Usability testing to evaluate user experience on cyclers for automated peritoneal dialysis

Thore Reitz[1], Stephanie Schwenke[1], Sebastian Hölzle[2] and Adelheid Gauly[2*]

Abstract

Background: In the development of medical devices usability is an important aspect standing alongside performance and safety. Peritoneal dialysis (PD) can be provided by use of automated PD (APD), assisted by a cycler performing the solution exchanges. The present study has been executed to simulate training on APD cyclers to evaluate learnability and usability through established questionnaires.

Methods: Usability of two APD cyclers (*sleep•safe harmony*, Fresenius Medical Care, Bad Homburg, Germany and HomeChoice Pro, Baxter International Inc., Deerfield (IL), USA) were evaluated with the User Experience Questionnaire (UEQ), the NASA TLX Questionnaire, and the System Usability Scale (SUS), both after training and after experience sessions.

Results: Lay persons ($n = 10$) and health care personnel (HCPs, $n = 11$) participated in the study. The respondents consistently gave positive ratings in the UEQ after training and experience session. The ratings from the NASA TLX Questionnaire were mostly below 50 points indicating a low workload. Lay users and HCPs gave high ratings in the SUS evaluation both after the training and experience sessions confirming a good learnability and usability of the devices.

Conclusions: The usability study to assess learnability and use-related safety revealed consistent results with all applied instruments, which demonstrated good learnability and ease-of-use of the studied APD cyclers.

Keywords: Usability, Learnability, Automated peritoneal dialysis, Cycler, Questionnaires

Background

Peritoneal dialysis is a home therapy, performed by the patient, either as continuous ambulatory peritoneal dialysis (CAPD) with manually performed solution exchanges, or as automated peritoneal dialysis (APD) assisted by a cycler to automatically perform solution exchanges, mostly during the night, while the patient is sleeping. Typical APD cyclers contain a set of features, firstly a mechanism to transport the peritoneal dialysis (PD) fluid in and out of the peritoneal cavity, structures to fix the tubing systems connected to the solution bags in the cycler, and a heating unit to ensure solutions are instilled at body temperature. Further, there is a control unit to program and control the therapy; nowadays, this occurs mostly via a touch screen. Many cyclers also contain a slot for a patient card to record treatment data or a LAN interface to transfer treatment data to the dialysis center. The patient-cycler interaction occurs predominantly in the set-up and dismantling of the disposables in the cycler and the treatment control at the touch screen of the cycler [1].

---

* Correspondence: Adelheid.gauly@fmc-ag.com
[2]Global Medical Office, Fresenius Medical Care Deutschland GmbH, Else-Kröner-Strasse 1, 61352 Bad Homburg, Germany
Full list of author information is available at the end of the article

Reitz *et al. Renal Replacement Therapy*   (2021) 7:20

Page 2 of 8

The interaction of a medical device and the user is an essential focus in the development process. While a broad range of technical functions is incorporated to ensure the device performs as intended and is safe, the user, i.e., a person interacting and operating the device is a further "component" to be considered. This is taken into account by human factors or usability engineering, which aims at "understanding how people interact with technology and studying how interface design affects the interactions people have with technology" [2].

Usability engineering sets its sights on developing a medical device that performs safely, as well as effectively and efficiently and that users are happy to use. It focuses on optimizing the user interface to facilitate training and learnability and to minimize the risk of misuse and associated hazards. It takes into account numerous aspects related to the intended users, the environment in which the device is to be used and the device's user interface [2, 3]. Usability testing is performed during the development process to evaluate use-related safety of the medical device but should also be pursued after market introduction with a focus on user experience with the product [4].

The present study was undertaken to address specifically usability and ease-of-use of the APD cycler *sleep•safe harmony* (Fresenius Medical Care, Bad Homburg, Germany) in comparison to another APD cycler in broad clinical use (HomeChoice Pro, Baxter International Inc., Deerfield (IL), USA), employing instruments commonly applied in collecting user and operational experience.

