



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Standards for Instrument Migration When Implementing Paper Patient-Reported Outcome Instruments Electronically: Recommendations from a Qualitative Synthesis of Cognitive Interview and Usability Studies

Willie Muehlhausen, DVM¹, Bill Byrom, PhD^{2,*}, Barbara Skerritt¹, Marie McCarthy, MBA¹, Bryan McDowell, MBA³, Jeremy Sohn, BA³

¹ICON Clinical Research, Dublin, Ireland; ²ICON Clinical Research, Marlow, Buckinghamshire, UK; ³Novartis Pharmaceuticals, Basel, Switzerland

ABSTRACT

Objectives: To synthesize the findings of cognitive interview and usability studies performed to assess the measurement equivalence of patient-reported outcome (PRO) instruments migrated from paper to electronic formats (ePRO), and make recommendations regarding future migration validation requirements and ePRO design best practice. **Methods:** We synthesized findings from all cognitive interview and usability studies performed by a contract research organization between 2012 and 2015: 53 studies comprising 68 unique instruments and 101 instrument evaluations. We summarized study findings to make recommendations for best practice and future validation requirements. **Results:** Five studies (9%) identified minor findings during cognitive interview that may possibly affect instrument measurement properties. All findings could be addressed by application of ePRO best practice, such as eliminating scrolling, ensuring appropriate font size, ensuring suitable thickness of visual analogue scale lines, and providing suitable instructions. Similarly, regarding solution usability, 49 of the 53 studies (92%) recommended no changes in display clarity, navigation, operation, and completion

without help. Reported usability findings could be eliminated by following good product design such as the size, location, and responsiveness of navigation buttons. **Conclusions:** With the benefit of accumulating evidence, it is possible to relax the need to routinely conduct cognitive interview and usability studies when implementing minor changes during instrument migration. Application of design best practice and selecting vendor solutions with good user interface and user experience properties that have been assessed in a representative group may enable many instrument migrations to be accepted without formal validation studies by instead conducting a structured expert screen review.

Keywords: cognitive interview, electronic patient-reported outcomes, ePRO, measurement equivalence, patient-reported outcomes, PRO.

Copyright © 2018, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Because of significant improvements in the integrity, quality, and timeliness of data collected and increased awareness of the potential benefits of electronic collection, a growing number of clinical trials are using electronic media (smartphones and tablets) to collect patient-reported outcomes (PROs). Because many PRO instruments were developed and validated on paper, care is needed when migrating them to electronic formats (ePRO) to ensure that the measurement properties of the instrument are unchanged and that the electronic version is easy to use in the target group of patients. In 2009, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) ePRO Good Research Practices Task Force published recommendations on the evidence needed to support measurement equivalence when

migrating from paper to electronic formats [1]. This task force recommended that minor changes to an instrument because of migration should require a cognitive interview and usability study in the target patient population to demonstrate measurement equivalence. Such minor changes include, for example, minor formatting changes such as presenting only a single question per screen or wording changes such as changing question response instructions from “tick or circle” on pen and paper to “select” on an electronic implementation. These recommendations have been largely adopted by the industry and regulators.

In this context, cognitive interviews typically involve developing a semistructured interview that is conducted by a trained qualitative interviewer to collect information about patient experience after they completed the instrument both on paper (or its original form) and in the electronic format. Structured, probing

* Address correspondence to: Bill Byrom, ICON Clinical Research, Unit 2 Globeside Business Park, Marlow, Buckinghamshire SL7 1HZ, UK. E-mail: bill.byrom@iconplc.com.

questions help to identify whether changes in format and presentation might affect the way patients respond to the questions and, thus, whether the modality provides equivalent patient responses. These studies are typically carried out in a small sample ($n = 5\text{--}10$) of the target patient population and interviews are transcribed and summarized qualitatively [2].

The purpose of our synthesis was to explore whether routine performance of cognitive interview and usability studies should remain a recommendation for all migrations requiring minor modifications or whether the benefit of growing evidence obtained from conducting these evaluations is supportive of other less arduous approaches. We also used our synthesis to confirm ePRO design best practice recommendations.

This is not the first review exploring learnings from previous migration studies. Two meta-analyses of equivalence studies performed on instruments migrated from original to electronic formats have been reported [3,4]. Both analyses concluded that there is no meaningful evidence that migration to alternative formats affects instrument measurement properties (the analyses considered 46 and 72 equivalence studies, respectively [3,4]).

