Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems

Aiyegbusi, Olalekan Lee

DOI:
DOI: 10.1007/s11136-019-02329-z

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Aiyegbusi, OL 2019, 'Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems', *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. https://doi.org/DOI: 10.1007/s11136-019-02329-z

Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems

Olalekan Lee Aiyegbusi

Quality of Life Research
An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation - An Official Journal of the International Society of Quality of Life Research

ISSN 0962-9343

Qual Life Res
DOI 10.1007/s11136-019-02329-z

Springer

Quality of Life Research
https://doi.org/10.1007/s11136-019-02329-z

REVIEW

Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems

Olalekan Lee Aiyegbusi[1]

Accepted: 3 October 2019

Abstract

Introduction Recent advances in information technology and improved access to the internet have led to a rapid increase in the adoption and ownership of electronic devices such as touch screen smartphones and tablet computers. This has also led to a renewed interest in the field of digital health also referred to as telehealth or electronic health (eHealth). There is now a drive to collect these PROs electronically using ePRO systems.

Method However, the user interfaces of ePRO systems need to be adequately assessed to ensure they are not only fit for purpose but also acceptable to patients who are the end users. Usability testing is a technique that involves the testing of systems, products or websites with participants drawn from the target population. Usability testing can assist ePRO developers in the evaluation of ePRO user interface. The complexity of ePRO systems; stage of development; metrics to measure; and the use of scenarios, moderators and appropriate sample sizes are key methodological issues to consider when planning usability tests.

Conclusion The findings from usability testing may facilitate the improvement of ePRO systems making them more usable and acceptable to end users. This may in turn improve the adoption of ePRO systems post-implementation. This article highlights the key methodological issues to consider and address when planning usability testing of ePRO systems.

Keywords Usability testing · Electronic patient-reported outcomes · PROs · ePROs · ePRO systems · ePROM · Digital health · eHealth · Telehealth · Electronic systems

Introduction

Recent advances in information technology and improved access to the internet have led to a rapid increase in the adoption and ownership of electronic devices such as touch screen smartphones and tablet computers. In 2017, about 77% of American adults reported owning a smartphone compared to 35% in 2011 [1]. The increase in ownership of electronic devices has also been observed worldwide albeit at a lower rate in developing countries [1] and the digital divide between younger and older populations has narrowed over the past decade [2].

These developments have in turn led to an upsurge of interest in digital healthcare also known as telehealth. It is now feasible to remotely collect patient-reported outcomes (PROs) using electronic devices. A PRO can be defined as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" [3]. An ePRO is therefore a PRO that is collected electronically. In the past, PROs were mainly collected using paper formats which were associated with significant administrative burden, missing data and data entry errors.

EPROs are increasingly used in clinical trials and cohort studies to appraise, from a patient perspective, the effectiveness and safety of interventions [4]. This is important as regulatory authorities are now paying greater attention to PRO data when making decisions about drug approvals [5–7]. The use of ePROs instead of paper formats in clinical trials could facilitate the robust analysis and reporting of PRO data which is often neglected or inadequate by making the data available in easily exportable formats with fewer errors and missing data [8].

Clinicians are now able to use interactive electronic patient-reported outcome (ePRO) systems to monitor and deliver healthcare to a considerable number of patients.

✉ Olalekan Lee Aiyegbusi
O.L.Aiyegbusi@bham.ac.uk

[1] Centre for Patient Reported Outcome Research, Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Published online: 18 October 2019

Patients can access ePRO systems using mobile devices to provide feedback on their health status and response to treatments in 'real time' [9, 10]. The use of telehealth could therefore facilitate patient engagement with care which is a key element of delivering patient-centred care. It has also been demonstrated that patient reports of their health could complement clinical and laboratory parameters in routine clinical practice [11, 12]. Recent research suggest that the use of ePRO systems could facilitate the remote monitoring of patients [13]; enhance efficiency by reducing the need for hospital appointments [14]; and improve patient outcomes such as quality of life and survival rates [15]. The number of health care providers developing ePRO systems has increased in recent years [16, 17] and is set to rise considerably in future as more evidence to support their use become available.

It is therefore crucial that the user-friendliness and usability of the ePRO user interfaces are adequately assessed and improved throughout system development to reduce attrition rates in clinical trials and enhance their adoption post-implementation in clinical practice.

This article highlights the important issues that need to be considered and addressed when planning the usability testing of ePRO systems. Although ePRO systems are the primary focus, majority of the issues discussed are relevant for usability testing of websites or other types of systems that involve human–computer interaction. This paper is focused on methodological considerations for planning usability tests in the context of ePRO systems rather than the design of user interface. However, guidance and recommendations for the design of user interfaces are available in various publications and guidelines [18–23].

Usability and usability testing

According to the International Organization for Standardization (ISO), usability is an outcome of use which can be defined as "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [23].

