Does the Type of Device Affect Usability or Conceptual Understanding of Patient-Reported Outcomes? Qualitative Evidence and Implications for Bring-Your-Own-Device (BYOD)

Dumais KM[1], Cascade E[1], Gary ST[1], Otero AV[1]

[1]ERT, Medford, MA, USA

ert.com

PRESENTED AT:

INTRODUCTION

Today, approximately 50% of clinical trials capture clinical outcomes assessments electronically (eCOA). As more clinical trials are using eCOA, patient-reported outcome (PRO) measures are being collected via a range of device types across and within studies. For the initial paper to electronic migration, cognitive debriefing and usability testing (CD/UT) is traditionally recommended to ensure the meaning of assessments are not altered.[1] However, what happens when we move between different electronic executions is less clear.

While historically, most studies used a single provisioned device so it was a single electronic execution, market trends demonstrate that the majority of studies will use more than one device. Changes in the Global Supply market, advances in hardware technology, and variation in country-specific import rules, are all driving the need for use of more than one device type for PRO collection across different sites and countries. With the desire to design more patient-centric trials and increased use of Bring-Your-Own-Device (BYOD) where trial participants complete PRO measures using their personal device, even more substantial variation in devices will be introduced.

Although there is some variation across license holders, a significant portion request CD/UT for each electronic device used in the clinical trial. The burden of continued CD/UT continues to increase with the need for multiple provisioned devices on a study, and conducting CD/UT for every device used in a study becomes impossible to execute for BYOD studies where patients use their own phone in the study.

METHODS

To help provide insight into the necessity for continued CD/UT testing across multiple device types / operating systems, ERT synthesized results from nine conceptual equivalence and usability studies conducted internally by ERT between 2017-2021. Of the nine studies, four studies included in-person qualitative interviews with adult participants and five used expert screen review. Each study included comparisons of different electronic modalities (handhelds, tablets, laptop, or desktop personal computer [PC]) that included generic pain or migraine diaries with common response-scale types, such as the Numerical Rating Scale (NRS), Visual Analog Scale (VAS), or Verbal Rating Scale (VRS).

For participant cognitive interviews, all studies were IRB-approved. Each participant completed the diary on each device sequentially, and without assistance from the interviewer as if they were participating in a clinical trial. The device order was randomized amongst the participants. After participants completed the diary on all devices independently, a semi-structured interview was conducted. The interview included yes-or-no as well as open-ended questions to collect qualitative data from the participant's perspective for each screen in the questionnaire. This allowed for both structured and spontaneous feedback from each participant during the interview. Each item was presented to participants on all devices, side-by-side, and participants were asked the following questions:

1. Is there anything [between the two] or [among the three] versions that would affect how you think about the question?

2. Is there anything [between the two] or [among the three] versions that would affect the answer you would choose? (if yes, please describe.)

Participants were also asked questions pertaining to the readability and usability of the devices.

For expert screen reviews, eCOA senior scientific advisors conducted expert screen reviews following methodology previously described.[2-3] Briefly, all devices were placed side-by-side and screens from generic pain diaries were reviewed. Screens were reviewed for direct comparison of the question and response options layout and formatting to assess whether items on each device were reasonably equivalent, and that the understanding and response to the items was not affected by the change in device (conceptually equivalent). Usability testing, including assessment of readability, was also conducted.

For this synthesis, dimensions for each device were compared and the largest to smallest difference in dimensions and display sizes compared was calculated. Participant and expert determination of conceptual equivalence, readability and usability were also compared across studies.

RESULTS

The nine studies are summarized in Table 1. The four participant cognitive interview studies each included 10-14 adult participants with chronic pain or migraine/headache. Demographics of the participants included in the cognitive interview studies are included in Table 2.

Table 1.

Table 1 – Summary of studies included in the analysis

Study	Device Type (#)	Analysis Type	Method	Assessment	Population
1	Handhelds (2)	UT+CET	Participant cognitive interviews	Pain & Migraine Diary	Adults with chronic pain (n=10) or migraine (n=4)
2	Handhelds (3)	UT+CET	Participant cognitive interviews	Migraine Diary	Adults with chronic migraine or headache (n=10)
3	Handhelds (3)	UT+CET	Participant cognitive interviews	Pain Diary	Adults with chronic pain (n=10)
4	Desktop PC (1), Laptop PC (1), Tablet (1)	UT+CET	Participant cognitive interviews	Pain Diary	Adults with chronic pain (n=10)
5	Handhelds (2)	Spec review, UT+CET	Expert Screen Review	Pain Diary	N/A
6	Handhelds (4)	Spec review, UT+CET	Expert Screen Review	Pain Diary	N/A
7	Tablets (2)	Spec review, UT+CET	Expert Screen Review	Pain Diary	N/A
8	Tablets (2)	Spec review, UT+CET	Expert Screen Review	Pain Diary	N/A
9	Tablets (2)	Spec review, UT+CET	Expert Screen Review	Pain Diary	N/A

Abbreviations: CET = Conceptual Equivalence Testing; N/A = Not Applicable; PC = Personal Computer; Spec = Specification; UT = Usability Testing

Table 2.