One of the fundamental aspects of our analysis has been to consider instruments as a collection of response scale types as opposed to a combination of items. Common response scale types include the following [5]:

1. *Verbal response scales (VRSs)*: These comprise a question prompt and an associated list of response options ordered in a logical scale order, for example, mild, moderate, and severe.
2. *Numeric response scales (NRSs)*: These scales combine question text with a horizontal list of ordered numbers reflecting the degree of association with the construct measured, such as severity or agreement. The scale interpretation is typically anchored using a text description to describe the first and the last number of the scale. An NRS to measure pain severity might, for example, ask the subjects to rate their pain on a scale from 0 to 10, where 0 represents “no pain” and 10 represents “worst possible pain.”
3. *Likert scales*: These scales measure a concept ranging from a positive to a negative rating, with the center option being neutral, for example, measuring satisfaction from very satisfied to very unsatisfied. These can be presented using a VRS or an NRS.
4. *Visual analogue scales (VAS)*: These scales use a straight horizontal line on which the respondents mark their assessment of a specific construct. The scale interpretation is typically anchored using a text description to describe each end of the horizontal line, for example, “no pain” to “worst possible pain.”

Additional response options sometimes included in electronic clinical outcome assessment instruments include yes/no fields, number entry fields, free-text fields, multiple choice fields, and time and date fields. Because these response types are common in everyday usage of a mobile device and personal computer (PC) applications, we did not consider it necessary to evaluate them specifically in this work.

The rationale for considering migration assessment by response scale types is founded in the hypothesis that potential changes in an instrument’s measurement properties, after minor formatting and layout changes due to migration, are primarily concerned with understanding whether subjects can interact with each response scale type appropriately and in the same way on both modalities, independent of the specific question item or construct that each item evaluates. Ensuring each item is an appropriate measure of the required construct has already been assessed thoroughly in the development and psychometric validation activity performed by instrument authors, and so

when changes are minor there is no requirement to re-assess this in cognitive interviews associated with migration assessment. This may mean that previous migration studies on instruments using the same response scale types can provide evidence of migration acceptability for new instruments, so long as ePRO design best practice standards are followed.

Methods

We synthesized findings from all cognitive interview and usability studies performed between 2012 and 2015 by a contract research organization (CRO) to which a number of the authors belong: 53 studies comprising 68 unique instruments and 101 instrument evaluations. These studies are rarely published in the scientific literature, but are routinely included in drug approvals by sponsor organizations to regulatory authorities to support the appropriate use of ePRO for clinical trials within the submission [1].

In all studies, cognitive interview and usability assessment was performed using a standardized semistructured interview conducted by an experienced qualitative interviewer. Patients were asked to read and complete both modes of instrument administration. Interviewers probed whether any perceived differences in the self-report task or aspects of the changes between formats—such as overall appearance, text size, instructional information, moving from question to question, and how responses were selected—may, in the patients’ perception, have caused them to potentially answer differently between formats. Usability questions explored the clarity of text and images, ease of navigation, use of touch screen, and whether participants felt they would be able to use the electronic solution without help. All interviews were recorded and transcribed, and findings were summarized.

For each study, we identified the instruments studied and the response scale types they contained, the patient population and sample size, and the electronic modality compared with the original paper instrument. Reported findings of the cognitive interviews were summarized, specifically identifying whether changes in the way patients responded to instrument items because of migration differences were reported, and any additional recommendations. For each usability testing report, we summarized findings relating to display clarity, navigation, use of touch screen/stylus (where applicable), and the ability of patients to use the electronic solution without help. We synthesized the findings across all studies included.

Results

Description of Studies and Instruments

The 53 studies were conducted in samples ranging from 5 to 30 patients (median sample size: 10 patients) and included patients from a broad range of therapeutic areas. Out of these studies, 6 studies (11%) were conducted in healthy volunteers; 9 studies (17%) included patients with respiratory conditions including asthma and chronic obstructive pulmonary disease; 7 studies (13%) included patients with gastrointestinal conditions such as ulcerative colitis, Crohn disease, and constipation; and 7 studies (13%) included oncology patients including those with breast cancer, melanoma, and gastric/bladder cancer. A further 6 studies (11%) included rheumatology patients all involving patients with osteoarthritis of the knee; 4 studies (8%) included central nervous disease indications including migraine; and 14 studies (26%) involved patients with other conditions (Fig. 1). Patients were aged between 5 and 84 years (Table 1). Four studies included

