VALUE IN HEALTH 21 (2018) 581–589

Available online at www.sciencedirect.com

ScienceDirect

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

PATIENT-REPORTED OUTCOMES

Measurement Equivalence of Patient-Reported Outcome Measure Response Scale Types Collected Using Bring Your Own Device Compared to Paper and a Provisioned Device: Results of a Randomized Equivalence Trial

Bill Byrom, PhD[1,*], *Helen Doll, DPhil*[1], *Willie Muehlhausen, DVM*[1], *Emuella Flood, BA*[1], *Cater Cassedy, MA*[1], *Bryan McDowell, MBA*[2], *Jeremy Sohn, BA*[2], *Kyle Hogan*[3], *Ryan Belmont, MBA*[3], *Barbara Skerritt*[1], *Marie McCarthy, MBA*[1]

[1]*ICON Clinical Research, USA, UK and Ireland;* [2]*Novartis Pharmaceuticals, Basel, Switzerland;* [3]*Clinical Ink, Winston-Salem, NC, USA*

ABSTRACT

Objectives: The aim of this study was to assess the measurement equivalence of individual response scale types by using a patient reported outcome measure (PROM) collected on paper and migrated into electronic format for use on the subject's own mobile device (BYOD) and on a provisioned device (site device). **Methods:** Subjects suffering from chronic health conditions causing daily pain or discomfort were invited to participate in this single-site, single visit, three-way crossover study. Association between individual item and instrument subscale scores was assessed by using the intraclass correlation coefficient (ICC) and its CI. Participant attitudes toward the use of BYOD in a clinical trial were assessed through use of a questionnaire. **Results:** In this study, 155 subjects (females 83 [54%]; males 72 [46%]) ages 19 to 69 years (mean ± SD: 48.6 ± 13.1) were recruited. High association between the modes of administration (paper, BYOD, site device) was shown with analysis of ICCs (0.79–0.98) for each response scale type, including visual analogue scale, numeric rating scale, verbal response scale, and Likert scale. Of the subjects, 94% (146 of 155) stated that they would definitely or probably be willing to download an app onto their own mobile device for a forthcoming clinical trial. Forty-five percent of subjects felt BYOD would be more convenient compared with 15% preferring a provisioned device (40% had no preference). **Conclusions:** This study provides strong evidence supporting the use of BYOD for PROM collection in terms of the conservation of instrument measurement equivalence across the most widely used response scale types, and high patient acceptance of the approach.

Keywords: electronic patient reported outcomes (ePRO), bring your own device (BYOD), measurement equivalence, patient acceptability.

There is a drive to design more patient-centric trials that make study participation more engaging and convenient. One approach is to leverage patients' own devices to enable the collection of self-report data ("Bring Your Own Device" [BYOD]) because this eliminates the burden of carrying and maintaining a second device for the duration of the study. Migrating an instrument from a paper-and-pencil format into a screen text format qualifies as a modification of the questionnaire that requires evidence to demonstrate that the instrument's measurement properties are unaffected by the change of format [1]. Although there is a growing body of evidence showing the equivalence of patient reported outcomes measures (PROMs) when migrated from the original format to the electronic format [2,3], there is no definitive study demonstrating that variable technical specification of the mobile device used does not affect the measurement properties of the instrument. This trial in patients suffering from diseases causing chronic pain explored the measurement equivalence of a PROM delivered on paper, PROM using a standardized provisioned device, and PROM using the patient's own mobile device (smartphone or tablet).

Methods

Subjects aged 18 to 70 years suffering from a chronic health condition causing daily pain or discomfort were invited to participate. The subjects provided written informed consent to participate, and the study was approved by the Salus Institutional Review Board (Austin, TX).

* *Address correspondence to:* Bill Byrom, ICON Clinical Research, 3rd Floor Marlow International, Parkway, Marlow, Buckinghamshire, SL7 1YL, Buckinghamshire, SL7 1HZ, UK. Tel: +44 (0)7795 090 645.

E-mail: bill.byrom@iconplc.com.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2017.10.008

Patients were requested to complete a PROM on three occasions in random order according to a William's Design balanced for first-order carryover [4]—once using a paper questionnaire, once electronically using a standard device provided by the study site, and a further electronic administration using an app installed on their own mobile device. The mobile app, SureSource Engage, was

