discussion

Task performance analysis generally showed little or no use errors in the various handling tasks for both UI designs (see Figure 4). The tasks with observed use errors were the insertion of the material, the connection and disconnection of the bag lines and the catheter. In the first task, most use errors occurred when inserting the bag lines into the inlet. For this task, Figure 5 A shows a category 1 pattern with longer mean DT and longer mean FD for the bag lines. Therefore, the results of gaze data analysis are consistent with the results of task performance. Gaze data shows more scrutinizing for D2 compared to D1in order to insert the bag lines and catheter into inlet and device. The longer and higher depth in visual perception indicates a higher mental workload for this task using D2.

When connecting and disconnecting the catheter, some participants missed to pull down the lever to connect the bag lines and the catheter again and to put a new cap on the catheter.

For connecting and disconnecting bag lines and catheter, the most important interface features show category 4 patterns (task 2, Figure 5B) and category 2 patterns (task 6, Figure 5C). This visual pattern indicates more skimming behaviour for task 2 and less skimming behaviour for task 6. This in turn indicates more visual controls when connecting the bag lines and the catheter in task 2 for D2. Compared to the task performance, this seems to result in slightly fewer use errors for D2 (2% vs. 4%). For task 6, the results indicate less visual search associated with the relevant features inlet and catheter for D2 when a new cap is placed onto the catheter. In a comparison of the two UI designs, the main difference between D1 and D2 is the position of the top window. With D2, the user can better see the inlet. This may help finding the important features while a new cap is placed on the catheter. Furthermore, the lever in task 6 shows a category 1 pattern associated with a longer and higher depth in visual perception for D2. Although the results show less use errors, handling the lever with D2 appears to be mentally more difficult than with D1.

When evaluating the total mental workload of the medical device system, the analysed UI features of the medical device showed shifts both in the mean FD and in the mean DT. On the one hand, the mean FD varied from 149 to 405 milliseconds in the critical tasks across all features (see Table 1). In a case study of a driving situation described by [32], the values for the mean FD were between 499 and 543 milliseconds. [31] reported in an evaluation of drug label designs that the FD varied between 260 and 392 milliseconds. [33] observed a mean FD of 477 milliseconds while reading a scientific text. Compared to these studies, the mean FD of the handling cycle is in the same range as reading a drug label. On the other hand, the mean DT in the critical tasks varied in a range from less than one second to 28.3 seconds (see Table 1). Especially in the first task, the insertion of the material in both UI designs required longer DT for bag lines and catheter compared to other tasks. This shows that this task requires special attention from the user.

Based on the results of this study, benchmarking D1 and D2 showed the following. Inserting the material seem to be challenging for both UI designs in general. Therefore, the guiding material (manual and quick starting guide) and training needs focusing on this task. The lever of D1 seems to result in lower mental workload. It has a more dominant appearance, compared to D2, where the lever is integrated into the housing for protection in case of a fall. The UI design D2 of the inlet seems to be easier to perceive visually. The higher position of the top window in D2 shows a positive impact on the task connecting and disconnecting bag lines and catheter.

The analysis of two dimensions of visual perception using eye tracking provided a detailed picture of the length and depth of the visual perception and therefore of the challenges in the user cognition and ease of use. The results highlighted the differences in information extraction for different UI features in single tasks. This information helped human factor engineering to focus the development on the critical UI features. Following this work, a summative study evaluated the final UI of the device. This final design and the instructions contained the results of this study, such as the detailed description of the insertion of the material and the colouring of the main UI features to guide the user's gaze. The summative study included a total of 48 patients, relatives, nurses and physicians and confirmed the safety, efficiency and effectiveness in use.

Due to the novelty of the medical device presented in this study, there are several limitations regarding the results. First, the participants were not patients in the real therapy. They were beginners who had no experience in this specific therapy or the associated tasks.