Methods

Participant profile
The user profile for an APD cycler includes both health care professionals (HCP), mostly nurses, as well as patients with kidney failure. In the present study, nurses who work in nephrology and/or dialysis and thus have some background on the concept of dialysis, but who were not directly familiar with either APD or any APD system, were recruited. To avoid any potential bias, the group of patients was represented by lay users who do currently not use dialysis, who were not foreseen to be treated with dialysis in the near future, and who had no experience with any APD device. To represent the user profile to the greatest extent possible, lay users were targeted to have an age between 50 and 80 years. They were also screened for technical affinity and empathy. The former was screened using an adapted version of a questionnaire on technical affinity [5] — the purpose of this screening was to try to avoid recruiting participants who either disliked technology so strongly or understood it so poorly that a physician would never prescribe APD to them. Empathy in this case referred to an ability to

put oneself in someone else's shoes and to imagine a situation that vividly; this characteristic was important, because participants did not have renal failure or the symptoms that went along with it and that could influence their use of the device. A modified version of the questionnaire on Interpersonal Reactivity Index was used to screen for empathy [6]. Lay users had to have no major visual impairment (unless manageable with visual aids).

Both HCPs and lay users had to confirm in writing their willingness to participate in the study and to be audio- and video-recorded during the test.

Study design and procedure
The study was performed by Use-Lab GmbH (Steinfurt, Germany), a company specialized in conducting usability studies. To allow both home and clinical settings to be simulated, the study was performed in Use-Lab's simulation lab, which provided a sufficiently realistic environment and the possibility to video- and audio-record test sessions.

The study consisted of a training session on both APD devices, and of the experience session scheduled for the subsequent day to start approx. 24 h after start of training. The order in which the two systems were trained and tested alternated between participants but was identical for training and experience session. The training session included general information on peritoneal dialysis and training on the devices. The trainer used a predefined script and method, with the essential information included in a PowerPoint presentation to perform a "guided exploration", i.e., a certain step was explained to the participant and then they were asked to carry it out themselves.

For the study, a *sleep•safe harmony* APD cycler, software version 2.2, and the associated instructions for use (IFU) were used. Regarding user interface, this software version had no significant differences compared to the latest released version 3.0. The HomeChoice Pro APD cycler was used with software version 10.4 and the associated IFU. Further, a dialysis solution bag designed for the respective cycler, filled with 1000 ml tap water, was used, a drainage bag and, in case of *sleep•safe harmony*, an organizer, a tool which is used as an aid to facilitate aseptic handling. To better simulate the situation and help participants engage in their roles, lay users were asked to wear a "dummy tummy" consisting of a patient catheter and an empty 500-ml bag (the peritoneal cavity); during sessions with HCPs, the moderator wore the dummy tummy and assumed the role of the patient.

The major tasks which were trained and tested were (i) all steps to prepare the system for the therapy until begin of the initial drain: turning the system on, placing all disposables in the device, connecting oneself or the

Reitz *et al. Renal Replacement Therapy*   (2021) 7:20

Page 3 of 8

"patient", (ii) skipping the initial drain, and (iii) ending the therapy prematurely and performing all steps to terminate the therapy and prepare it for its next use. No programming of treatment schedules was trained, i.e. participants worked on predefined settings. Although importance of antiseptic working and the risk of infections was part of the training, participants did not have to adhere to hygienic standards.

A questionnaire on basic demographic data of the participant was completed before the training. After completion of the training session, participants completed several questionnaires, as described below.

For the experience session on the second day, the participants were asked to perform the tasks they had been trained on the previous day. Participants were observed as they performed their tasks and were permitted to use the IFU at any time; actions, including IFU use, were recorded. After the experience session, the participants were asked to complete a set of questionnaires, similar to those given after training.

Participants had individual training and experience sessions; all sessions and documents were present in German.

Objective measurements

Time recordings exact to the second were derived from video recordings performed during training and experience session.

To capture difficulties in executing device operations and thereby to obtain a further measure on the ease-of-use, all mistakes and difficulties observed during the experience sessions were captured. Steps were considered successfully completed only if they were completed without any observed difficulties; requesting assistance and using the IFU were considered difficulties.