Therefore, usability testing can be described as the formal assessment of the extent to which interaction with a product or system is effective, efficient and perceived as satisfactory by users. It allows end users to actually test ePRO systems and provides developers the opportunity to evaluate usability.

Based on the ISO definition, the three measures of usability are effectiveness, efficiency and satisfaction. Effectiveness refers to the ability of participants to perform tasks in order to achieve pre-determined goals completely and accurately without negative consequences [23, 24]. A negative consequence in the case of an ePRO system might be the accidental selection of a questionnaire option (due to suboptimal interface layout) which would send a red alert to clinicians. Efficiency relates to the amount of resources required by participants to achieve the pre-specified goals. An effective and efficient system or product could be considered as one that offers a better way of achieving specific goals compared to the current manner [25]. Satisfaction refers to the subjective opinions of participants based on their experience interacting with a system or product [23]. Some authors consider satisfaction with a system or product as equivalent to desirability the presence of which might actually facilitate the adoption of a system or product with flawed effectiveness and efficiency [25]. However, it can be argued that a system with effectiveness or efficiency issues would soon be abandoned regardless of initial desirability. An ePRO system needs to be rated highly on the three measures of usability to be considered fit for purpose. In turn, a system perceived as fit for purpose has a better chance of adoption [26] by patients and clinicians post-implementation.

The context of use refers to the characteristics of the users, tasks, equipment and the physical and social setting in which a system or product is used [23].

Evaluation of the measures of usability may be achieved by recruiting participants from the target population to perform pre-determined tasks using the product or system and provide feedback on their user experiences. Usability testing also provides ePRO developers the opportunity to detect and fix issues early during system development. It is important that usability testing is conducted iteratively [27] during system development to ensure that issues are detected and addressed adequately prior to full-scale implementation. This ensures that the final product or system is fit for purpose and may reduce attrition rates post-implementation [28].

Key points to consider when planning a usability test

Complexity of the user pathway and the ePRO system

An ePRO system is typically nestled within a broader IT system and usually requires users to perform a number of tasks before the ePRO questionnaires can be accessed. These may include tasks such as navigating webpages by following url links, entering personal details for verification before gaining access to the ePRO portal and login out of the system. Usability testing should assess the entire pathway and identify potential issues as its user-friendliness is crucial for user adoption. The complexity of an ePRO system also needs to be assessed when planning its usability testing. Most

ePRO systems involve the adaptation of existing paper PRO questionnaires [24, 29]. A basic adaptation keeps the electronic version as identical to the paper version as possible. It involves minor modifications to format or questions and such systems usually have a low level of complexity. Moderate adaptations may include subtle changes to meanings and format such as text font, colour or size. An extensive adaptation entails substantial changes such as the removal of items or the addition of functions such as drop down menus leading to the development of a more complex or sophisticated system [19, 30]. The greater the modification and in turn the complexity of an ePRO system, the larger the overall sample size and the number of test cycles that might be required [31]. In addition, the greater the degree of modification of an existing paper questionnaire, the greater the likelihood that additional studies such as psychometric validation might be required to evaluate the electronic version [19, 30].

Stage of system development

While the focus of this article is usability testing, it is worth mentioning that it is one of a number of study methods utilised during system development. System development can be divided into five stages, namely (i) planning, (ii) analysis, (iii) design, (iv) implementation and (iv) support [32]. Various study methods can be utilised during these stages to ensure that user requirements are met and the ePRO system is therefore fit for purpose. For instance, interviews, focus groups and surveys may be conducted with stakeholders during the planning and analysis stages, while usability testing and inspection techniques such as heuristic evaluation and cognitive walkthroughs may be conducted during design and implementation stages [32]. Figure 1 depicts the relationship between the stages of system development and the methods applicable.

It is important to consider the stage of system development as this would determine the type and depth of usability testing to conduct. Broadly speaking, there are two types of usability testing: formative and summative testing [33]. Formative testing is usually performed during the early stages of system development and the aim is to identify major (usually technical) issues. Formative usability testing may be conducted before any development work is done during the design phase using wireframes (mock-up screens) [34]. Formative test sessions are often less formal with greater interaction between participants and moderator. Formative tests collect mainly qualitative data especially during the earlier cycles, although some quantitative data such as error rate may also be collected [35]. After a series of iterative formative testing, the number of which might depend on the amount of issues detected by participants interacting with the system, summative testing may be conducted. The aim of summative testing is to obtain definitive