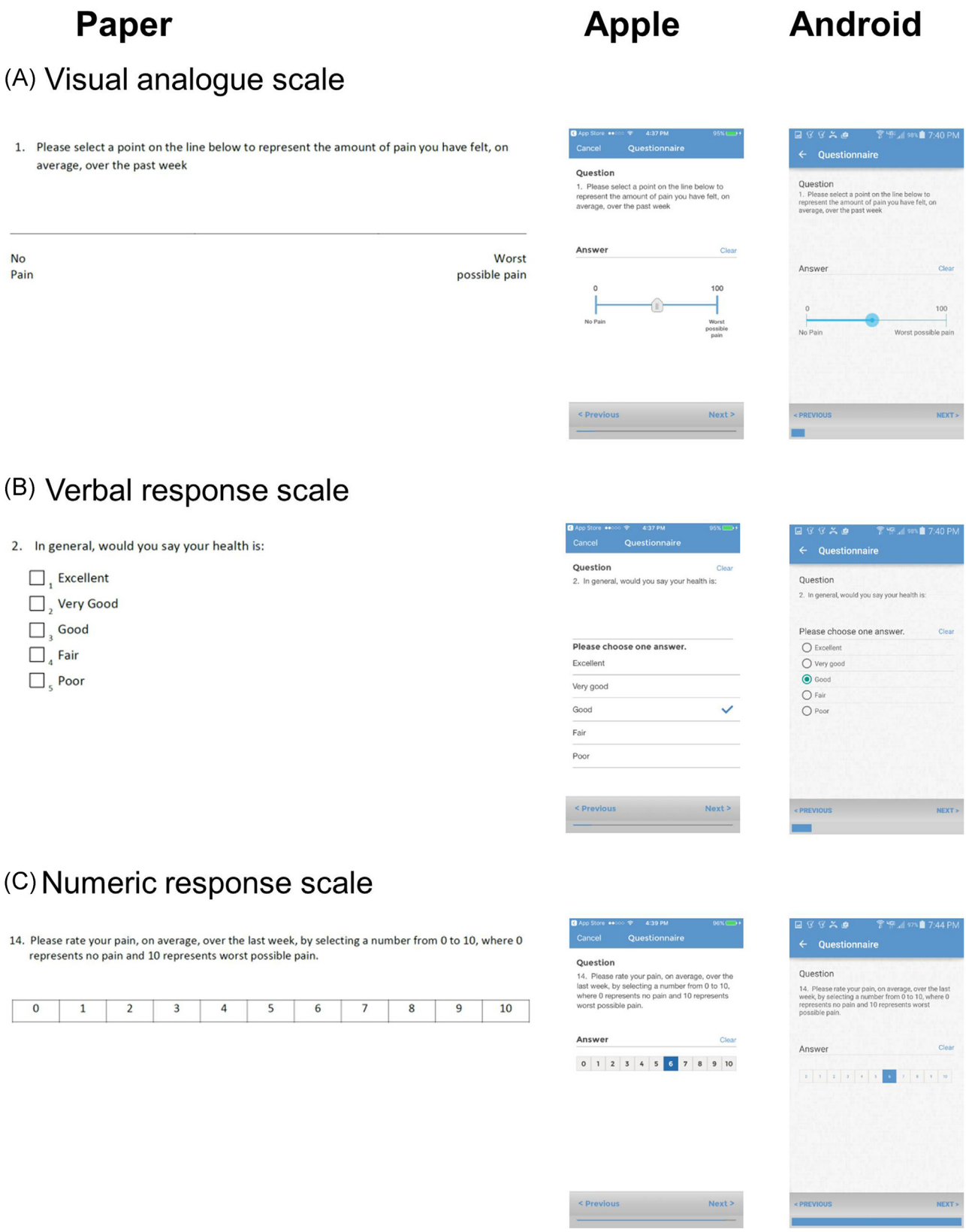

Fig. 1 – Differences in instrument display format between paper and electronic formats.

provided by Clinical Ink (Winston-Salem, NC). PROM administrations were conducted on the same site visit, separated by a 30- to 60-minute washout period in which subjects completed two distraction tasks of attention, processing and working memory using the Paced Visual Serial Addition Test [5] of 60 serial additions; and a spatial memory test of 10 levels of increasing difficulty. Distraction tasks were delivered on an iPad Mini using an Apple Research Kit app developed by ICON Clinical Research and mProve Health (Arlington, VA) and took approximately 10 minutes to complete.

A range of site devices were available: Samsung Galaxy J1 Mini (4-inch screen), Motorola Moto G4 (5.5-inch screen), Apple iPhone 5 (4-inch screen), Apple iPhone 6 plus (5.5-inch screen), and Apple iPad Mini 2 (7.9-inch screen). At the time of the study, the ePRO app was not compatible with Android tablet devices, and therefore an Android device of screen size exceeding 7 inches was not included. Subjects were randomized to use a site device of a different size category to their own mobile device according to a predefined randomization list. We categorized devices as follows: normal: 3- to 5-inch screen, Large 5- to 7-inch screen, X-large: greater than 7-inch screen, broadly in line with Android developer definitions [6]. Where possible, subjects owning Apple devices were provided an Android site device, and vice versa.

The PROM investigated was chosen to comprise a set of commonly used question-and-answer response types, enabling the additional investigation of whether the measurement properties of each response type are maintained independent of the device used. The instrument used in this study was the RAND 20-Item Health Survey (SF-20) 1.0 (RAND Health, Santa Monica, CA) [7], supplemented by a visual analogue scale (VAS) and a numeric response scale (NRS) for pain assessment at the start and end of the questionnaire, respectively. This ensured that questions with the following response types were included: VAS, NRS, verbal response scale (VRS), Yes/No scale, and Likert scale. Differences in visual representation among paper, Apple, and Android administrations are presented in Figure 1 for (1) VAS, (2) five-category VRS (VRS-5), and (3) 11-point NRS response types. The SF-20 items were scored in the way described by the instrument developers, including the calculation of the Physical Functioning, Role Functioning, Social Functioning, Mental Health, Health Perceptions, and Pain subscales [7,8]. The migration to electronic format followed ePRO design good practice guidelines [9,10].

Patient preferences and attitudes toward BYOD use in clinical trials were also collected by using an end-of-study questionnaire.

The target study sample size of 150 was determined on the basis of ensuring at least 80% power, 95% significance, and a true underlying ICC of 0.85 for comparing each BYOD device size category (target: 43 subjects per device size category), assuming that the difference we wish to rule out equates to an effect size of one-quarter of the standard deviation and a lower bound for ICC of 0.70.