Subjective measurements
User Experience Questionnaire (UEQ)

The UEQ allows quick assessment of user experience on products with a product-user interface [7]. Respondents are presented a total of 26 pairs of words that represent opposite polls of a scale and are asked to rate their experience with the device on a seven-step scale between these word pairs. The groups of word pairs are allocated to three dimensions: use quality, which includes three subscales (dependability, perspicuity and efficiency), design quality, including two subscales (stimulation, novelty), and attractiveness; those filling out the questionnaire, are not made aware of the dimensions.

NASA TLX Questionnaire

The National Aeronautics and Space Administration Task Load Index (NASA TLX) questionnaire is a tool to assess workload of a specific task as subjectively perceived by the individual filling it out [8]. The questionnaire consists of six dimensions: mental demand, physical demand, temporal demand, effort, performance, and frustration level. Each of the six dimensions is rated with a score from 0 to 100 in steps of five, with 0 being the lowest, and 100 being the highest perceived task workload.

System Usability Scale (SUS)

The SUS consists of ten five-point Likert-scale items anchored at either end with disagree or agree. The questions address how learnable and how usable a system is. The answers result in a system usability score represented by scores between 0 and 100, with 0 for the least and 100 for the best perceived usability [9].

Statements

Furthermore, participants were presented several statements and were asked to rate their agreement with these statements on five-point Likert scales anchored at either end with "I agree" and "I disagree". These statements addressed how easy it has been to use the system and find necessary information and how satisfied participants were with the system overall (see Additional File 1).

Sample size considerations and analysis

The goal of participant recruitment was 12 persons in both, the group of HCPs and of lay users. This allowed to execute the study with a pilot participant to adjust, if necessary, procedures, and/or to allow a drop-out, with the aim to include at least 10 participants each in the analysis. According to Faulkner, a sample size of 10 in a usability test allows to identify at least 82% of total known usability problems, and to find in average 95% of total known usability problems [10].

All questionnaires were analyzed descriptively with frequency, mean ± SD, or median, as appropriate. To test for statistical significance a Wilcoxon signed rank-test was used and the probability level of 95% was applied [11]. The statistical analysis was performed using the statistical programming language R using R Statistical Software (https://www.R-project.org/).

Results
Participants

Data from ten lay users and eleven HCPs was evaluated. Data from two additional lay users was excluded due to changes to the protocol and abnormalities during an experience session.

Data on demographics, profession and, for the lay users, technical affinity and empathy are given in Table 1. None of the lay users was employed in the healthcare sector, HCPs were associated with nephrology or dialysis, but

Table 1 Demographic data of participants — lay users and health care personnel (HCP)

	Lay users N = 10	HCPs N = 11
Age (years, mean, range)	63 (52–75)	40 (24–52)
Male/female (n)	7/3	2/9
Technical affinity (mean, range)[a]	4.1 (3.6–4.7)	n/a
Empathy (mean, range)[b]	3.9 (3.3–4.2)	n/a
Professional work experience (years, mean, range)	24.7 (6–45)	18.5 (2–30)
Profession (n)	Technical profession (4)	Nurse (8)
	Clerk (2)	Diabetes advisor (1)
	Craftsman (1)	Medical assistant (2)
	Pensioner (3)	