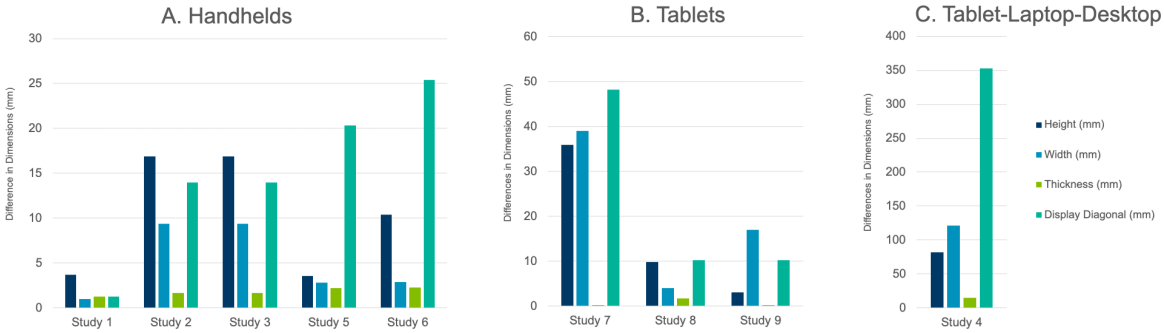
Table 2 – Demographics of participants included in cognitive interview studies

Gender	N (%)
Female	29 (66%)
Male	15 (34%)
Age (years)	
20-29	2 (5%)
30-39	1 (2%)
40-49	6 (14%)
50-59	14 (32%)
60-70	20 (45%)

70+	1 (2%)
Education	
High School Graduate	8 (18%)
Some College or Technical Degree	18 (41%)
College Degree	12 (27%)
Advanced Degree	4 (9%)
Did Not Respond	2 (5%)
Household Income	
Less than $20,000	20 (45%)
Between $20,000 and $49,999	14 (32%)
Between $50,000 and $99,999	4 (9%)
Did not respond	5 (11%)
I don't know	1 (2%)
Occupational Status	
Employed part-time	14 (32%)
Employed full-time	3 (7%)
Retired	10 (23%)
On Disability	9 (20%)
On Leave of Absence	1 (2%)
Unemployed	6 (14%)
Homemaker	1 (2%)

For each study, the largest to smallest difference in device height, width, thickness, and diagonal display size were compared (Figure 1). For handheld devices, Study 6 included the largest difference in screen display sizes at a difference of 25 mm from the smallest device screen display to the largest device screen display (Figure 1A). For tablet devices, Study 7 included the largest difference in screen display sizes at a difference of 48 mm (Figure 1B). The largest difference in screen display sizes of all studies was Study 4 at a difference of 353 mm (Figure 1C). This is expected, since Study 4 included comparisons of a tablet to a laptop and desktop monitor.

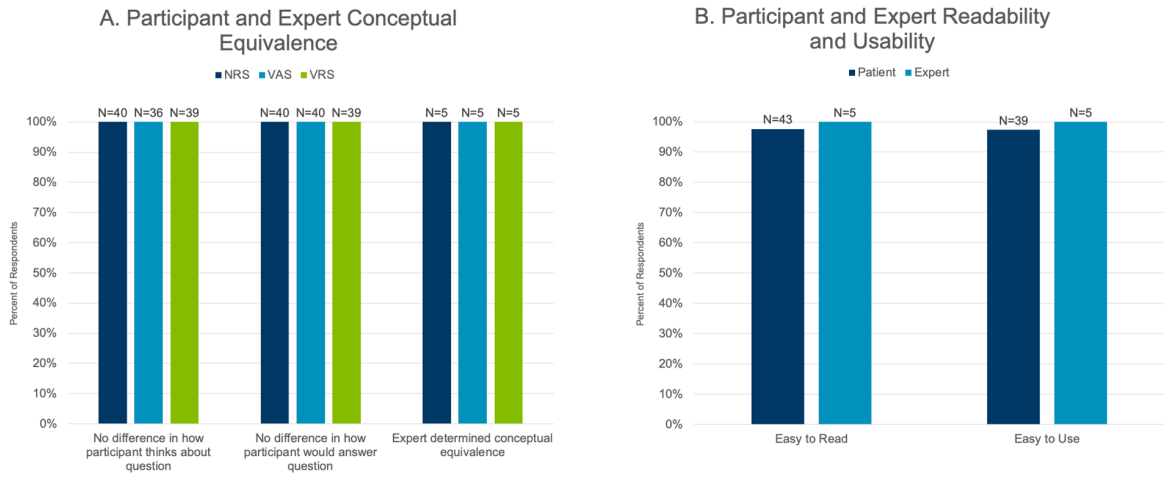
Figure 1.