Furthermore, the device was not used in the real therapy application, but in a simulation. These factors provide information on how forgetfulness or even dementia influence the use of the medical device. Second, when the medical device is used at a later date, individual training of the user is mandatory and labelling material supports the user. This support was not provided in the present study. Instead, a presentation with an additional low-level representation of the user interface and a neutral text of the seven tasks guided the participants through the handling cycle. Consequently, the focus was on intuitive task performance and the perception of information depending on the different UI designs. Third, the design of the two different top-level designs was similar due to a unicolored representation. This is not a strong contrast between the main UI functions and the rest of the medical device. As stated in the methods, this was chosen to eliminate the influences of different colouring as an additional influencing factor. At the level of gaze data analysis, the representation of the combination of mean FD and mean DT is the first published. Further research is needed to assess whether the identified patterns apply to different usability studies with different tasks and stimuli.

Conclusion

Studying the impact on the usability of alternatives of different UI designs is crucial to understand, which best supports the user. Traditional methods such as observation, interviews or questionnaires tend to give feedback only at the level of the user interface as a whole. Furthermore, when it comes to reporting usability issues or first impressions of the medical device during interviews or questionnaires, several challenges arise. Test participants may forget to report their impressions, or adapt their answers to social expectations [28, 36]. This makes it difficult to identify the root causes of usability problems and thus the necessary changes in UI design. In alignment with [27] and [28], this study showed that mobile eye tracking provides objective quantitative results based on physiological measurements related to individual UI features. These results can be used to evaluate usability in much more detail compared to traditional methods.

This information is crucial in order to be able to adapt the design of a product to the needs of the users. Therefore, the results of usability testing must be more detailed than just a yes-or-no result of use errors. On the contrary, the evaluation of each individual feature of the user interface promises to achieve the best possible UI design by combining the best features found. This combined solution would therefore offer the highest level of usability. In this way, manufacturers can develop products, which can be used even by untrained users without prior knowledge. This would allow home care to be provided not only by highly qualified nurses and caregivers, but also by patients themselves, partners, children or neighbours. This would contribute to removing barriers to home care and thus to a higher quality of life and normalisation of everyday life, which is less dominated by illness for patients.

Acknowledgements

The authors express their gratitude to all participants of the handling study and to the administration team of the retirement home Limmat in Zurich, Switzerland, for their support. The study was reviewed by the Swiss Ethics Committees on research involving humans (Req-2017-00832). It was assessed to not fall under the Human Research Act (HRA).

Conflicts of Interest

Sandra Neumann and Jean-Claude Gröbli are employed by Peripal AG. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Abbreviations

FD: Fixation duration
DT: Dwell time
M: Mean
PD: Peritoneal dialysis
UI: User interface

References

1. Rygh E, Arild E, Johnsen E, Rumpsfeld M. Choosing to live with home dialysis-patients' experiences and potential for telemedicine support: a qualitative study. BMC Nephrol. 2012; 13:13-20. DOI: 10.1186/1471-2369-13-13
2. Brown EA, Johansson L, Farrington K, Gallagher H, Sensky T, Gordon F, Da Silva-Gane M, Beckett N, Hickson M. Broadening Options for Long-term Dialysis in the Elderly (BOLDE): differences in quality of life on peritoneal dialysis compared to haemodialysis for older patients. Nephrol Dial Transplant. 2010; 25: 3755-3763. DOI: 10.1093/ndt/gfq212
3. Alloatti S, Manes M, Paternoster G, Gaiter AM , Molino A , Rosati C. Peritoneal dialysis compared with hemodialysis in the treatment of end-stage renal disease. J Nephrol. 2000; 13:331-342. PMID:11063136
4. Devoe DJ, Wong B, James MT, Ravani P, Oliver MJ, Barnieh L, Roberts DJ, Pauly R, Manns BJ, Kappel J, Quinn RR. Patient Education and Peritoneal Dialysis Modality Selection: A Systematic Review and Meta-analysis. Am J Kidney Dis. 2016; 68:422-433. DOI: 10.1053/j.ajkd.2016.02.053
5. ERA-EDTA Registry: ERA-EDTA Registry Annual Report 2016. Amsterdam UMC, location AMC, Department of Medical Informatics, Amsterdam, the Netherlands, 2018. ISBN 978-90-817480-9-4
6. Wilson SE, Ouédraogo CT, Prince L, Amadou Ouédraogo A, Sonja Y. Hess SY, Rouamba N, Ouédraogo JB, Vosti SA, Brown KH. Caregiver Recognition of Childhood Diarrhea, Care Seeking Behaviors and Home Treatment Practices in Rural Burkina Faso: A Cross-Sectional Survey. PLoS One. 2012; 7(3):e33273. DOI: 10.1371/journal.pone.0033273
7. Jones R, Jordan S. The Implementation of Crisis Resolution Home Treatment Teams in Wales: Results of the National Survey 2007-2008. Open Nurs J. 2010; 4:9-19. DOI: 10.2174/1874434601004010009
8. Dijkstra NE, Sino CGM, Heerdink ER, Schuurmans MJ. Development of eHOME, a Mobile Instrument for Reporting, Monitoring, and Consulting Drug-Related Problems in Home Care: Human-Centered Design Study. JMIR Hum Factors. 2018; 5(1):e10. DOI: 10.2196/humanfactors.8319