Statistical analysis of equivalence was determined by calculation of mean scores along with the intraclass correlation coefficient (ICC [1,2]) derived from a mixed effects model with subject considered a random effect and mode, device size (BYOD and site administration) and administration period as fixed effects [11]. When estimating the impacts of mode, device size, and administration period, the model included factors for baseline/screening characteristics: gender, education, type of health condition, and difficulty washing/dressing. Analysis was conducted in Stata v14 (StataCorp LLC, College Station, TX). Statistical significance was taken at the two-sided 5% level ($P < 0.05$) throughout, although on account of the number of repeated comparisons performed those differences significant at the 1% level ($P < 0.01$) are considered most reliable.

Results

In this study, 156 subjects were enrolled. One subject presented with a device associated with a platform not supported by the study app (Blackberry) and was excluded because the individual chose not to participate in any study period.

The remaining 155 subjects (females 83 [54%]; males 72 [46%]) were ages 19 to 69 years (mean ± SD: 48.6 ± 13.1) and presented with a variety of health conditions that caused some form of pain or discomfort on a daily basis (Table 1). Of the subjects, 42% (65/155) suffered from some form of back pain or joint pain, with a further 15%, 8%, and 7% suffering from rheumatoid arthritis, generalized pain, and fibromyalgia, respectively. The majority of subjects had difficulties walking and performing usual activities or felt anxious or depressed because of their health conditions (62%, 77%, and 77%, respectively). With regard to ethnicity, 70% (109 of 155) of the subjects were white, and 23% (35 of 155) were black. Subjects came from a cross-section of educational backgrounds, as described in Table 1.

Table 1 – Baseline demographics of subjects.

Variable	N = 155
Age	
Range	19 – 69
Mean ± SD	48.6 ± 13.1
Gender	
Female	83 (54%)
Male	72 (46%)
Racial group	
Black	35 (23%)
Asian	3 (2%)
Native American or Alaska Native	1 (1%)
White	109 (70%)
Other	4 (3%)
Declined to answer	3 (2%)
Education	
Did not complete high school/High school diploma	11 (7%)
Some college	32 (21%)
2-year Associate's degree/Technical training/ 4-year Bachelor's degree	76 (49%)
Master's degree/Doctorate/Professional degree	36 (23%)
Disease indication	
Fibromyalgia	11 (7%)
Generalized pain	12 (8%)
Hallux rigidus	2 (1%)
Joint/back pain	65 (42%)
Myalgia	2 (1%)
Neuropathy	7 (5%)
Plantar fasciitis	3 (2%)
Psoriatic arthritis	3 (2%)
Rheumatoid arthritis	24 (15%)
Sciatica	3 (2%)
Scoliosis	6 (4%)
Other	17 (11%)
Difficulty walking	
Yes	96 (62%)
No	59 (38%)
Problems washing or dressing	
Yes	36 (23%)
No	119 (77%)
Problems doing usual activities	
Yes	120 (77%)
No	35 (23%)
Feeling anxious/depressed because of health condition	
Yes	119 (77%)
No	36 (23%)

Sixteen subjects (10%) were unable to download the study app by using their Android or Apple mobile devices and did not complete the BYOD administration (Table 2). The reasons included forgotten App Store credentials (8 of 16), inability to run app or unknown error message (3 of 16), inability to locate the downloaded app on the Android device (3 of 16), insufficient storage space (1 of 16), and having an Android tablet device that was not compatible with the study app (1 of 16). Three of these subjects did not compete the site device administration period and were excluded. A further five subjects provided only one electronic administration, and definitive determination of whether this was completed using site or BYOD device was not possible, so their data were not included in the equivalence comparison. One subject could not log onto the site device by using the ePRO account and did not provide data for the site device administration period.

In summary, 155 subjects completed the study questionnaires regarding attitudes toward BYOD and experience with the app in this study; and 147 subjects provided evaluable data for the equivalence comparison: 133 providing paper, site device, and BYOD administration; one providing only paper and BYOD administration; and 13 providing paper and site device administration only.

Of the subjects, 98 (63%) brought Apple devices, the remaining 57 (37%) brought Android devices. Subjects presented with a range of device sizes: 79 (51%) normal, 52 (34%) large, and 24 (15%) X-large (see Table 2). Larger BYOD device sizes were associated with increased reported difficulty washing/dressing (response to screening question), increasing age, and lower educational attainment.

Table 2 – Summary of BYOD mobile devices and app experience.

Variable	N = 155*
BYOD mobile device type	
Apple	98 (63%)
Android	57 (37%)
BYOD mobile device size	
Normal	79 (51%)
Large	52 (34%)
X-large	24 (15%)
Site mobile device size (randomly assigned to a different category to BYOD)	
Normal	37 (24%)
Large	51 (33%)
X-large	66 (43%)
Familiarity downloading and using apps on their mobile device	
Yes	149 (96%)
No	6 (4%)
BYOD device: ability to download and run study app on own mobile device	
Yes	141 (91%)
No	16 (10%)
Android: Tablet device (not supported)	1 (6%)
Android: Unknown Google Play ID	2 (13%)
Android: Could not locate app after download	3 (19%)
Android: Unknown error message on opening app	2 (13%)
Apple: App would not run	1 (6%)
Apple: Unknown Apple ID	6 (38%)
Apple: Insufficient storage space	1 (6%)

* One additional subject presented with a device of an ineligible platform (Blackberry) and was excluded.