Comparison of the largest to smallest difference in device height (mm), width (mm), thickness (mm), and diagonal display size (mm) for studies comparing handheld devices (A), tablet devices (B), and for a study comparing a tablet, laptop PC, and desktop PC (C). For desktop PC (Study 4), only display size was included in the comparison. Height, width and thickness were not included since a desktop PC is not intended to be portable.

In response to the questions: “Is there anything [between the two] or [among the three] versions that would affect how you think about the question?” or “Is there anything [between the two] or [among the three] versions that would affect the answer you would choose?”, 100% of participants responded “No” when asked to compare NRS, VAS, and VRS on the different devices (Figure 2A). Similarly, 100% of expert screen reviews found the devices included in the studies to be conceptually equivalent (Figure 2A). All qualitative interview studies and expert screen reviews found that the devices were easily and similarly usable and readable by participants, with 98% of participants reporting devices were easy to read, and 97% of participants reporting devices are easy to use (Figure 2B). Representative participant quotes and expert findings are provided in Table 3.

Figure 2.



A. For each response-scale type (NRS, VAS, VRS), participants were asked: “Is there anything [between the two] or [among the three] versions that would affect how you think about the question?” or “Is there anything [between the two] or [among the three] versions that would affect the answer you would choose?” The graph represents the percent of participants that responded “No”, or the percent of eCOA experts that determined each response-scale type was conceptually equivalent across devices compared.

B. The graph represents the percent of respondents (participant or expert) that responded the devices were easy to read or easy to use.

Table 3.

Table 3 – Representative participant quotes and expert findings

	Participant Quotes		Expert Findings	
	Study 1	Study 3	Study 6	Study 8
NRS	"To me there's no difference in the way that the question or the way it was presented or anything. To me it's the same."	"It's not like they're asking me different questions. They're all the same."	"The NRS are similar on all handhelds…"	"The NRS are similar on both tablets…"
VAS	"No, I would answer it the same."	"No, my answer was the same."	"The vertical VAS is smaller on [handheld 1] compared to [handheld 2, 3 and 4]. These differences in VAS length are unlikely to affect equivalence as equivalence between visual analog scales with much greater differences has previously been shown."	"The vertical VAS appears approximately the same length on both tablets… The horizontal VAS is similar in length on both tablets although very slightly smaller on [tablet 2]. This very slight difference in length of the VAS on both tablets is unlikely to affect equivalence as equivalence between visual analog scales with much greater differences has previously been shown."
VRS/Likert	"I would think the same thing with the second version as I did with the first."	"Nope, cause they're asking me the same thing in all questions. None of them are different."	"The VDS or Likert scales appear similar…"	"The spacing and layout of questions are very similar…"
Selectable Image	N/A	N/A	The selectable graphic of a body diagram... appears similar between all handhelds, however, the need for scrolling may be different between the devices due to differences in screen sizes."	"The single-select graphic of a body diagram ... appear similar between the two tablets although they are slightly smaller on [tablet 2]"
Number Pad	N/A	N/A	"The layout of the number pad is similar between all devices."	"The number pad is slightly different between [tablet 1] and [tablet 2]...However, it is important to note that the user is not permitted to enter the symbols. On both tablets only the numbers 0 through 9 are permitted for entry..."
Date/Time	N/A	N/A	"The layout of the calendar and clock function for setting the date/time are the same…"	"The calendar and clock function for setting the date/time appear identical between the tablets."
Readability	"Even with my eyesight, they were easy to read."	"It was easy to read."	"The font size is easily readable on all four devices."	"The font size is the same between [tablet 1] and [tablet 2] (14 pixels). There may be minor differences in actual font size due to the small difference in pixel density, although it is not perceptible and the text size appears identical." "It was easy to read... on both [tablets]."
Usability	"They were very easy."	"I think it was very easy."	"All devices were similarly responsive to tapping the touch screen to select a response option, move along a VAS, scroll down, and move forward and back. It was easy to read and navigate screens on the [handhelds]...it took a similar amount of time to respond to questions on all four handhelds. Overall, usability testing found that [handheld 1, handheld 2, handheld 3, and handheld 4] were very easy to use."	"Both [tablets] were similarly responsive to tapping the touch screen to select a response option, move along a VAS, scroll down, and move forward and back.....It was easy to read and navigate screens on both [tablets]....it took a similar amount of time to respond to questions on these tablets. Overall, usability testing found that both [tablet 1] and [tablet 2] were very easy to use."