[a]Technical affinity as evaluated with a questionnaire as given in [5]
[b]Empathy as screened with the Interpersonal Reactivity Index [6]
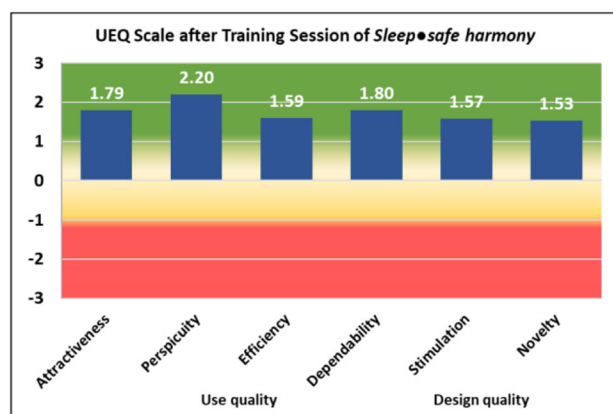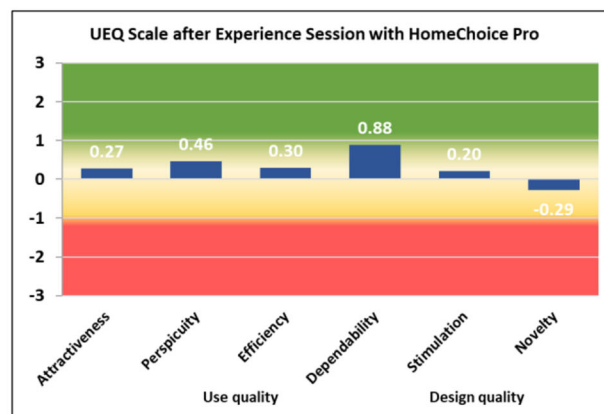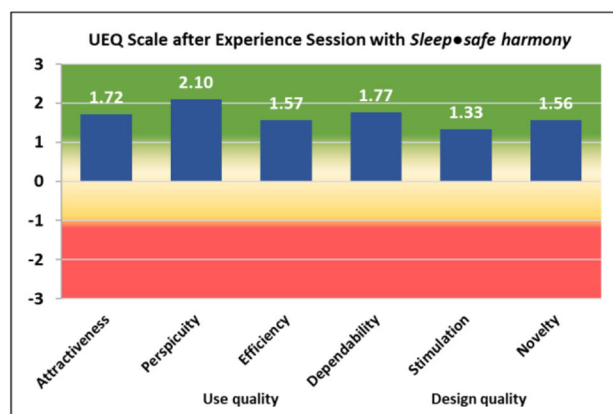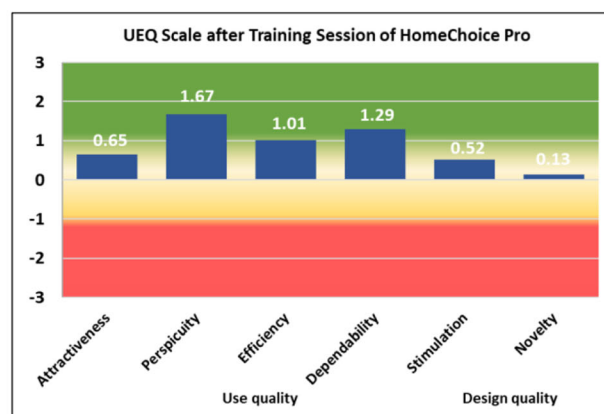
Fig. 1 User Experience Questionnaire (UEQ) dimensions after training and experience session with *sleep•safe harmony* (**a**) and HomeChoice Pro cyclers (**b**)

none of them had working experience with peritoneal dialysis.

Operation time and difficulties

The mean time to execute the test tasks with *sleep•safe harmony* was 25:36 ± 3:11 min for the lay users and 21:36 ± 3:28 min for the HCPs, and for HomeChoice Pro it was 24:24 ± 3:19 min for the lay users and 23:30 ± 4:10 min for the HCPs.

In total, participants were asked to complete nine individual tasks with each APD system, leading per system to a total of 189 tasks observed (90 for lay user, 99 for HCPs). The cumulative number of operational difficulties, i.e., failure to complete defined steps adequately during all experience sessions was 38 for 10 lay users, 28 for 11 HCPs with the *sleep•safe harmony*, and 35 and 40 for lay users and HCPs with the HomeChoice Pro, respectively.

The difference in time needed showed a correlation to the number of difficulties encountered with the *sleep•safe harmony* ($r = 0.79$; $p < 0.001$), but not with the HomeChoice Pro ($r = 0.28$, $p = 0.238$).

User Experience Questionnaire

Figure 1 shows the results of the UEQ after training and after the experience session. Scores in the red region reflect more negative responses, while those in the green region reflect more positive responses. All ratings on user experience with the *sleep•safe harmony* are in the green region, representing a perception of positive attributes both after training and after experience session. Overall, there were no major changes in the ratings from training to experience session. With the HomeChoice Pro cycler, the ratings are in the yellow/green region for training and mostly in the yellow region after the experience session.