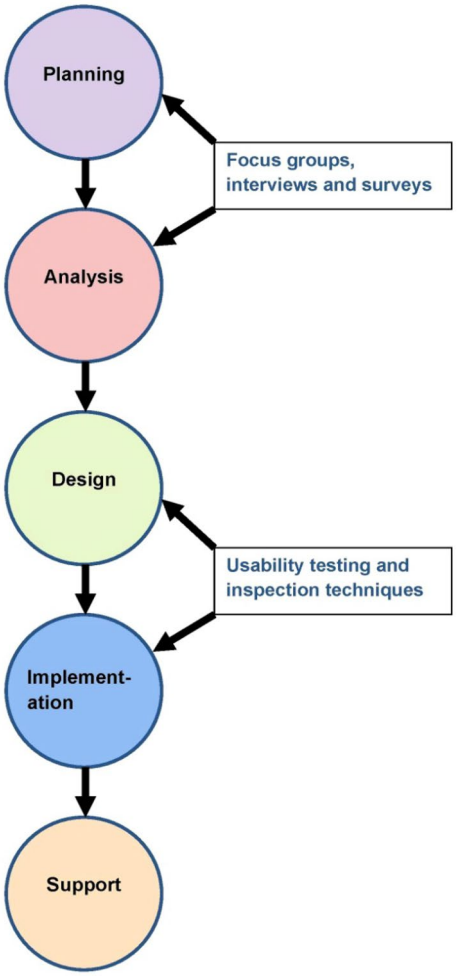

Fig. 1 Relationship between the stages of system development and applicable methods

evidence of usability [25] which may be used to support government regulatory claims or marketing campaigns. Summative test sessions are usually more formal with little or no interaction with a moderator.

Summative tests usually involve closer observation and recording of participants' actions as well as the collection of more quantitative data such as success or failure on tasks, average time on task, completion rates and error rates for statistical calculations [25]. As summative testing involves more statistical analyses, it requires more participants than formative testing. The issue of sample sizes is discussed further in the dedicated section.

The stage of ePRO development would also determine whether tests are conducted on-site or off-site (often at participants' homes). During the early stages, on-site moderated tests are more appropriate as these provide the opportunity to observe how well participants interact with a system. However, later on testing should be done remotely within participants' own environment. Remote

testing may be synchronous or asynchronous. In synchronous, the session is facilitated and data are collected by the evaluators in real time, while in asynchronous the session is not facilitated and the evaluator only has access to the data after the session has ended [36]. As off-site testing more closely resembles real life use, a successful test may provide ePRO developers the assurance that a system is indeed usable. However, a number of studies have demonstrated that remote synchronous usability tests may provide comparable results to traditional on-site tests of the same website or system, while participants of asynchronous tests may require more time to complete tasks [37, 38]. Remote testing may help developers detect potential internet, software or hardware compatibility issues.

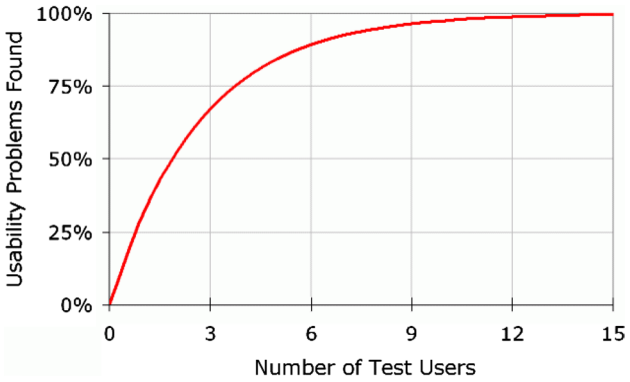

Fig. 2 Sample size for usability cycles. Reproduced with the kind permission of the Nielsen Norman Group [53]

Usability metrics to measure

Usability metrics to measure may be grouped into three categories: self-reported, observer-reported and implicit [39]. Self-reported metrics come directly from participants and include satisfaction and difficulty ratings. Observer-reported relates to assessments of participants' actions by the evaluator. Observer-reported metrics include time to complete tasks. Self- and observer-reported metrics may suffer from bias as participants often consider their responses and are conscious of their actions and may not act as they would in real life [40, 41]. Implicit metrics which are less commonly used may provide the most unbiased data as they measure participants' unconscious behaviours and physiology [35]. These include eye tracking and pupillary dilation [42].

Usability metrics relevant to ePRO systems are linked to the measures of usability (i.e. effectiveness, efficiency and satisfaction). Relevant quantitative metrics for effectiveness include error rates and completion rates. Time required for completing tasks, numbers of clicks to complete tasks, and cost effectiveness are appropriate metrics for efficiency. Overall satisfaction rates and proportion of users reporting complaints can be used to assess satisfaction [35]. While effectiveness, efficiency and satisfaction are often assessed quantitatively, they could also be assessed qualitatively. For example, effectiveness could be assessed by discussing errors and successful task completions with participants. Participants could also describe their satisfaction with the system in their own words [35].