For each response-scale type (NRS, VAS, VRS), participants were asked: "Is there anything [between the two] or [among the three] versions that would affect how you think about the question?" or "Is there anything [between the two] or [among the three] versions that would affect the answer you would choose?". To assess readability and usability, participants were asked "Were the devices easy or difficult to read?" and "Were the devices easy or difficult to use?"
Abbreviations: N/A = Not Applicable; NRS = Numerical Rating Scale; VAS = Visual Analog Scale; VRS = Verbal Rating Scale

In summary, participants reported no differences in their understanding of or user experience with the diary content. All expert screen review studies similarly concluded no differences that would change the meaning of diary items, and that cognitive understanding and diary responses were maintained between devices.

CONCLUSIONS

This data supports conceptual equivalence and usability of common response scale types (NRS, VAS, VRS) across various devices, demonstrating that device type would not influence how participants think about or answer items. This is in line with mounting evidence in the literature that supports equivalency of PRO measures across devices and modes of administration. This evidence includes two quantitative meta-analyses which concluded that the mode of PRO administration (paper vs. electronic) does not impact measurement equivalence, and a randomized trial reporting equivalence of the most common PRO response scale types using BYOD.[4-6] A previous qualitative synthesis of cognitive debrief and usability studies further concluded that the industry can start relaxing the need to routinely conduct cognitive interviews and usability studies when migrating to electronic systems, as long as eCOA best practices are followed and user experience properties have been assessed in a representative group.[2]

Therefore, together with existing evidence, our results suggest that conducting routine CD/UT for each device may not be necessary (or even feasible in the case of BYOD) when using proper eCOA migration and best practices. Obtaining an industry consensus on CD/UT requirements for electronic-to-electronic migrations is critically urgent given the increase in home assessments due to the COVID-19 pandemic and resulting uptake in BYOD studies where there is little control over device type.

DISCLOSURES

Funding for all studies included in analysis was provided by ERT. All authors are employees of ERT, a company specializing in clinical trial endpoint collection, including eCOA.

ABSTRACT

OBJECTIVES: Today, around 50% of clinical trials capture clinical outcomes assessments electronically. For the initial paper to electronic migration, cognitive debriefing and usability testing (CD/UT) is typically conducted to ensure the meaning of assessments are not altered. However, what happens when we move between different electronic executions is less clear. While there is some variation across license holders, a significant portion request CD/UT for each electronic device. Whether this is needed, and how it would work using Bring-Your-Own-Device (BYOD) where there is significant variation in devices used, is unknown.

METHODS: To better understand the level of variation between different electronic devices, we synthesized results from 7 conceptual equivalence and usability studies conducted internally by ERT between 2017-2019. To assess equivalency of generic diaries using a range of response scales across different devices (handhelds, tablets, and PCs), 4 studies included in-person qualitative interviews with adult participants and 3 used expert screen review.

RESULTS: All qualitative interview studies found that devices were easily and similarly usable by participants. Participants reported no differences in their understanding of or user experience with the diary content. All expert screen review studies similarly concluded no differences that would change the meaning of diary items, and that cognitive understanding and diary responses were maintained between devices.

CONCLUSIONS: This data supports conceptual equivalence and usability of common diaries across various devices, demonstrating that device type would not influence how participants think about or answer items. Together with mounting evidence in the literature, this suggests that conducting routine CD/UT for each device may not be necessary when using proper eCOA migration and best practices. Obtaining an industry consensus on CD/UT requirements for electronic-to-electronic migrations is critically urgent given the increase in home assessments due to the COVID-19 pandemic and resulting uptake in BYOD studies where there is little control over device type.

REFERENCES

1 Coons et al, 2009. Value in Health 12(4):419-429.

2 Muehlhausen et al., 2018. Value in Health 21:41-48.

3 Gary et al, 2020. Int J Clin Trials 7(4):271-7.

4 Gwaltney et al, 2008. Value in Health 11: 322-333.

5 Muehlhausen et al, 2015. Health Qual Life Outcomes 13:167-187.

6 Byrom et al, 2018. Value in Health 21(5):581-589.