NASA TLX Questionnaire

The NASA TLX Questionnaire measures several dimensions of workload, in this case with regards to a workflow with a device, as perceived by the device's user. Figure 2 shows the distribution of ratings for the six dimensions given by both groups, lay users and HCPs. Only few exceptions rated the task load in any dimension higher than 50 points.

Comparing the perceived workload after training and after experience session shows nearly similar ratings with the *sleep•safe harmony*, some participants rated the perceived workload higher after experience than after training. With HomeChoice Pro, the perceived workload showed increases in some categories from training to experience session.
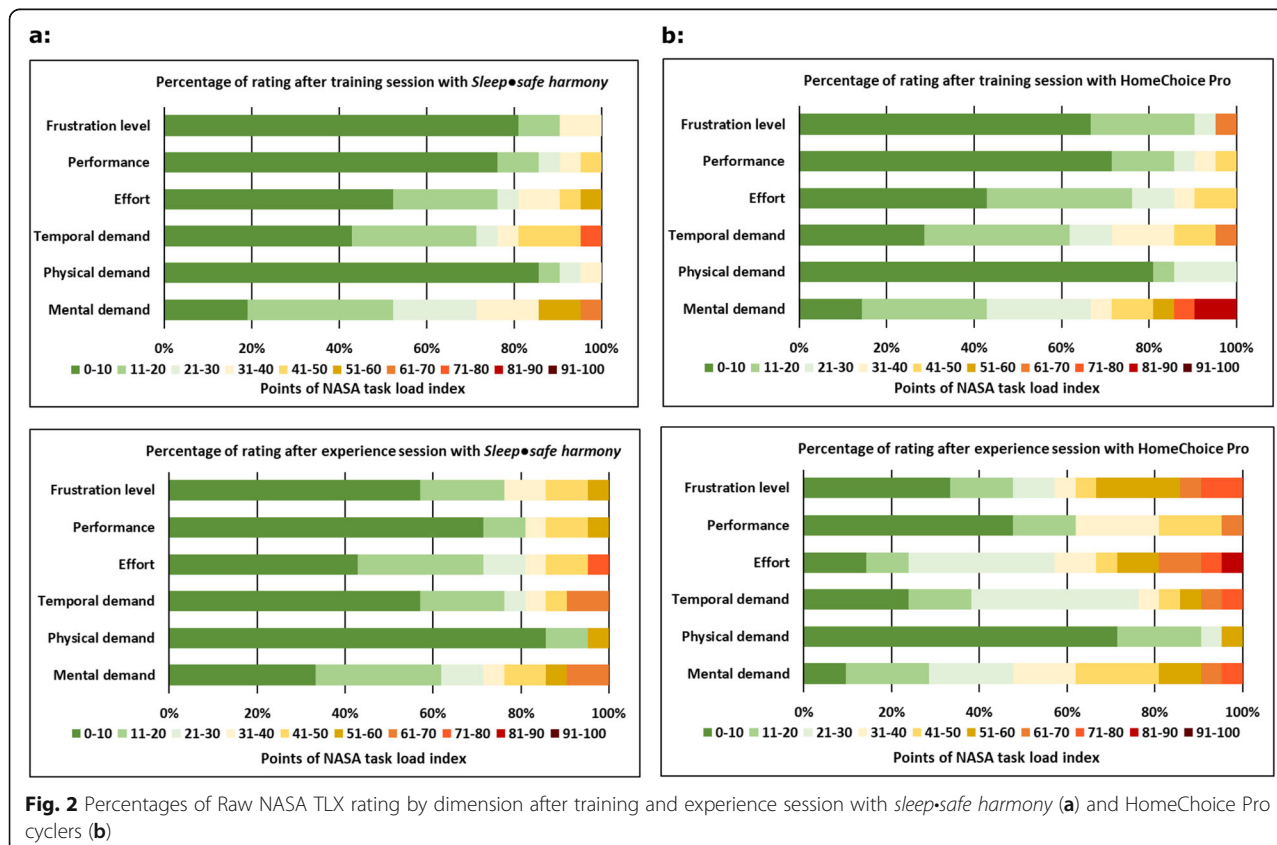
Fig. 2 Percentages of Raw NASA TLX rating by dimension after training and experience session with *sleep•safe harmony* (**a**) and HomeChoice Pro cyclers (**b**)

System Usability Scale (SUS)

Overall, the SUS score for the *sleep•safe harmony* APD cycler was high reaching a median of 87.5 points on a scale from 0 to 100 where 100 represents the greatest subjective usability. Comparing the individual ratings given after training and after experience, no significant change could be observed (Fig. 3; $p = 0.122$). With HomeChoice Pro the ratings after training with a median of 77.5 decreased to a median of 55 after test for the entire cohort (Fig. 3; $p < 0.001$).