Attitudes toward BYOD

Ninety-six percent of subjects (149 of 155) were familiar with downloading and using apps on their mobile device (see Table 2). Ninety-two percent of subjects (142 of 155) felt that after participating in this study, they could have definitely or probably downloaded the study app at home by using an instruction sheet without additional assistance from the study team; and 94% (146 of 155) stated that they would definitely or probably be willing to download an app on their own mobile device for a forthcoming clinical trial (Figure 2A). When asked if they had any concerns using their own mobile device in a forthcoming trial, 135 subjects (87%) reported no concerns. Of those citing a concern, common concerns reported included uncertainties about the use of personal data (7 of 20) and the effect on available device storage capacity (4 of 20). When considering BYOD, 115 subjects (74%) identified reimbursement for data charges as important, very important, or essential; 78%, 90% and 97% reported ensuring data privacy, ease of installation and use, and no interference with other device functions as important or greater, respectively (Figure 2B). Forty-five percent of subjects felt that using their own device would be more convenient compared with 15% preferring a provisioned device (40% had no preference).

Measurement Equivalence

The primary analysis of ICCs showed very high correlation between the three modes of administration for all question items and response scale types. For each of the 22 items, the overall comparison ICCs ranged from 0.82 to 0.98, with the lower bound of the 95% confidence interval (CI) exceeding 0.75 in all cases (Table 3). Comparing paper and BYOD, ICCs ranged from 0.81 to 0.98, with only one lower bound of the 95% CIs dropping below 0.75 (item 10, 6-category VRS; 95% CI 0.74–0.86). ICCs for paper

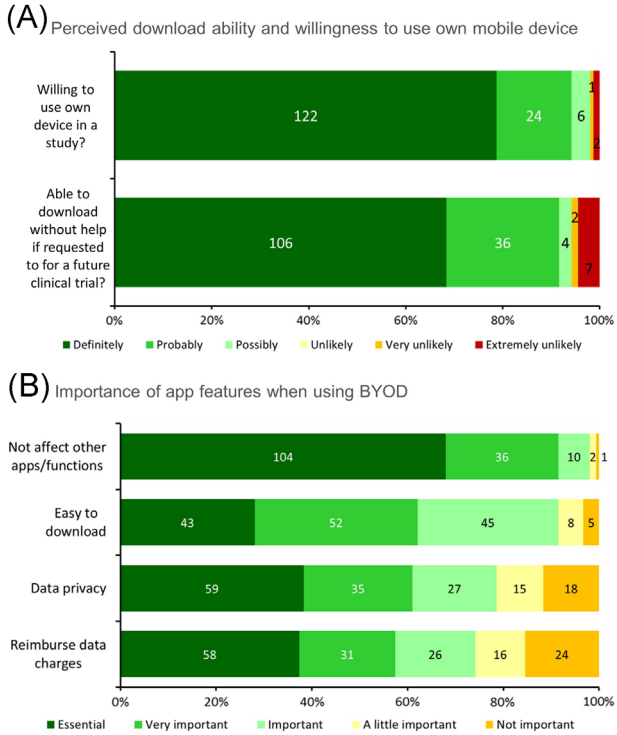

Fig. 2 – Subject attitudes toward bring-your-own-device (BYOD). A, Perceived download ability and willingness to use own mobile device. B, Importance of app features when using BYOD.

Table 3 – Intraclass Correlations for questionnaire items between paper, BYOD and Site device administration.