The choice and number of metrics to measure may be influenced by the type of usability testing being conducted. As mentioned earlier, formative testing may involve the measurement of fewer quantitative metrics; relying more on qualitative feedback from participants while summative testing, which often involves more statistical analyses, tends to require the measurement of more quantitative metrics.

The use of usability questionnaires

Developers of ePRO systems could use usability questionnaires to capture and quantify participants' subjective opinion and satisfaction with their ePRO interfaces. Some questionnaires are designed for specific interfaces such as the Website Analysis and Measurement Inventory for websites [43]. Others, such as the System Usability Scale (SUS) [44], are more generic and can be used across interfaces. The use of such scales provides the opportunity to generate additional data which could be analysed to generate useful statistics about ePRO systems. Participants' scores from each test cycle may be compared with previous scores to confirm any improvements in satisfaction with the system. However, not all developers perceive usability scales as pertinent and some studies have suggested that a qualitative approach might be more useful especially in studies involving older participants [17, 45]. Once again it is vital that the goals of the developers and stakeholders are considered when making decisions about using usability scales.

Sample size

There has been considerable debate about the appropriate sample size for usability testing [31, 46–50]. Testing with more participants than necessary would increase costs and project time [51]. On the other hand, important issues might go undetected if inadequate sample sizes are used. In reality, there are no magic formulas for calculating sample sizes. The decision needs to be based on the careful consideration of a number of factors, namely (i) iterative nature of usability testing, (ii) homogeneity of target end users, (iii) complexity of the system and (iv) type of usability testing.

Iterative nature of usability testing

Studies have shown that five participants are required per (formative) test cycle to detect over 80% of issues (Fig. 2) [27, 52, 53]. However, as Spool and Schroeder demonstrated in their study, up to 15 participants might be required before serious usability issues are found [31]. Many system developers, usability personnel and researchers struggle to accept the recommendation of five users per test cycle as they are more familiar with larger sample requirements for most qualitative and quantitative studies. However, improving the usability of any system should be an iterative process which would allow developers the opportunity to detect and correct issues after each test cycle [27, 54]. It is therefore more sensible, for instance, to test with five participants per cycle and have the opportunity to detect issues and improve a system over four test cycles than to conduct a single cycle with 20 participants with no way of telling if subsequent changes to the system has improved its usability. It should be noted that this estimate of five participants per test cycle does not take into account the other three factors.

While the general expectation with iterative testing is that fewer issues will be detected with each test cycle until no substantial benefit is gained from further testing [54], it is quite possible that changes made on the basis of the results obtained from a cycle might inadvertently introduce fresh issues. Therefore, each cycle checks and assesses the changes made to the system. A 'stopping rule' needs to be agreed between the development team and commissioning body at the start of the project to prevent interminable testing [33]. An option is to stop further testing once the test results from a summative test meet predetermined targets [33].

Homogeneity of target end users

The estimate of five participants per iteration is only appropriate if the target end users are reasonably homogenous in their socio-demographic characteristics. For instance, studies have shown that the age of participants might have a significant influence on usability experiences [21, 55]. Therefore, it is very likely that younger and older end users would have different satisfaction levels if they interact with the same system [56]. For this reason, if a system is being designed for use by both age groups, each should be treated as a distinct group when estimating sample sizes. There is a suggestion that fewer participants may be recruited per group per cycle as some overlap in participant experience is bound to occur [53]. However, decisions about sample sizes may also be dependent on the complexity of the system.

Design of ePRO systems

The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) suggests that the complexity of the physical and cognitive tasks to be performed during a usability test may inform decisions about sample sizes [19]. The complexity of the tasks is influenced by the characteristics of the PRO questionnaire. Questionnaires that utilise matrices and drop down options are more complex compared to those with simple line by line formats. The current recommendation is 5 to 10 participants for simple ePRO systems and up to 20 for more physically and cognitively demanding systems [19]. However, these ranges are fairly wide and it is not clear whether the recommendation refers to each test cycle or the entire test [19].

How the system would be accessed may also influence sample size requirements as different platforms may have different usability issues [56]. The rapid developments in mobile technology have led to an increasing number of people accessing websites and web applications using mobile devices such as smartphones, tablets and phablets rather than 'traditional' desktops and laptops. Producing a version of an ePRO system for each type of device is probably impossible given the vast number of variations in screen sizes and resolutions. Responsive web design (RWD) an approach that allows dynamic adaptations to various screen sizes, resolutions and orientations is regarded as a solution [57]. Usability testing of websites or ePRO systems designed using RWD should be done across multiple platforms [58]. However, it is impractical to conduct usability testing for all types of devices. Therefore, developers have to decide which key platforms to test based on the degrees of similarity or differences between groups of devices. The recommendation of five subjects per test cycle should be applied to each device type (i.e. five subjects per device type per cycle) as user experiences may completely differ from one view of the ePRO system to another [58].