Ratings that were more than 30 points lower after the experience session were observed with *sleep•safe harmony* for three HCPs, with HomeChoice Pro for two HCPs and two lay users. There was no apparent relation to professional background or work experience, except in one case of an HCP with only 2 years of professional experience showing such lower ratings after the experience session with both APD cyclers.

Statements

Participants were given a few statements on the learnability and usability of the APD cyclers and asked to rate their agreement with them. Most participants agreed or somewhat agreed with the concluding statement, indicating their satisfaction with the devices: with the *sleep•safe harmony* system, after both training (16 agree; 5 somewhat agree) and experience (17 agree; 3 somewhat agree) sessions, and with the HomeChoice Pro as well after both training (10 agree; 7 somewhat agree) and experience (5 agree; 10 somewhat agree) sessions.

Discussion

The usability study presented here employed multiple instruments to assess the subjective ease-of-use and general usability of two APD cyclers. Results after simulated use are satisfactory.

The participants were divided into two groups: Lay users representing patients who had to learn to use such devices and healthcare professionals, primarily nurses who use the cycler for in-hospital treatments and who train patients to use the device. Both groups can be considered characteristic of their peer group, with the lay users' age between 52 and 72 years, which is consistent with the age of patients starting kidney replacement therapy [12], and the HCPs with a wider range of age and years of professional experience. Although for the lay user group it would have been conceivable to recruit patients with chronic kidney disease, who might be confronted with kidney replacement therapy in the near future to take psychological and pathophysiological effects of the disease into account, a simulation study is justifiable as persons never confronted with kidney replacement therapies are less biased than patients who might have dealt with it to a variable extent. Furthermore, ethical considerations strongly favor recruiting individuals who are not actual patients to avoid causing further distress. Simulation studies are also an accepted instrument for regulatory purposes [2].

The operation time and operational difficulties were expectedly lower for HCPs than for lay users. Although HCPs had no experience with automated PD and had never been trained on an APD cycler, they were nevertheless familiar with technical equipment commonly used in hospital and patient care, which may facilitate learning and speed of operation.

The User Experience Questionnaire, in which respondents could attribute positive or negative words to the study object resulted in consistently positive ratings of the *sleep•safe harmony* APD cycler both after the training and after the experience session. The fact that the rating did not decrease from training to test is an indicator that the training was effective, and learnability and user experience with the system were good. With the HomeChoice Pro cycler, lower ratings in most categories were observed after the experience than after the training session. Comparing the results to a benchmark study including UEQ responses from a broad variety of mostly digital applications, the mean rating given to the