Item number subscale	Response type*	Paper		BYOD		Site		ICC (95% CI)			
		Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Overall	Paper vs BYOD	Paper vs site	BYOD vs site
1	**VAS**	**58.86 (19.90)**	**155**	**60.38 (19.82)**	**134**	**60.09 (20.76)**	**146**	**0.94 (0.93–0.96)**	**0.93 (0.91–0.95)**	**0.94 (0.92–0.96)**	**0.95 (0.94–0.97)**
2	VRS-5	3.07 (0.99)	155	3.06 (1.00)	134	3.07 (0.97)	146	0.97 (0.97–0.98)	0.97 (0.96–0.98)	0.98 (0.97–0.99)	0.97 (0.95–0.98)
3a	VRS-3	1.21 (0.53)	155	1.19 (0.48)	134	1.21 (0.53)	146	0.86 (0.82–0.89)	0.90 (0.86–0.93)	0.82 (0.76–0.86)	0.88 (0.84–0.92)
3b	VRS-3	1.72 (0.85)	154	1.68 (0.82)	133	1.75 (0.86)	146	0.88 (0.84–0.90)	0.91 (0.88–0.93)	0.88 (0.83–0.91)	0.84 (0.78–0.88)
3c	VRS-3	1.95 (0.95)	154	1.90 (0.92)	134	1.92 (0.91)	146	0.90 (0.88–0.93)	0.94 (0.91–0.96)	0.88 (0.83–0.91)	0.91 (0.87–0.93)
3d	VRS-3	1.80 (0.88)	154	1.80 (0.91)	134	1.73 (0.87)	146	0.82 (0.77–0.86)	0.82 (0.76–0.87)	0.81 (0.74–0.86)	0.82 (0.76–0.87)
3e	VRS-3	2.38 (0.89)	153	2.36 (0.87)	134	2.33 (0.89)	146	0.91 (0.89–0.93)	0.94 (0.91–0.95)	0.91 (0.88–0.94)	0.89 (0.85–0.92)
3f	VRS-3	2.49 (0.83)	154	2.51 (0.83)	134	2.51 (0.81)	146	0.89 (0.86–0.92)	0.91 (0.87–0.93)	0.87 (0.82–0.90)	0.91 (0.88–0.94)
3a-3f	**VRS-3**	**1.92 (0.60)**	**155**	**1.91 (0.58)**	**134**	**1.91 (0.58)**	**146**	**0.96 (0.94–0.97)**	**0.96 (0.95–0.97)**	**0.95 (0.93–0.96)**	**0.96 (0.94–0.97)**
4	VRS-6	4.26 (0.77)	155	4.22 (0.82)	134	4.21 (0.83)	146	0.91 (0.88–0.93)	0.96 (0.95–0.97)	0.87 (0.83–0.91)	0.91 (0.88–0.94)
5	Y>3, Y<3, N	2.17 (0.96)	155	2.16 (0.96)	134	2.13 (0.96)	146	0.91 (0.89–0.93)	0.93 (0.91–0.95)	0.94 (0.91–0.95)	0.87 (0.82–0.90)
6	Y>3, Y<3, N	1.74 (0.90)	155	1.75 (0.91)	134	1.71 (0.87)	146	0.82 (0.78–0.86)	0.82 (0.76–0.87)	0.82 (0.76–0.87)	0.86 (0.81–0.90)
5-6	**Y>3, Y<3, N**	**1.95 (0.82)**	**155**	**1.95 (0.84)**	**134**	**1.92 (0.93)**	**146**	**0.91 (0.89–0.93)**	**0.93 (0.90–0.95)**	**0.92 (0.90–0.94)**	**0.89 (0.86–0.92)**
7	VRS-6	4.52 (1.45)	155	4.38 (1.52)	134	4.43 (1.51)	146	0.92 (0.89–0.94)	0.91 (0.88–0.94)	0.91 (0.88–0.93)	0.93 (0.90–0.95)
8	VRS-6	4.37 (1.42)	155	4.31 (1.40)	134	4.40 (1.40)	146	0.88 (0.85–0.91)	0.90 (0.86–0.93)	0.88 (0.84–0.92)	0.88 (0.83–0.91)
9	VRS-6	3.61 (1.23)	155	3.67 (1.20)	134	3.67 (1.27)	146	0.88 (0.84–0.91)	0.85 (0.79–0.89)	0.89 (0.85–0.92)	0.90 (0.86–0.93)
10	VRS-6	4.38 (1.24)	155	4.28 (1.26)	134	4.29 (1.31)	146	0.82 (0.77–0.86)	0.81 (0.74–0.86)	0.86 (0.81–0.90)	0.79 (0.72–0.85)
11	VRS-6	3.39 (1.18)	155	3.51 (1.19)	134	3.49 (1.19)	146	0.85 (0.81–0.89)	0.81 (0.75–0.86)	0.91 (0.88–0.94)	0.82 (0.76–0.87)
12	VRS-6	5.04 (1.18)	155	4.99 (1.20)	134	4.99 (1.21)	146	0.90 (0.87–0.92)	0.91 (0.88–0.93)	0.90 (0.86–0.93)	0.89 (0.85–0.92)

continued on next page

Table 3 – *continued*

Item number subscale	Response type*	Paper		BYOD		Site		ICC (95% CI)			
		Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Overall	Paper vs BYOD	Paper vs site	BYOD vs site
4, 7-12	VRS-6	4.01 (0.88)	155	3.94 (0.88)	134	3.96 (0.92)	146	0.96 (0.95–0.97)	0.95 (0.94–0.97)	0.96 (0.95–0.97)	0.96 (0.95–0.97)
2, 3a-3f, 4, 7-12	VRS	3.04 (0.62)	155	3.00 (0.63)	134	3.01 (0.65)	146	0.97 (0.96–0.98)	0.97 (0.96–0.98)	0.97 (0.95–0.97)	0.97 (0.96–0.98)
13a	Likert-5	2.84 (1.40)	154	2.84 (1.37)	134	2.78 (1.33)	146	0.88 (0.85–0.91)	0.89 (0.85–0.92)	0.87 (0.83–0.91)	0.90 (0.86–0.93)
13b	Likert-5	3.38 (1.23)	152	3.40 (1.20)	134	3.43 (1.17)	146	0.91 (0.88–0.93)	0.92 (0.89–0.94)	0.92 (0.90–0.94)	0.89 (0.85–0.92)
13c	Likert-5	3.45 (1.22)	152	3.51 (1.20)	134	3.44 (1.24)	146	0.91 (0.88–0.93)	0.89 (0.85–0.92)	0.92 (0.89–0.94)	0.91 (0.88–0.94)
13d	Likert-5	2.59 (1.24)	152	2.51 (1.17)	134	2.56 (1.17)	146	0.85 (0.81–0.88)	0.85 (0.80–0.89)	0.84 (0.79–0.88)	0.89 (0.85–0.92)
13a-d	Likert-5	2.63 (1.04)	155	2.61 (1.03)	134	2.62 (0.99)	146	0.96 (0.94–0.97)	0.96 (0.94–0.97)	0.96 (0.94–0.97)	0.96 (0.94–0.97)
14	NRS-11	6.17 (1.94)	154	6.16 (1.97)	134	6.16 (1.93)	146	0.98 (0.97–0.98)	0.98 (0.97–0.98)	0.98 (0.97–0.99)	0.98 (0.97–0.98)
	Average ICC across individual items							0.89	0.90	0.89	0.89
Physical Function	VRS-3	46.29 (30.04)	154	45.31 (29.22)	134	45.38 (29.05)	146	0.96 (0.94–0.97)	0.96 (0.95–0.97)	0.95 (0.93–0.96)	0.96 (0.94–0.97)
Role Function	Y>3, Y<3, N	47.74 (41.13)	155	47.57 (42.13)	134	45.89 (41.42)	146	0.91 (0.89–0.93)	0.93 (0.90–0.95)	0.92 (0.90–0.94)	0.89 (0.86–0.92)
Social Function	VRS-6	70.32 (28.95)	155	67.61 (30.32)	134	68.63 (30.16)	146	0.92 (0.89–0.94)	0.91 (0.88–0.94)	0.91 (0.88–0.93)	0.93 (0.90–0.95)
Mental Health	VRS-6	63.15 (20.56)	155	61.61 (19.93)	134	62.11 (21.18)	146	0.96 (0.94–0.97)	0.95 (0.93–0.96)	0.96 (0.95–0.97)	0.96 (0.94–0.97)
Health Perceptions	VRS/Likert-5	44.16 (25.04)	152	43.11 (25.19)	134	43.26 (24.03)	146	0.97 (0.96–0.98)	0.97 (0.96–0.98)	0.97 (0.96–0.98)	0.97 (0.96–0.98)
Pain	VRS-6	34.71 (15.47)	155	35.67 (16.34)	134	35.89 (16.60)	146	0.91 (0.88–0.93)	0.96 (0.95–0.97)	0.87 (0.83–0.91)	0.91 (0.88–0.94)