Touch screen devices such as smartphones and tablets may be easier to use and control than desktops or laptops which require keyboards and mouse. However, they usually have smaller screens which might influence the visual display and font sizes of ePROs. This could be an issue for participants with poor eyesight and this should be considered when selecting study participants [24]. The United States Access Board and the World Wide Web Consortium have published detailed guidelines to improve IT accessibility for all individuals regardless of disability [59–61].

Type of usability testing

Sample size requirements would also be influenced by the type of testing being conducted which would in turn be determined by the objectives of the developers. As the

data collected during formative usability testing (especially during the early stages) tends to be more qualitative than quantitative, sample size will be influenced by the theoretical approach and the achievement of thematic or data saturation [62, 63]. Summative usability testing would need more participants to ensure that statistical tests are adequately powered and the results meaningful [49]. Sample size calculations for controlled experiments would depend on study design, estimates of the variance and the desired level of precision (which includes the size of the critical difference and the chosen confidence level) [51, 64]. However, detailed discussion of sample size calculations for statistical tests is outside the scope of this present article.

Task scenarios

The setting for usability testing is by nature artificial and moderator-controlled. Despite this situation, participants are expected to interact with a system, website or product as they would 'normally' do without observation or guidance. It should be expected that in reality people will often behave differently during moderated on-site test sessions and unmoderated off-site testing. It is therefore necessary for the moderator to set the scene by providing suitable scenarios which give context and meaning to the tasks to be performed in order to achieve pre-determined goals. Scenarios should ideally mirror the types of outcome that may be obtained in real life. The platforms participants use for their tests will determine their pathway to accessing the ePRO system and in turn the applicable scenarios. For example, participants using a laptop might have to carry out some initial navigation by following web links, whereas smartphone users might only have to tap the icon for the ePRO application on their phones. The number of scenarios to use for a particular test session will depend on the number of possible outcomes for interaction with the ePRO system. This is especially relevant for ePRO systems that employ conditional branching (skip logic) where the sequence of questions is determined by participants' responses [65]. As there are a higher number of possible paths patients may take, there may be a need for more scenarios especially as not all questions may be formatted the same way. For instance, an item on 'pain' might have an initial 'yes or no' option. Individuals who click 'no' would move on to the next symptom, whereas those who select 'yes' would have a further item such as a visual analogue scale (VAS) ruler appear. Therefore, participant A's interaction and experience with the system may not be the same as that of participant B. Crafting realistic scenarios requires skill and a delicate balance has to be achieved with information provision. Participants should be given just enough information to execute the tasks [66].

Moderator's duties and choice of moderating technique(s)

Usability test sessions, particularly formative ones, need to be effectively moderated in order to derive useful insights which can subsequently be used for system improvement. Moderating usability tests is a skilled task that requires excellent judgement and observational skills. The degree of interaction between the moderator and the participants should be decided prior to the start of the testing cycle. As discussed earlier this would generally be determined by the type of usability testing to be conducted. The moderator needs to clarify before each test session that the purpose of the session is to evaluate the *interface* of the system and not to assess the *meaning and relevance* of the individual questions of the ePRO questionnaire. It is important that the moderator understands and makes this distinction as participants may confuse the two activities. For instance, they may comment on the clarity or suitability of individual questions rather than the font size of the interface. Content validation to evaluate the meaning and relevance of questions should be separately conducted for newly developed or extensively modified existing questionnaires.

There are four moderating techniques described in the literature [67], namely (i) concurrent think aloud (CTA), (ii) retrospective think aloud (RTA), (iii) concurrent probing (CP) and (iv) retrospective probing (RP).

In CTA, participants are encouraged to 'think aloud' and vocalise their thoughts on the user interface as they interact with the system or website and execute the predetermined tasks. The moderator employs minimal prompts to keep participants talking. With RTA, the test sessions are usually video recorded and participants complete their tests in silence. The moderator then asks them afterwards to recall and vocalise their thoughts during the test usually with the aid of the video recording [68]. The only technique in which the moderator plays an active role during test sessions is CP. In CP, the moderator asks probing or follow-up questions to participants' comments, nonverbal cues or noteworthy actions. When using RP, participants are allowed to complete their tests before being questioned by the moderator.