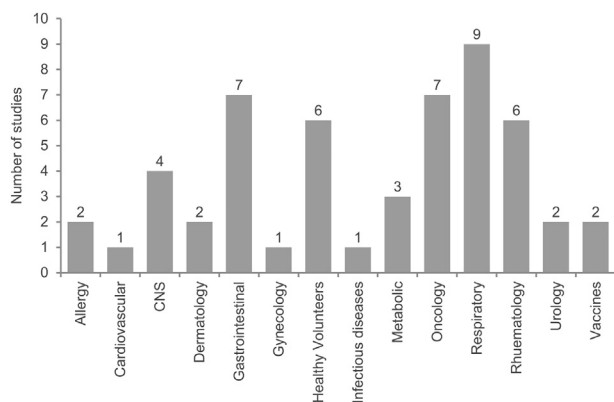


Fig. 1 – Therapy areas included in the 53 cognitive interview and usability studies. CNS, central nervous system.

children and were conducted with the aid of parent/caregivers. Almost half (45%) of the studies included some elderly participants (older than 65 years). Overall, studies contained a similar number of male and female patients (51.5% and 48.5%, respectively; 540 patients in total). The ethnicity of the participants is presented in Table 1. In addition, patients were recruited across a broad range of educational backgrounds and technology familiarity.

The studies examined 68 different PRO instruments and totaled 101 instrument evaluations (Table 2). Individual studies investigated between 1 and 5 instruments in the same group of patients (median = 1, observed in 27 of 53 [51%] studies). Most common instruments evaluated included the three-level [6] and the five-level [7] EuroQol Five-Dimensional Questionnaire, the Asthma Control Questionnaire [8], the Asthma Quality of Life Questionnaire [9], the European Organization for Research and Treatment of Cancer QLQ-C30 [10], the Short Form 36 Health Survey, Version 2 [11], and the Knee injury and Osteoarthritis Outcome Score [12].

Table 1 – Demographic characteristics of subjects included in cognitive interview studies.

Characteristic	Value
Age range (y)	5–84
Number of studies (%) by age category*	
Child (up to 11 y)	4 (7.5)
Adolescent (12–17 y)	4 (5.5)
Adult (18–65 y)	50 (94)
Elderly (> 65 y)	24 (45)
Sex, n (%)	
Female	278 (51.5)
Male	262 (48.5)
Ethnicity, n (%)	
White	396 (73.3)
Black/black British/African American	69 (12.8)
Asian	17 (3.1)
Hispanic	10 (1.9)
Arab	1 (0.2)
Mixed race (black/white)	27 (5.0)
Mixed race (other)	17 (3.1)
Not stated	3 (0.6)

* Not mutually exclusive.

Table 2 – Summary of instruments included in the synthesis of migration studies.

Instrument	No. of studies
Three-level EuroQol Five-Dimensional Questionnaire [6]	7
Five-level EuroQol Five-Dimensional Questionnaire [7]	5
Asthma Control Questionnaire [8]	5
Asthma Quality of Life Questionnaire [9]	4
The EORTC QLQ-C30 [10]	4
Short Form 36, version 2 [11]	3
Knee injury and Osteoarthritis Outcome Score [12]	3
Asthma Symptom Utility Index [13]	2
COPD Assessment Test [14]	2
St George’s Respiratory Questionnaire [15]	2
St George’s Respiratory Questionnaire (Change) [15]	2
Western Ontario and McMaster Universities Osteoarthritis Index [16]	2
Proprietary Vaccine Reaction Report Card	2
Inflammatory Bowel Disease Questionnaire [17]	2
Pain Visual Analogue Scale	2
The EORTC Breast Cancer-Specific Quality of Life Questionnaire [18]	2
Other instruments (n = 52)	1 (n = 52)
Total number of evaluations	101

COPD, chronic obstructive pulmonary disease; EORTC, European Organisation for Research and Treatment of Cancer.
* Version for age 12 y and older.

Twenty-seven studies (51%) examined migration to tablets, 25 (47%) to handhelds, and 1 (2%) to PC/laptops, corresponding to 46, 54, and 1 instrument evaluation, respectively.

Seventy-three percent (74 of 101) of instrument migrations displayed a single question and response item per screen, including all the handheld device (smartphone) migrations. Twenty-four percent (24 of 101) of migrations used multiple questions per screen (tablet or PC modalities only). A further 3% of studies did not have screenshot reports available within the cognitive interview study report, and this property could not be evaluated post hoc.