BYOD, bring your own device; CI, confidence interval; ICC, intraclass correlation coefficient; SD, standard deviation.
* VAS, visual analogue scale; VRS-n, n-category verbal response scale; Likert-5, 5-category Likert scale; NRS-11, 11-point numeric rating scale.

versus site device ranged from 0.81 to 0.98, with the lower limit of the 95% CI dropping below 0.75 for a single item (item 3d, 3-category VRS; 95% CI 0.74–0.86) . ICCs for BYOD vs site device ranged from 0.79 to 0.98, with the lower limit of the 95% CI dropping below 0.75 for a single item (item 10, 6-category VRS; 95% CI 0.72–0.85). All ICCs for SF-20 instrument subscores exceeded 0.87, with CI lower limits exceeding 0.75.

Mean differences for each questionnaire item, SF-20 subscales, and mean scores across questions of common response type are presented in Table 4. Although there was a suggestion that the mean VAS scores were a little lower on paper compared with the electronic administrations ($P = 0.035$), the numerical difference was small, the difference did not reach the 1% level of significance, and the values measured using each mode still showed high agreement, with lower limit of the 95% CI for ICC well above the required threshold of 0.7 [1]. Similarly, there was a suggestion of lower scores on BYOD for item 2 (5-category VRS, $P = 0.025$) and for both electronic administrations for item 4 (6-category VRS, $P = 0.039$). Again, these numerically small and nonsignificant at the 1% level, and in both cases with the ICCs showing high agreement in the scores recorded, with lower limits of the CI for ICC above the required threshold of 0.7.

No differences as a result of BYOD size category were observed on the basis of the mean differences analysis (see Table 4).

Table 4 – Mean differences and overall factor effects for questionnaire items between paper, BYOD and Site device administration.