Fig. 3 Subjective Usability Score (SUS) (median) after training and experience session of the *sleep•safe harmony* and the *HomeChoice Pro* cyclers
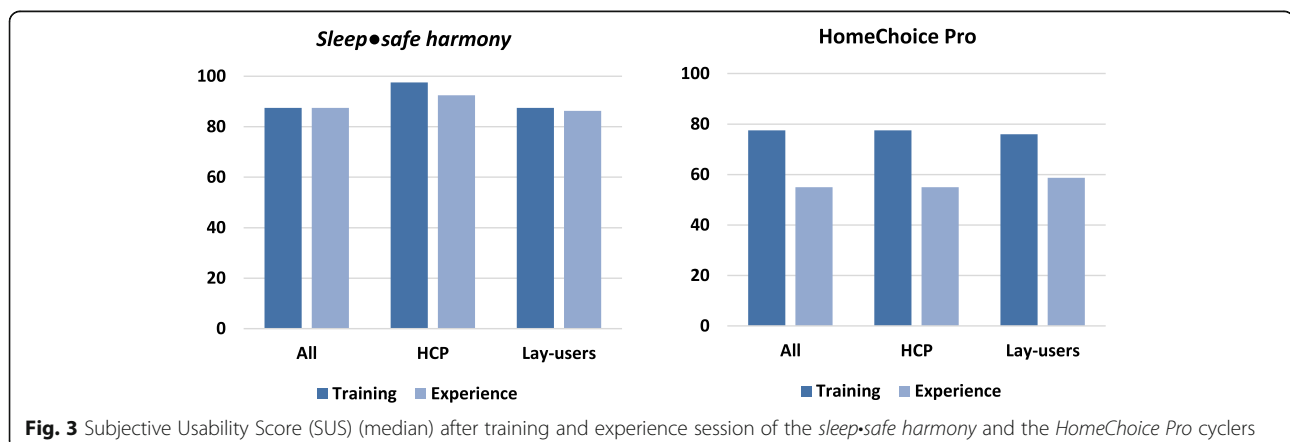
sleep•safe harmony device was for all categories in the "good" or mostly the "excellent" range, for the Home-Choice Pro this was more heterogeneous and would have fallen in the categories "good," "above average," "below average," and "bad" [13].

The NASA TLX Questionnaire showed that both HCPs and lay users perceived low task load using the *sleep•safe harmony* APD cycler, which was in most cases unchanged between training and test. For the Home-Choice Pro again, more heterogeneous and higher ratings of task load after the experience session were observed. Still for both cyclers, ratings indicating a high task load were too few to derive associations, e.g., to participant characteristics, such as age or professional experience.

Although a benchmark is lacking, the ratings mostly below 50 points indicate to an overall low workload. Compared to a similar analysis of a medical device in the field of peritoneal dialysis, where about one third of ratings were higher than 50 points [14], the distribution of ratings in our study was rather in the lower range for both devices.

As the participants were only trained once and had only one opportunity to apply their knowledge on the device, higher ratings after test were not unexpected and would have been expected to decrease with repeated application of the device. Usually, the APD devices are trained for several days or training sessions [15] allowing the patient to become increasingly familiar with the device operation.

The System Usability Scale is an instrument developed in 1996 and includes 8 items on usability and 2 items on learnability [9]. The SUS is a questionnaire provided to the test user immediately after use without, e.g., an interview in between, thus catching the most intuitive impression on the device [16]. Again, the level of the SUS score as such and the consistency of feedback given after training and after experience session confirms a good learnability and usability of the *sleep•safe harmony* device, whereas lower ratings with HomeChoice Pro after the experience than after the training session indicate to the possible necessity to deepen learnings by repeated application, as it is usually common practice.

The learnability and ease-of-use as assessed with the instruments used in this study could also be confirmed in direct questions to the participants on agreement/disagreement with certain statements, such as overall satisfaction with the system. Although these statements were given by persons currently not confronted with the disease and the respective treatment option using such an APD cycler, this supports the systems being easy-to-use for actual patients as well.

The study has certain limitations. With the simulation approach, we missed the psychological and pathophysiological effects of chronic kidney disease, which might impact cognitive and physical functions [17]. Furthermore, the devices were tested in a simulated rather than in its actual setting. Training and experience sessions did not encompass a full treatment, where other aspects, e.g., handling of unforeseen alarms need to be trained. A second experience session would have been desirable to allow the participant further learning from application and possible difficulties. Although a potential bias cannot be completely ruled out, as there may have been different levels of knowledge about the two devices, this was contained as much as possible by commissioning an external provider independent from the manufacturer to recruit participants and to execute the training and experience sessions. Over time, updated versions of both devices may have been or will be introduced that take usability into account, along with other features.

Conclusions

The study showed that systematic usability analysis is a useful and necessary instrument in the development and life cycle of a medical device in general and that the *sleep•safe harmony* APD cycler proved to be an easy-to-use system for participants. Usability studies performed with the target users and use environment of the device are warranted.

Abbreviations
APD: Automated peritoneal dialysis; CAPD: Continuous ambulatory peritoneal dialysis; HCP: Healthcare personnel; IFU: Instructions for Use; LAN: Local area network; NASA TLX: National Aeronautics and Space Administration Task Load Index; PD: Peritoneal dialysis; UEQ: User Experience Questionnaire; SUS: System Usability Scale

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s41100-021-00340-0.