9. Alnosayan N, Chatterjee S, Alluhaidan A, Lee E, Feenstra LH. Design and Usability of a Heart Failure mHealth System: A Pilot Study. JMIR Hum Factors. 2017; 4(1): e9. DOI: 10.2196/humanfactors.6481
10. Beatty AL, Magnusson SL, Fortney JC, Sayre GG, Whooley MA. VA FitHeart, a Mobile App for Cardiac Rehabilitation: Usability Study. JMIR Hum Factors. 2018; 5(1):e3. DOI: 10.2196/humanfactors.8017
11. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, Lei J. Usability Study of Mainstream Wearable Fitness Devices: Feature Analysis and System Usability Scale Evaluation. JMIR Mhealth Uhealth. 2018; 6(11):e11066. DOI: 10.2196/11066
12. George D, Hassali MA, Hss AS. Usability Testing of a Mobile App to Report Medication Errors Anonymously: Mixed-Methods Approach. JMIR Hum Factors 2018; 5(4):e12232. DOI:10.2196/12232)
13. Rotondi AJ, Spring MR, Hanusa BH, Eack SM, Haas GL. Designing eHealth Applications to Reduce Cognitive Effort for Persons With Severe Mental Illness: Page Complexity, Navigation Simplicity, and Comprehensibility. JMIR Hum Factors. 2017; 4(1): e1. DOI: 10.2196/humanfactors.6221
14. Medicines and Healthcare products Regulatory Agency. Human Factors and Usability Engineering – Guidance for Medical Devices Including Drug-device Combination Products. 1st edition. London, United Kingdom: MHRA; 2017.
15. Rosenblum LD. See what I'm saying: The extraordinary powers of our five senses. 1st edition. New York, NY: W.W. Norton; 2010. ISBN: 0393067602
16. Miller GA. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. Psychological Review. 1956; 101(2):81–97. DOI: 10.1037/h0043158
17. Peterson L, Peterson MJ. Short-term retention of individual verbal items. Journal of Experimental Psychology. 1959; 58(3):193-198. DOI: 10.1037/h0049234
18. Wu Y, Cheng J, Kang X. Study of Smart Watch Interface Usability Evaluation Based on Eye-Tracking. A. Marcus, editor. Design, User Experience, and Usability: Technological Contexts. DUXU 2016. Lecture Notes in Computer Science. 2016 July 17-22; Toronto, Canada. Springer, Cham; 2016. DOI: 10.1007/978-3-319-40406-6_10
19. Groen M. and Noyes J. Using eye tracking to evaluate usability of user interfaces: Is it warranted? IFAC Proceedings Volumes. 2010; 43(13):489-493. DOI: 10.3182/20100831-4-FR-2021.00086
20. Wedel M, Pieters R. A Review of Eye-Tracking Research in Marketing. Review of Marketing Research. 2008; 123-147. DOI: 10.1108/S1548-6435(2008)0000004009
21. Chynal P, Sobecki J, Szymanski JM. Remote Usability Evaluation Using Eye Tracking Enhanced with Intelligent Data Analysis. A. Marcus, editor. Design, User Experience, and Usability. Design Philosophy, Methods, and Tools. DUXU 2013. Lecture Notes in Computer Science. 2013 July 21-26; Las Vegas, NV, USA. Springer, Berlin, Heidelberg; 2013. DOI: 10.1007/978-3-642-39229-0_24