Each technique has its own advantages and disadvantages therefore the choice of technique to employ would depend on which qualities are important to system developers and stakeholders. Table 1 summarises these advantages and disadvantages [67]. A number of studies have compared moderating techniques [68]. It has been suggested that both 'think aloud' and retrospective approaches produce similar results which are prone to positive bias [18]. Participants in CTA sessions may take more time and complete fewer tasks compared to those recruited for sessions moderated by retrospective methods [18]. An

Table 1 Pros and cons of moderating techniques

Moderating techniques	Pros	Cons
Concurrent think aloud (CTA)	Understand participants' thoughts as they occur and as they attempt to work through issues they encounter Elicit real-time feedback and emotional responses	Can interfere with usability metrics, such as accuracy and time on task
Retrospective think aloud (RTA)	Does not interfere with usability metrics	Overall session length increases Difficulty in remembering thoughts from up to an hour before = poor data
Concurrent probing (CP)	Understand participants' thoughts as they attempt to work through a task	Interferes with natural thought process and progression that participants would make on their own, if uninterrupted
Retrospective probing (RP)	Does not interfere with usability metrics	Difficulty in remembering = poor data

Reproduced with the kind permission of Dr Jennifer Romano Bergstrom [67]

option is to employ more than one technique and RP is particularly suitable for combining with any of the others.

It is important to note that aside from moderating technique, moderator skills may have a significant impact on the conduct and outcome of tests. For instance, RP relies heavily on the moderator's ability to observe and note participants' actions, verbal and non-verbal cues during the tests for subsequent probing.

Conclusions

As advances in information technology continue and the adoption of mobile devices increases, digital healthcare will play a more prominent role in patient care. The development and use of ePRO systems could enhance the quality of clinical trials, and facilitate the remote monitoring and timely delivery of healthcare to patients. They could also promote patient involvement which is a crucial element of patient-centred care. However, the usability of the user interface of these systems needs to be adequately assessed by individuals drawn from the target population. The insights obtained from usability tests may be used to optimise ePRO systems to ensure that they are fit for purpose and acceptable to the end users.

Author contribution OLA is the sole author of the manuscript. He performed all the analysis of the research material and the writing of the manuscript.

Funding This work is funded as part of the Health Foundation's Improvement Science Programme (Ref: 7452). The Health Foundation is an independent charity working to improve the quality of healthcare in the UK. The Health Foundation was not involved in any other aspect of the project.

Compliance with ethical standards

Conflict of interest The author declares that there is no conflict of interest with respect to the research, authorship and/or publication of this article.

Ethical approval Not required as the study does not involve human subjects or animals.

References

1. Perrin A. *10 facts about smartphones as the iPhone turns 10: Pew Research Center*; 2017. Retrieved from October 2018 http://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones/.
2. Hong, Y. A., & Cho, J. (2017). Has the digital health divide widened? Trends of health-related internet use among older adults from 2003 to 2011. *The Journals of Gerontology: Series B, 72*(5), 856–863.
3. FDA. Guidance for industry. *Patient-reported outcome measures: Use in medicinal product development to support labeling claims*. Silver Spring, MD: US Department of Health and Human Services Food and Drug Administration; 2009.
4. Kyte, D., Bishop, J., Brettell, E., Calvert, M., Cockwell, P., Dutton, M., et al. (2018). Use of an electronic patient-reported outcome measure in the management of patients with advanced chronic kidney disease: The RePROM pilot trial protocol. *British Medical Journal Open, 8*(10), e026080.
5. Basch, E., Geoghegan, C., Coons, S. J., Gnanasakthy, A., Slagle, A. F., Papadopoulos, E. J., et al. (2015). Patient-reported outcomes in cancer drug development and US regulatory review:

Perspectives from industry, the food and drug administration, and the patient. *JAMA Oncology, 1*(3), 375–379.
6. FDA. *Food and Drug Administration Safety and Innovation Act*. Food and Drug Administration; 2012.
7. EMA. *Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man—The use of patient-reported outcome (PRO) measures in oncology studies*. European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP), 2016 EMA/CHMP/292464/2014.
8. Vidal-Fisher, L., Vidal Boixader, L., Andrianov, V., Curtis, K. K., Shepshelovich, D., & Moss, K. R. (2019). Reporting of patient reported outcome (PRO) in clinical trials: A systematic review of clinical trials. *Journal of Clinical Oncology, 37*(15_suppl), 6590.
9. Krawczyk, M., & Sawatzky, R. (2018). Relational use of an electronic quality of life and practice support system in hospital palliative consult care: A pilot study. *Palliat Support Care., 17*(2), 1–6.
10. Ginsberg, J. S., Zhan, M., Diamantidis, C. J., Woods, C., Chen, J., & Fink, J. C. (2014). Patient-reported and actionable safety events in CKD. *Journal of the American Society of Nephrology, 25*(7), 1564–1573.
11. Aiyegbusi, O. L., Kyte, D., Cockwell, P., Anderson, N., & Calvert, M. (2017). A patient-centred approach to measuring quality in kidney care: Patient-reported outcome measures and patient-reported experience measures. *Current Opinion in Nephrology and Hypertension, 26*(6), 442–449.
12. Bryan, S., Davis, J., Broesch, J., Doyle-Waters, M. M., Lewis, S., McGrail, K., et al. (2014). Choosing your partner for the PROM: A review of evidence on patient-reported outcome measures for use in primary and community care. *Healthcare Policy, 10*(2), 38–51.
13. Basch, E., Deal, A. M., Kris, M. G., Scher, H. I., Hudis, C. A., Sabbatini, P., et al. (2016). Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. *Journal of Clinical Oncology, 34*(6), 557–565.
14. Schougaard, L. M., Larsen, L. P., Jessen, A., Sidenius, P., Dorflinger, L., de Thurah, A., et al. (2016). AmbuFlex: Tele-patient-reported outcomes (telePRO) as the basis for follow-up in chronic and malignant diseases. *Quality of Life Research, 25*(3), 525–534.
15. Basch, E., Deal, A. M., Dueck, A. C., et al. (2017). Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA, 318*(2), 197–198.
16. Cox, C. E., Wysham, N. G., Kamal, A. H., Jones, D. M., Cass, B., Tobin, M., et al. (2016). Usability testing of an electronic patient-reported outcome system for survivors of critical illness. *American Journal of Critical Care, 25*(4), 340–349.
17. Steele Gray, C., Gill, A., Khan, A. I., Hans, P. K., Kuluski, K., & Cott, C. (2016). The electronic patient reported outcome tool: Testing Usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. *JMIR mHealth and uHealth, 4*(2), e58.
18. *The Research-Based Web Design & Usability Guidelines*, Enlarged/Expanded edition. Washington: U.S. Government Printing Office; 2006.
19. Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., et al. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value in Health, 12*(4), 419–429.
20. Zbrozek, A., Hebert, J., Gogates, G., Thorell, R., Dell, C., Molsen, E., et al. (2013). Validation of electronic systems to collect patient-reported outcome (PRO) data-recommendations for clinical trial teams: Report of the ISPOR ePRO systems validation good research practices task force. *Value in Health, 16*(4), 480–489.
21. Zaphiris, P., Kurniawan, S., & Ghiawadwala, Bulsara M. (2006). A systematic approach to the development of research-based web design guidelines for older people. *Universal Access in the Information Society, 6,* 59–75.
22. Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2009). *Designing for older adults: Principles and creative human factors approaches* (2nd ed.). Boca Raton: FL CRC Press.
23. ISO. ISO 9241-11:2018(en). *Ergonomics of human–system interaction—Part 11: Usability: Definitions and concepts*; 2018.
24. Aiyegbusi, O. L., Kyte, D., Cockwell, P., Marshall, T., Dutton, M., Walmsley-Allen, N., et al. (2018). Development and usability testing of an electronic patient-reported outcome measure (ePROM) system for patients with advanced chronic kidney disease. *Computers in Biology and Medicine, 101,* 120–127.
25. Barnum, C. M. (2011). 1—Establishing the essentials. In C. M. Barnum (Ed.), *Usability testing essentials* (pp. 9–23). Boston: Morgan Kaufmann.
26. Szajna, B. (1996). Empirical evaluation of the revised technology acceptance model. *Management Science, 42*(1), 85–92.
27. Bailey GD, editor. *Iterative methodology and designer training in human-computer interface design*. INTERCHI; 1993.
28. Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7*(1), e11.
29. Schick-Makaroff, K., & Molzahn, A. (2015). Strategies to use tablet computers for collection of electronic patient-reported outcomes. *Health and Quality of Life Outcomes, 13,* 2.
30. Shields, A., Gwaltney, C., Tiplady, B., Paty, J., & Shiffman, S. (2006). Grasping the FDA's PRO Guidance. *Applied Clinical Trials, 15*(8), 69.
31. Spool J, Schroeder W. (2001). Testing web sites: Five users is nowhere near enough. *CHI '01 Extended Abstracts on Human Factors in Computing Systems* (pp. 285–286). Seattle, Washington: ACM.
32. Kushniruk, A. (2002). Evaluation in the design of health information systems: Application of approaches emerging from usability engineering. *Computers in Biology and Medicine, 32*(3), 141–149.
33. Lewis, J. R. (2012). Usability testing. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (4th ed., pp. 1267–1312). New York: John Wiley.
34. Brown, D. M. (2011). *Wireframes communicating design: Developing web site documentation for design and planning* (2nd ed., pp. 166–200). Berkeley, CA: New Riders.
35. Geisen, E., & Romano, Bergstrom J. (2017). Chapter 5—Developing the usability testing protocol. In E. Geisen & J. Romano Bergstrom (Eds.), *Usability testing for survey research* (pp. 111–129). Boston: Morgan Kaufmann.
36. Bastien, J. M. C. (2010). Usability testing: A review of some methodological and technical aspects of the method. *International Journal of Medical Informatics, 79*(4), e18–e23.
37. Madathil KC, Greenstein JS. (2011). Synchronous remote usability testing: A new approach facilitated by virtual worlds. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2225–2234). Vancouver, BC: ACM.
38. Brush AJB, Ames M, Davis J, editors. (2004). A comparison of synchronous remote and local usability studies for an expert interface. *CHI*. Vienna/New York: ACM.
39. Romano Bergstrom JC, Strohl J, editors. (2013). *Improving government websites and surveys with usability testing: A comparison of methodologies*. Washington, DC.
40. Natesan, D., Walker, M., & Clark, S. (2016). Cognitive bias in usability testing. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care., 5*(1), 86–88.