Instruments varied in length, containing between 1 and 52 items (median = 10 items). Most commonly, instruments used a single item response scale type throughout (31 of 68 [46%]), for example, the Asthma Control Questionnaire [8] and the Asthma Quality of Life Questionnaire [9], both of which use 7-option VRS, and the Western Ontario and McMaster Universities Osteoarthritis Index [16], which uses the VAS response scale type. Other instruments included multiple response scale types, such as the Asthma Control Diary [19], which combines VRS and numeric entry (for capture of peak flow rate). The composition of one instrument was not retained by the CRO because this instrument was proprietary to the sponsor organization.

Description of Instrument Response Scale Types

The evaluated response scale types, and their frequencies, are presented in Table 3.

VRSs (ranging from three to nine response options) were the most commonly used response type (49 of 68 [72%] instruments studied) (Table 3).

NRSs (ranging from 6- to 11-point scales) were used in 21% of instruments (14 of 68) and 18 instrument evaluations (Table 3). These scales combine question text with a horizontal list of

Table 3 – Summary of instrument response option types evaluated in the synthesis of migration studies.

Response option type	No. of instruments	No. of evaluations
Verbal response scale	49	84
3-option scale	9	21
4-option scale	15	24
4-option scale + NA option	1	1
5-option scale	31	41
5-option scale + NA option	1	1
6-option scale	7	10
7-option scale	8	16
9-option scale	1	1
Numeric response scale	14	18
6-point scale (L/R anchors)	1	2
7-point scale (L/R anchors)	2	5
8-point scale (L/R anchors)	1	1
8-point scale (3 anchors)	1	1
11-point scale (L/R anchors)	10	10
11-point scale (5 anchors)	1	1
Visual analogue scale (L/R anchors)	4	6
Vertical 101-point scale (EQ-5D)	2	12
Likert scale	7	7
5-option verbal scale	2	2
7-option verbal scale	4	4
7-option verbal scale + NA option	1	1
7-point numeric scale (L/R anchors)	1	1
Single-selection list	7	8
2-option list	2	2
3-option list	3	4
4-option list	3	3
5-option list	1	1
6-option list	2	2
8-option list	1	1
Multiple-selection list	4	4
4-option list, any number selected	1	1
5-option list, any number selected	3	3
8-option list, any number selected	1	1
10-option list, any number selected	1	1
11-option list, 3 to be selected	1	1
Yes/no response	17	19
Yes/no	17	19
Yes/no + NA option	1	1
Number field	8	9
Free-text field	3	4
Time field	8	9
Date field	6	7
Total	68	101

EQ-5D, EuroQol five-dimensional questionnaire; L/R; anchors, anchor text for left and right ends of the scale. NA, not applicable.

numbers reflecting the degree of association with the construct measured, such as severity or agreement. All scales used anchor text providing a description of the minimum and maximum points on the scale (Fig. 2A). One instrument, the Psychosexual

Daily Questionnaire [20], used a third anchor to differentiate “none” from “very low” (Fig. 2B), and a second instrument, a sponsor specific daily diary, used anchor ranges to identify bands of severity (Fig. 2C). Another instrument, the Diabetes Treatment Satisfaction Questionnaire (Change) [21], used a Likert scale in the form of an NRS with values ranging from 3 to –3, with anchors at either end of the scale (Fig. 2D). In terms of how users functionally operate this scale, we do not differentiate these variants from other forms of NRS.

Likert scales were used in seven (10%) instruments and seven evaluations. Besides the numeric response Likert scale used in the Diabetes Treatment Satisfaction Questionnaire (Change) [21], six additional instruments used Likert scales using a VRS with five or seven options, with one instrument also containing an additional “not applicable” option.

VASs were used in four instruments and evaluated 6 times through cognitive debrief studies (Table 3). In all cases, the scale interpretation was anchored using a text description to describe each end of the horizontal line.

Other response option types contained in the instruments evaluated included the vertical 101-point scale (specific to the three-level [6] and the five-level [7] EuroQol Five-Dimensional Questionnaire); a single-selection list (identical to the VRS except that categories do not exhibit a logical ordering or size); and a multiple-selection list ranging from 4 to 11 options, with most not limiting the number of options that could be selected (exceptions included the Core Lower Urinary Tract Symptom Diary [22], which asked respondents to select a maximum of three items, for example). Yes/no (17 instruments), number entry fields (8 instruments), free-text fields (3 instruments), and time and date fields (8 and 6 instruments, respectively) were also included in some instruments.