22. Harezlak K, Kasprowski P. Application of eye tracking in medicine: A survey, research issues and challenges. Comput Med Imaging Graph. 2018; 65:176-190. DOI: 10.1016/j.compmedimag.2017.04.006
23. Barkana DE, Acik A. Improvement of design of a surgical interface using an eye tracking device. Theor Biol Med Model. 2014; 11:48-65. DOI: 10.1186/1742-4682-11-S1-S4
24. Guo Q, Xue C, Lin Y, Niu Y, Chen M. A Study for Human-Machine Interface Design of Spacecraft Display & Control Device Based on Eye-Tracking Experiments. In: Harris D, editor. Engineering Psychology and Cognitive Ergonomics: Cognition and Design. EPCE 2017. Lecture Notes in Computer Science, 2017 July 9-14; Vancouver, BC, Canada. Springer, Cham; 2017. DOI: 10.1007/978-3-319-58475-1_16
25. Dill ET, Young SD. Analysis of Eye-Tracking Data with Regards to the Complexity of Flight Deck Information Automation and Management – Inattentional Blindness, System State Awareness, and EFB Usage. 15th AIAA Aviation Technology, Integration, and Operations Conference. 2014 June 22-26; Dallas, TX, USA. DOI: 10.2514/6.2015-2901
26. Obrist M, Bernhaupt R, Beck E, Tscheligi M. Focusing on Elderly: An iTV Usability Evaluation Study with Eye-Tracking. Cesar P, Chorianopoulos K, Jensen JF, editors. Interactive TV: a Shared Experience. EuroITV 2007. Lecture Notes in Computer Science. 2007 May 24-25; Amsterdam, The Netherlands. Springer, Berlin, Heidelberg; 2007. DOI: 10.1007/978-3-540-72559-6_8
27. Lohmeyer Q, Schneider A, Jordi C, Lange J, Meboldt M. Toward a new age of patient centricity? The application of eye-tracking to the development of connected self-injection systems. Expert Opin Drug Deliv. 2019; 16(2):163-175. DOI: 10.1080/17425247.2019.1563070
28. Koester T, Brøsted JE, Jakobsen JJ, Malmros HP, Andreasen NK. The Use of Eye-Tracking in Usability Testing of Medical Devices. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care. 2017; 6(1):192–199. DOI: 10.1177/2327857917061042
29. International Electrotechnical Commission. IEC 62366-1 Medical devices - Part 1: Application of usability engineering to medical devices. 1st edition. Geneva, Switzerland: IEC Central Office; 2015. ISBN 978-2-8322-2281-2
30. Godfroid A, Spino LA. Reconceptualizing Reactivity of Think-Alouds and Eye Tracking: Absence of Evidence Is Not Evidence of Absence. Language Learning. 2015; 65(4):896-928. DOI: 10.1111/lang.12136
31. Bojko A, Gaddy C, Lew G, Quinn A, Israelski E. Evaluation of Drug Label Designs Using Eye Tracking. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2005; 49(11):1033–1037. DOI: 10.1177/154193120504901107
32. Velichkovsky BM, Dornhoefer SM, Pannasch S, Unema PJA. Visual Fixations and Level of Attentional Processing. Proceeding ETRA '00 Proceedings of the 2000 symposium on Eye tracking research & applications. 2000 November 06-08; Palm Beach Gardens, Florida, USA. ACM New York, NY, USA; 2000. DOI: 10.1145/355017.355029

33. Just MA, Carpenter PA. A Theory of Reading: From Eye Fixations to Comprehension. Psychological Review. 1980; 87(4):329-354. DOI: 10.1037/0033-295X.87.4.329
34. Kim JH, Zhao X, Du W. Assessing the performance of visual identification tasks using time window-based eye inter-fixation duration. Int J Ind Ergon. 2018; 64:15-22. DOI: 10.1016/j.ergon.2017.09.002
35. Sarter NB, Arbor A, Mumaw RJ, Wickens CD. Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data. Hum Factors. 2007; 49(3):347-357. DOI: 10.1518/001872007X196685
36. Wiklund M, Kendler J, Strochlic A. Usability Testing of Medical Devices. 2nd edition. Boca Raton, FL: CRC Press, Taylor & Francis Group; 2015. ISBN: 9781466595880