41. Sauro J. (2012). *9 Biases in Usability Testing: MeasuringU*. Retrieved 16 August 2019 from https://measuringu.com/ut-bias/.
42. Bergstrom, J. R., & Schall, A. (2014). *Eye tracking in user experience design* (p. 400). San Francisco: Morgan Kaufmann Publishers Inc.
43. Kirakowski, J., & Cierlik, B. (1998). Measuring the usability of web sites. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42*(4), 424–428.
44. Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry*. London: Taylor and Francis.
45. Cornet, V. P., Daley, C. N., Srinivas, P., & Holden, R. J. (2017). User-centered evaluations with older adults: Testing the usability of a mobile health system for heart failure self-management. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting., 61*(1), 6–10.
46. Nielsen, J. (1994). *Usability engineering* (p. 165). Cambridge, MA: Academic Press Inc.
47. Macefield, R. (2009). How to specify the participant group size for usability studies: A practitioner's guide. *J Usability Studies, 5*(1), 34–45.
48. Turner, C., Lewis, J., & Nielsen, J. (2006). Determining usability test sample size. *International Encyclopedia of Ergonomics and Human Factors, 3*(2), 3084–3088.
49. Nielsen J. How many test users in a usability study?: Nielsen Norman Group; 2012. Retrieved November 2018 from https://www.nngroup.com/articles/how-many-test-users/.
50. *Industry usability reporting: National Institute of Standards and Technology*. Retrieved November 2018 from https://www.nist.gov/itl/iad/industry-usability-reporting.
51. Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user experience* (2nd ed.). San Francisco, CA: Elsevier, Morgan Kaufmann.
52. Bailey B. *Determining the correct number of usability test participants: usability.gov*; 2006. Retrieved November 2018 from https://www.usability.gov/get-involved/blog/2006/09/correct-number-of-test-participants.html.
53. Nielsen J. *Why you only need to test with 5 users: Nielsen Norman Group*; 2000. Retrieved November 2018 from https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/.
54. Romano Bergstrom, J., Olmsted-Hawala, E., Chen, J. M., & Murphy, E. (2011). Conducting iterative usability testing on a web site: Challenges and benefits. *Journal of Usability Studies, 7*, 9–30.
55. Becker, S. A. (2004). E-Government visual accessibility for older adult users. *Social Science Computer Review, 22*(1), 11–23.
56. Geisen, E., & Romano, Bergstrom J. (2017). Chapter 4—Planning for usability testing. In E. Geisen & J. Romano Bergstrom (Eds.), *Usability testing for survey research* (pp. 79–109). Boston: Morgan Kaufmann.
57. Marcotte E. Responsive web design: A list apart; 2010. Retrieved 14 August 2019 from https://alistapart.com/article/responsive-web-design/.
58. Schade A. *Responsive web design (RWD) and user experience: Nielsen Norman Group*; 2014. Retrieved 14 August 2019 from https://www.nngroup.com/articles/responsive-web-design-definition/.
59. Section 508 Standards for Electronic and Information Technology: United States Access Board. Retrieved 10 August 2019 from https://www.access-board.gov/guidelines-and-standards/communications-and-it/about-the-section-508-standards/section-508-standards.
60. Information and Communication Technology (ICT) Final Standards and Guidelines: United States Access Board. Retrieved 10 August 2019 from https://www.access-board.gov/guidelines-and-standards/communications-and-it/about-the-ict-refresh/final-rule.
61. Web Content Accessibility Guidelines (WCAG) 2.1. Retrieved 10 August 2019 from https://www.w3.org/TR/2018/REC-WCAG21-20180605/.
62. Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. *Field Methods., 18*(1), 59–82.
63. Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.
64. Haas, J. P. (2012). Sample size and power. *American Journal of Infection Control, 40*(8), 766–767.
65. Norman K. *Implementation of Conditional Branching in Computerized Self-Administered Questionnaires*; 2001.
66. Sauro J. (2013). *Seven tips for writing usability task scenarios*. Retrieved November 2018 from https://measuringu.com/task-tips/.
67. Bergstrom J. (2013). *Moderating usability tests: usability.gov*. Retrieved November 2018 from https://www.usability.gov/get-involved/blog/2013/04/moderating-usability-tests.html.
68. Van DenHaak, M., De Jong, M., & Jan, Schellens P. (2003). Retrospective vs concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology, 22*(5), 339–351.