Item number / subscale	Response type*	Modality comparison							Device size and period comparisons		
		Paper		BYOD		Site		P-value	Regression coefficient (*P* value)†		
		Mean (SD)	n	Mean (SD)	n	Mean (SD)	n		BYOD size	Site size	Period
1	**VAS**	**58.86 (19.90)**	**155**	**60.38 (19.82)**	**134**	**60.09 (20.76)**	**146**	**0.035**	**–0.265 (0.915)**	**0.212 (0.926)**	**0.916 (0.001)**
2	VRS-5	3.07 (0.99)	155	3.06 (1.00)	134	3.07 (0.97)	146	0.025	–0.008 (0.949)	0.066 (0.552)	–0.022 (0.021)
3a	VRS-3	1.21 (0.53)	155	1.19 (0.48)	134	1.21 (0.53)	146	0.428	–0.056 (0.357)	0.068 (0.228)	0.003 (0.788)
3b	VRS-3	1.72 (0.85)	154	1.68 (0.82)	133	1.75 (0.86)	146	0.101	0.063 (0.489)	–0.018 (0.836)	–0.004 (0.829)
3c	VRS-3	1.95 (0.95)	154	1.90 (0.92)	134	1.92 (0.91)	146	0.790	–0.083 (0.438)	–0.230 (0.019)	0.008 (0.635)
3d	VRS-3	1.80 (0.88)	154	1.80 (0.91)	134	1.73 (0.87)	146	0.324	–0.017 (0.865)	–0.081 (0.367)	0.056 (0.013)
3e	VRS-3	2.38 (0.89)	153	2.36 (0.87)	134	2.33 (0.89)	146	0.689	0.015 (0.879)	–0.103 (0.259)	0.023 (0.139)
3f	VRS-3	2.49 (0.83)	154	2.51 (0.83)	134	2.51 (0.81)	146	0.093	–0.019 (0.817)	–0.097 (0.209)	–0.015 (0.354)
3a–3f	**VRS-3**	**1.92 (0.60)**	**155**	**1.91 (0.58)**	**134**	**1.91 (0.58)**	**146**	**0.246**	**–0.016 (0.797)**	**–0.077 (0.185)**	**0.011 (0.123)**
4	VRS-6	4.26 (0.77)	155	4.22 (0.82)	134	4.21 (0.83)	146	0.039	–0.099 (0.309)	–0.058 (0.517)	0.044 **(0.002)**
5	Y>3, Y<3, N	2.17 (0.96)	155	2.16 (0.96)	134	2.13 (0.96)	146	0.523	0.145 (0.177)	0.036 (0.719)	–0.013 (0.430)
6	Y>3, Y<3, N	1.74 (0.90)	155	1.75 (0.91)	134	1.71 (0.87)	146	0.883	–0.058 (0.557)	0.049 (0.597)	–0.012 (0.578)
5, 6	**Y>3, Y<3, N**	**1.95 (0.82)**	**155**	**1.95 (0.84)**	**134**	**1.92 (0.93)**	**146**	**0.857**	**0.044 (0.638)**	**0.042 (0.623)**	**–0.013 (0.362)**
7	VRS-6	4.52 (1.45)	155	4.38 (1.52)	134	4.43 (1.51)	146	0.127	0.144 (0.395)	–0.019 (0.905)	0.017 (0.511)
8	VRS-6	4.37 (1.42)	155	4.31 (1.40)	134	4.40 (1.40)	146	0.659	–0.114 (0.485)	–0.289 (0.054)	0.044 (0.118)
9	VRS-6	3.61 (1.23)	155	3.67 (1.20)	134	3.67 (1.27)	146	0.215	–0.055 (0.711)	–0.134 (0.332)	0.134 (<0.001)
10	VRS-6	4.38 (1.24)	155	4.28 (1.26)	134	4.29 (1.31)	146	0.276	–0.081 (0.581)	–0.245 (0.072)	0.147 (<0.001)
11	VRS-6	3.39 (1.18)	155	3.51 (1.19)	134	3.49 (1.19)	146	0.085	0.043 (0.763)	–0.084 (0.522)	0.088 (0.001)
12	VRS-6	5.04 (1.18)	155	4.99 (1.20)	134	4.99 (1.21)	146	0.612	0.034 (0.813)	–0.199 (0.134)	–0.011 (0.634)
4, 7–12	**VRS-6**	**4.01 (0.88)**	**155**	**3.94 (0.88)**	**134**	**3.96 (0.92)**	**146**	**0.123**	**–0.017 (0.873)**	**–0.147 (0.141)**	**0.066 (<0.001)**
2, 3a–3f, 4, 7-12	**VRS**	**3.04 (0.62)**	**155**	**3.00 (0.63)**	**134**	**3.01 (0.65)**	**146**	**0.401**	**–0.014 (0.851)**	**–0.099 (0.142)**	**0.039 (<0.001)**
13a	Likert-5	2.84 (1.40)	154	2.84 (1.37)	134	2.78 (1.33)	146	0.937	–0.001 (0.997)	–0.028 (0.857)	–0.071 (0.011)
13b	Likert-5	3.38 (1.23)	152	3.40 (1.20)	134	3.43 (1.17)	146	0.498	–0.108 (0.473)	–0.054 (0.696)	–0.008 (0.712)
13c	Likert-5	3.45 (1.22)	152	3.51 (1.20)	134	3.44 (1.24)	146	0.112	–0.078 (0.600)	–0.078 (0.571)	–0.031 (0.157)
13d	Likert-5	2.59 (1.24)	152	2.51 (1.17)	134	2.56 (1.17)	146	0.864	–0.138 (0.322)	–0.108 (0.400)	0.023 (0.397)
13a-d	**Likert-5**	**2.63 (1.04)**	**155**	**2.61 (1.03)**	**134**	**2.62 (0.99)**	**146**	**0.212**	**–0.084 (0.505)**	**–0.069 (0.554)**	**–0.023 (0.074)**
14	**NRS-11**	**6.17 (1.94)**	**154**	**6.16 (1.97)**	**134**	**6.16 (1.93)**	**146**	**0.677**	**–0.029 (0.904)**	**–0.040 (0.859)**	**0.007 (0.679)**

continued on next page

Table 4 – *continued*

Item number / subscale	Response type*	Modality comparison							Device size and period comparisons		
		Paper		BYOD		Site		P-value	Regression coefficient (*P* value)†		
		Mean (SD)	n	Mean (SD)	n	Mean (SD)	n		BYOD size	Site size	Period
Physical Functioning	VRS-3	46.29 (30.04)	154	45.31 (29.22)	134	45.38 (29.05)	146	0.253	–0.800 (0.798)	–3.818 (0.186)	0.574 (0.121)
Role Functioning	Y>3, Y<3, N	47.74 (41.13)	155	47.57 (42.13)	134	45.89 (41.42)	146	0.857	2.179 (0.638)	2.105 (0.623)	–0.662 (0.362)
Social Functioning	VRS-6	70.32 (28.95)	155	67.61 (30.32)	134	68.63 (30.16)	146	0.127	2.876 (0.395)	–0.372 (0.905)	0.338 (0.511)
Mental Health	VRS-6	63.15 (20.56)	155	61.61 (19.93)	134	62.11 (21.18)	146	0.093	–0.664 (0.789)	–3.811 (0.097)	1.611 (<0.001)
Health Perceptions	VRS/Likert-5	44.16 (25.04)	152	43.11 (25.19)	134	43.26 (24.03)	146	0.103	–1.561 (0.612)	–0.914 (0.747)	–0.563 (0.033)
Pain	VRS-6	34.71 (15.47)	155	35.67 (16.34)	134	35.89 (16.60)	146	0.039	–1.982 (0.309)	–1.167 (0.517)	0.884 (0.002)