Additional file 1:. Qualitative statements on learnability and operation of the APD cyclers.

Reitz *et al. Renal Replacement Therapy*   (2021) 7:20

Page 8 of 8

Declarations

Ethics approval and consent to participate
Because this usability study simulating the use of a CE certified medical device was not a clinical investigation on performance and safety of a medical device and in the absence of diagnostic or therapeutic interventions, consultation with an ethics committee was not considered. Both HCPs and lay users had to confirm in writing their willingness to participate in the usability study and to be audio- and video-recorded during the test.

Consent for publication
Not applicable

Competing interests
TR and SS are fulltime employees of Use-Lab GmbH, and SH and AG are full-time employees of Fresenius Medical Care.

Author details
[1]Use-Lab GmbH, Steinfurt, Germany. [2]Global Medical Office, Fresenius Medical Care Deutschland GmbH, Else-Kröner-Strasse 1, 61352 Bad Homburg, Germany.

References
1. Ronco C, Amerling R, Dell'Aquila R, Rodighiero MP, Di Loreto P. Evolution of technology for automated peritoneal dialysis. Contributions Nephrol. 2006;150:291–309. https://doi.org/10.1159/000093624.
2. Applying human factors and usability engineering to medical devices. Guidance for industry and food administration staff. USA: Food and Drug Administration; 2016.
3. International Organization for Standardization. Ergonomics of human-system interaction - Part 110: Interaction principles. ISO/DIS 9241-110:2020, 2020.
4. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on Medical Devices. Off J Europ Union. 2017:60: L117.
5. Karrer K, Glaser C, Clemens C, Bruder C. Technikaffinität erfassen–der Fragebogen TA-EG. Der Mensch im Mittelpunkt technischer Systeme. 2009;8:196–201.
6. Paulus C. Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. 2009; http://psydok.sulb.uni-saarland.de/volltexte/2009/2363/. Accessed on 26 June 2020.
7. Laugwitz B, Held T, Schrepp M. Construction and evaluation of a user experience questionnaire. In: Holzinger A, editor. HCI and usability for education and work. Berlin: Springer Berlin Heidelberg; 2008. p. 63–76.
8. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Advanc Psychol. 1988;52. Elsevier:139–83.
9. Brooke J. SUS-A quick and dirty usability scale. Usability Evaluation Industry. 1996;189(194):4–7.
10. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behavior Res Methods Instruments Computers. 2003;35(3):379–83. https://doi.org/10.3758/BF03195514.
11. Wilcoxon F. Individual comparisons of grouped data by ranking methods. J Econ Entomol. 1946;39(2):269–70. https://doi.org/10.1093/jee/39.2.269.
12. ERA-EDTA Registry Annual Report 2017. Amsterdam UMC, Location AMC, Department of Medical Informatics, Amsterdam, the Netherlands, 2019.
13. Schrepp M, Hinderks A, Thomaschewski J. Construction of a benchmark for ther User Experience Questionnaire (UEQ). Int J Interactive Multimedia Artificial Intelligence. 2017;4(4):40–4. https://doi.org/10.9781/ijimai.2017.445.
14. Hess S, Dubach M, Meboldt M, Foggensteiner L. Evaluating patient safety and ease of use of a novel connection-assist device for peritoneal dialysis. Patient Preference Adherence. 2019;13:1785–90. https://doi.org/10.2147/PPA.S218663.
15. Sharma S, Kattamanchi S, Gonzales MG, Sloand JA, Uribarri J. Patient-centric user-interface in automated peritoneal dialysis: impact on training and outcomes at a single center. Blood purification. 2019;48(2):138–41. https://doi.org/10.1159/000495341.
16. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Human–Computer Interaction. 2008;24(6):574–94. https://doi.org/10.1080/10447310802205776.
17. Weiner DE, Seliger SL. Cognitive and physical function in chronic kidney disease. Curr Opinion Nephrol Hypertension. 2014;23(3):291–7. https://doi.org/10.1097/01.mnh.0000444821.87873.7b.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.