Cognitive Interview Findings

Three studies evaluated usability only because no paper instrument existed. The remaining 50 studies of 98 instrument assessments reported cognitive interview results. Forty-five (90%) studies involving 91 (93%) instrument assessments reported that migration was appropriate. Five studies (10%) involving 9 (10%) instruments identified minor findings, some of which were thought may affect measurement properties (Table 4).

One study (study A) reported that some patients found the use of the PC format difficult. Multiple questions were displayed per screen and questions and their response options could be revealed only, along with the “back” and “next” navigation buttons, by scrolling. This caused usability issues for 2 of 10

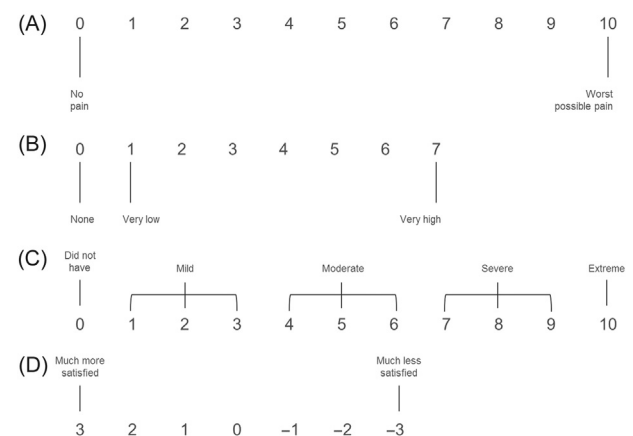


Fig. 2 – Anchor text variants for numeric response scale (NRS) types observed amongst the instruments studied.

patients. In addition, these patients had difficulty locating the cursor and differentiating between left and right mouse clicks. Although these patients did not believe this affected the responses they gave, it does draw into question the appropriateness of this migration on the PC format. Because this was a site-based instrument, it is likely that such issues could be resolved through instruction and supervision. Following design best practice and using a touch screen/tablet format would likely also eliminate these potential issues, such as scrolling.

A second study (study B) found that 3 of 10 patients felt they may have answered differently on three instruments when considering paper and tablet implementations. In all cases, the reasons cited for this were a result of presenting multiple questions per page on the tablet version. Because of the smaller display size, the smaller font size, and reduced spacing between paragraphs on the tablet format, these patients did not always believe their answers would be the same between paper and electronic formats. This could be resolved by breaking pages within the tablet format or displaying a single item per page. Interestingly, the last two questions of one questionnaire were split onto two pages on the paper version and combined on the tablet version. One patient felt that presenting the questions together helped them to think of them both as a group when answering them and may have resulted in a different answer compared with the paper version. Again, this potential difference could have been resolved by using a single question per screen format on the tablet version.

A further study (study C) found that the font size should be increased and the VAS line thickened in three instruments using a VAS. This was a usability/clarity of display finding and patients did not believe it made a difference to the way they used the

scale between paper and tablet versions. In addition, 3 of 10 patients were unable to understand the terminology used to distinguish between a target knee and a contralateral knee when answering questions about their pain. This was an issue for both paper and electronic formats and more a result of lack of instructional text in both modalities of the instrument.

Study D migrated a proprietary paper diary to a handheld device format. In this study, 3 of 10 adults found it difficult to use the up and down arrows to select number entry, and 1 found it hard to understand how to enter free text. Again, these were usability issues, but could affect the responses made between paper and electronic formats if insufficient instruction was given. In addition, 2 of 10 adults found that the changes in the phrasing of one question on medication usage may have affected their responses. The nature of this wording change was substantive and should be considered a moderate or major change to the instrument, and so did not permit equivalence to be assessed. It was likely, however, that the wording change used on the handheld version was superior.

A final study (study E) presented yes/no questions without providing a definition of terms used in the question. On the paper instrument, the definition was clearly presented alongside the question text, but on the electronic version the definition could be revealed only by tapping on the word. Three out of 15 subjects felt their answers may have differed between formats because of this. It should be noted that although this migration was assessed by cognitive interview and usability testing, the nature of this change is more appropriately considered a moderate change to the instrument. In such cases, the ISPOR recommendations would propose an equivalence study to assess the acceptability of instrument migration [1].

Table 4 – Summary of findings from cognitive interview and usability instrument evaluations.

Study	Electronic media	Finding type	No. of instruments	Response scale type	Items per screen	Detail
<i>Cognitive interview instrument evaluations (n = 98)</i>						
A	PC	Scrolling	1	VRS	Multiple	Remove the need to scroll to reveal all the response options.
B	Tablet	Multiple questions per screen	3	VRS	Multiple	3 of 10 subjects felt they may answer differently. Multiple questions per screen led to a small font size and compressed paragraphs. The last two pages were combined on the tablet, whereas split on paper.
C	Tablet	Display clarity	3	VAS	Single (2) Multiple (1)	Terminology on one question not understood. Enlarge font size and darken VAS line.
D	Handheld	Information presentation	1	Number field Free-text field	Single	Add instructions on how to use numeric entry and text entry fields.
E	Handheld	Information presentation	1	Y/N	Not recorded	Add definition of terms on screen as opposed to clicking on them to reveal text.
<i>Usability test instrument evaluations (n = 101)</i>						
A	PC	Scrolling	1	VRS	Multiple	Remove the need to scroll to reveal all the response options.
C	Tablet	Display clarity	3	VAS	Single (2) Multiple (1)	Terminology on one question not understood. Enlarge font size and darken VAS line.
D	Handheld	Information presentation	1	Number field Free-text field	Single	Add instructions on how to use numeric entry and text entry fields.
F	Handheld	Display clarity	1	VRS	Single	5 of 6 patients felt the device was too small.

PC, personal computer; VAS, visual analogue scale; VRS, verbal response scale; Y/N, yes/no selection.

All findings could be addressed by application of ePRO design best practice guidelines [5,6] including eliminating scrolling, ensuring appropriate font size and VAS line thickness, presenting a single question per screen, and providing clear and comprehensive instructions.

Usability Study Findings

Usability findings were reported for 4 of 53 (8%) studies and referred to 6 of 101 (6%) instrument evaluations. Findings for the first three studies (studies A–C) were the same as reported earlier in the cognitive interviews. A fourth study (study F) reported display clarity as an issue with a smartphone migration of one instrument. This was part of a larger study also exploring a tablet format. Because of the unavailability of the smartphone at the time of the study, smartphone screens were displayed on a tablet with the sizing of screens consistent with the target smartphone. This caused navigation confusion and made the smartphone display appear small when presented within the larger tablet screen. This is not recommended for assessment of usability, and it is our opinion that the smartphone solution was likely to be adequate when deployed properly on that device. This was the only migration assessed in this way.

Conclusions

ePRO Design Best Practice

Very few adverse findings were reported in our review of 101 instrument migrations across 53 studies and a broad range of therapeutic areas. All adverse findings could have been avoided by thorough application of good ePRO design and by selecting solutions with good product design and UI/UX properties. The low proportion of usability findings in our studies likely arises from a number of reasons. The first reason may be the fact that the instruments migrated were deployed on platforms provided by vendors with mature ePRO products that have already been used extensively and, in many cases, subjected to extensive usability testing as part of the product development process. Second, the studies were conducted from 2012 onward and used commercially available smartphones and tablets that a growing number of patients are already familiar with. And finally, the training materials associated with modern ePRO applications are generally fit for purpose. On the basis of our review of cognitive interview and usability studies, we identify a number of elements of ePRO design best practice that should be followed when developing an instrument or migrating an instrument to an electronic format. We detail these here, combined with other published guidance from the Critical Path Institute's ePRO Consortium [5,23]:

1. Provide robust instructions on use. This may include information on how to enter free text and use the number up and down buttons. This could be achieved through a tutorial with example questions using the same widgets as the study instrument(s) or via a simple paper training guide or guided instruction at the clinical site.
2. Ensure font size is suitable, clear, and readable for all questions, response options, and instructional text.
3. Present a single question and response scale option per screen. Both the question and its response scale should be completely visible on a single screen without the need to scroll.
4. Take care not to modify the original instrument text beyond minor changes to adapt to the format. If the display size is such that all instructional text or terminology cannot be

presented alongside the question and response option item, this should be presented on the immediately preceding screen.

5. Ensure the screen area used to select each option is of equal size and that the font and line spacing for each response option are equal (e.g., for NRSs and VRSs).
6. Use indicator arrows to identify the location of anchor text if needed because of the display size available.
7. Present the recall period with each item as opposed to only in initial instrument instructions.
8. If the recall period is described for a series of questions, either repeat this within each question prompt or present the instructional text after every three to five questions.

Ensuring Usability

In terms of usability testing, the main usability features such as navigation are aspects of each product/platform as opposed to being instrument-specific. Our findings would support the idea that once usability has been demonstrated for well-developed products/platforms, across the range of instrument response scale types, and within a sample representative of the target population, then there is little value in repeating it for multiple studies.

This brings us to the question of what form of usability testing should be required to provide sufficient evidence to meet the evidentiary requirements of most future studies. The ISPOR Task Force guidance states that usability testing should be performed in the target patient population for all migrations requiring minor instrument modification. Our review covered a broad range of patient populations, and all usability findings could be resolved through application of good product design and the ePRO design best practice principles mentioned earlier. On the basis of this, we recommend that usability testing can be generalized on the basis of testing in "representative groups" as opposed to each target patient population. Testing should include the range of response scale types required by most instruments, in particular the ones identified in our synthesis. Representative groups should include patients or healthy volunteers of a similar age range to the target population, and from a range of educational backgrounds and socioeconomic status, and may include any of the following additional representative groups when appropriate to the target population: children/adolescents, dexterity-challenged subjects, technology-naive subjects (e.g., very elderly subjects), cognitively challenged subjects, and partially sighted subjects. Additional testing is recommended, however, when populations cannot be adequately generalized by representative groups, when instruments contain response scale types not included in previous usability testing, or when new platform releases impact the instrument display and navigation properties.

Expert Screen Review

The ISPOR ePRO Good Research Practices Task Force recommendations [1] have been enormously valuable in providing a clear and scientific approach to instrument migration from paper to electronic formats and have encouraged the continued uptake of ePRO in clinical trials. Nevertheless, the accumulating experience of validation studies since the publication of the recommendations in 2009, for migrations involving both minor and moderate changes, provides additional scientific evidence and practical experience, indicating an opportunity to reflect and reconsider some of the current migration validation recommendations. With the benefit of this accumulating evidence we believe it is possible to relax the need to routinely conduct cognitive interview studies when implementing minor changes during instrument migration. We recommend that an expert screen review is sufficient in most instances. On the basis of our synthesis, we recommend

that an expert screen review should cover the following broad areas, in addition to individual instrument author requirements:

1. *Overall instructional information*: assessment of the completeness of instructional information (including device usage), ensuring faithful representation of the original instrument instructions, ability to remember instructions when a single question per screen is displayed, recall period representation, and adherence to specific author requirements including display of copyright information.
2. *Usability, including font size and navigation*: assessment of the overall usability of the solution, with the target patient population(s) in mind including clarity, navigation, font size, and device usability. This assessment should include, but not be limited to the following:
 - instructions specific to device usage;
 - screen-by-screen review of clarity and font size, type, and color;
 - consistency, visibility, and size of controls during navigation;
 - screen layout changes resulting in changes in device orientation, where enabled;
 - clarity of content and operation of an end of questionnaire review screen, where implemented.
 - Ability to go back to previous instrument items.
3. *Item-by-item migration review*: display clarity; the need to scroll; ease of navigation; question skipping options (when appropriate); clarity of recall period; size, and consistency of response options; placement of anchor text; and language modifications made. This assessment should include, but not be limited to the following:
 - single instrument item (question and response options) and navigation controls clearly visible on each screen;
 - faithful representation of recall period;
 - no changes introduced to the core wording of questions and their response options;
 - consistent use of bold and underlining, when used in the original version (where possible);
 - question skipping capability consistent with the requirements of the instrument authors;
 - VRS: equally spaced response options and equal font size and screen area used by each response item;
 - VAS: sufficient space at scale edges to enable the full range of values to be selected, and anchor text is located so that it is clear which position on the scale it corresponds to;
 - NRS: responses are equally sized and spaced, have sufficient size to enable easy selection, and, where appropriate, anchor text is located so that it is clear which position on the scale it corresponds to.

The implementations in our review investigated the use of a single specific device within each study, and devices ranged from smartphones to tablets and PCs. There is a growing interest in leveraging the patients' own mobile devices and hardware for use in clinical trials ePRO (bring your own device [BYOD]) because this may reduce cost and logistical issues and improve convenience and usability (familiarity) for the patient. Improved convenience may also impact data quality and ePRO compliance. Because of the breadth of studies included, and the small number of findings from cognitive interviews, we believe that our findings extend to the BYOD setting if design best practice can be ensured on all devices. This assurance may be achieved by expert screen review for a minimum screen size/resolution for the chosen instruments and by ensuring that patients without the approved minimum specification would need to be provided with a suitable device. In addition, it is recommended that a BYOD ePRO solution should not permit the user to over-ride specific app display settings including the font size

and the language of the instrument and the rotation properties of the screen during instrument completion.

The findings of all reviews should be considered in the light of potential bias toward publication of positive findings that may affect the robustness of conclusions. Our review of 53 cognitive interview and usability studies is not subject to publication bias because these represent the full number of studies conducted by the CRO between 2012 and 2015.

As an industry we continue to learn and adapt our processes on the basis of increased understanding and accumulating evidence. We welcome the relaxation of hurdles that may limit the uptake and use of ePRO in clinical trials, when robust scientific evidence supports this.

REFERENCES

- [1] Coons SJ, Gwaltney CJ, Hays RD, et al. 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 Health* 2009;12:419–29.
- [2] Beatty PC. Cognitive interviewing: the use of cognitive interviews to evaluate ePRO instruments. In: Byrom B, Tiplady B, eds. *ePRO: Electronic Solutions for Patient-Reported Data*. London: Gower, 2010. p. 23–49.
- [3] Muehlhausen W, Doll H, Quadri N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health Qual Life Outcomes* 2015;13:167–87.
- [4] Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 2008;11:322–33.
- [5] Critical Path Institute ePRO Consortium. Best practices for electronic implementation of patient-reported outcome response scale options. Available from: <https://c-path.org/programs/epro/#section-5648>. [Accessed February 1, 2017].
- [6] The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- [7] Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–36.
- [8] Juniper EF, O'Byrne PM, Guyatt GH, et al. Development and validation of a questionnaire to measure asthma control. *Eur Respir J* 1999;14:902–7.
- [9] Juniper EF, Guyatt GH, Epstein RS, et al. Evaluation of impairment of health-related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax* 1992;47:76–83.
- [10] Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organisation for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;85:365–76.
- [11] Ware JE, Kosinski M, Keller SD. *SF 36 Physical and Mental Health Summary Scales: A User's Manual*. Boston, MA: The Health Institute, New England Medical Center, 1994.
- [12] Roos EM, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes* 2003;1:64–71.
- [13] Revicki DA, Leidy NK, Brennan-Diemer F, et al. Integrating patient preferences into health outcomes assessment: the multiattribute asthma symptom utility index. *Chest* 1998;114:998–1007.
- [14] Jones PW, Harding G, Berry P, et al. Development and first validation of the COPD Assessment Test. *Eur Respir J* 2009;34:648–54.
- [15] Jones PW, Quirk FH, Baveystock CM, et al. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992;145:1321–7.
- [16] Bellamy N. *WOMAC Osteoarthritis Index User Guide (Version V)*. Brisbane, Australia, 2002.
- [17] Irvine EJ. Development and subsequent refinement of the inflammatory bowel disease questionnaire: a quality-of-life instrument for adult patients with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 1999;28:S23–7.
- [18] Sprangers MAG, Groenvold M, Arraras JI, et al. The European Organisation for Research and Treatment of Cancer: Breast Cancer Specific Quality of Life Questionnaire Module: first results from a three-country field study. *J Clin Oncol* 1996;14:2756–68.
- [19] Juniper EF, O'Byrne PM, Ferrie PJ, et al. Measuring asthma control: Clinic questionnaire or daily diary? *Am J Respir Crit Care Med* 2000;162:1330–4.

- [20] Lee KK, Berman N, Alexander GM, et al. A simple self-report diary for assessing psycho-sexual function in hypogonadal men. *J Androl* 2003;24:688–98.
- [21] Bradley C. The Diabetes Treatment Satisfaction Questionnaire (DTSQ): change version for use alongside status version provides appropriate solution where ceiling effects occur. *Diabetes Care* 1999;22:530–2.
- [22] Okamura K, Kimura K, Mizuno H, et al. Core lower urinary tract symptom score questionnaire: a psychometric analysis. *Int J Urol* 2014;21:1151–4.
- [23] Critical Path Institute ePRO Consortium. Best practices for migrating existing patient-reported outcome instruments to a new data collection mode. Available from: <https://c-path.org/programs/e-pro/#section-5648>. [Accessed February 1, 2017].