BYOD, bring your own device; SD, standard deviation.
* VAS, visual analogue scale; VRS-n, n-category verbal response scale; Likert-5, 5-category Likert scale; NRS-11, 11-point numeric rating scale.
† Adjusted for baseline/screening characteristics: age, gender, education, type of health condition, difficulty washing/dressing.

One item (3c, 3-category VRS) indicated minor reductions in mean item score with increasing site device size ($P = 0.019$). All other comparisons across site device size categories were not significant.

Conclusions

This study has been effective in demonstrating the measurement equivalence of the response scale types contained within the SF-20, the SF-20 subscores, and VAS and NRS, between paper and electronic administration using BYOD and a site-provisioned device. In addition, there was no evidence of a difference in measurement properties of each response scale type across different sizes of BYOD devices (screen sizes: normal smartphone: 3–5 inch; large smartphone: 5–7 inch; tablet: >7 inch). Instead of simply considering equivalence by comparison of instrument overall and subscale scores, we explored associations on the basis of a per-item and per-response-scale-type basis. This enabled us to generate measurement equivalence evidence on a response-scale-type basis, which we feel is generalizable to measurement equivalence of other instruments using similar response scales. The purpose of an equivalence study is not to assess the psychometric properties or content validity of the instrument but simply to assess whether the change in format may result in changed responses. We argue, therefore, that the ability of subjects to provide the same responses to the different question-and-answer response formats examined in this study provides strong evidence to infer the acceptability of BYOD for other instruments, assuming that principles of ePRO design good practice are followed—such as those reported by the Critical Path Institute's ePRO Consortium [9,10]. This applicability represents the majority of PROMs. For example, of the 114 instruments included in the meta-analysis of measurement equivalence studies reported by Muehlhausen et al [2], 91 (80%) were composed entirely of verbal response scale, Likert scale, VAS, and NRS items, as examined in this study.

Despite the differences in appearance and operation among paper, Apple, and Android devices, the agreement between modes in VAS scores remained high and above the acceptance level. Between devices, VAS length varied from less than 40 mm in some of the smaller normal-category devices, to over 80 mm on some tablets. This compared with the standard 100 mm line on the paper version. In all cases, the number of horizontal pixels comprising each line enabled an integer score of any value between 0 and 100 to be recorded, in common with the manual measurement of the paper version to the nearest millimeter. These results support currently understood properties of the VAS, whereby scale length is understood not to affect measurement properties (see [12], for example, comparing a 21 mm VAS on a mobile phone to the 100 mm paper version).

Some subjects (16 of 155) in this study were unable to complete the BYOD assessment because of their inability to download or run the app. Half of these subjects (8 of 16) could not remember the App Store login credentials to download the app. As this study was conducted on a single morning or afternoon, these subjects did not have time to establish or reset credentials. In a clinical trial, it is anticipated that subjects could be given instructions and time to rectify this. In addition, three subjects could not locate the app on their Android phones after successful downloading. Again, we feel this could be mitigated by training, information from site staff, or an external support line. In only three cases out of 155 (<2%), the device or operating system was incompatible with the app, and this could not be mitigated in some way; and in another case, device capacity was not sufficient. These figures are supportive in demonstrating the broad applicability of ePRO apps in a BYOD setting.

We acknowledge some limitations in our study. Although the inclusion of an additional period to enable measurement of the test–retest reliability of the paper instrument in this sample would have been helpful, it was not practical to include a fourth study period. That said, the measures of association were overwhelmingly in favor of equivalence. The washout period in this study was 30 to 60 minutes in length, in line with the 15 single-day studies reported in the meta-analysis of Muehlhausen et al. [2], which reported a mean washout period of 39 minutes (range 1–180 minutes). In our study, we used the Paced Visual Serial Addition Test and a spatial memory test to provide an additional distraction task between

administration periods. Both tests require significant attention, processing, and working memory, and we feel that the inclusion of these facilitated the effective washout of subjects between the three modes of administration in this study.

This study supplements earlier meta-analysis work that demonstrated the general equivalence of migrations from paper to a standardized provisioned device type [2,3]. This study extends that evidence to provide the first comprehensive formal assessment of the equivalence of BYOD compared with both paper and standard provisioned devices. Furthermore, our study explores the measurement equivalence of individual response scale types as opposed to total instrument scores. PROMs are constructed from a number of instrument items, representing a collection of common response scale types. Our study has uniquely demonstrated the measurement equivalence of commonly used response scale types when migrated from paper to electronic formats in a BYOD setting. This extends the growing body of evidence supporting the use of BYOD to collect ePRO data in clinical trials. We recommend that researchers and regulatory agencies utilize this evidence when they consider how to apply BYOD models more effectively for the collection of ePRO data to support regulatory drug development programs.

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(4):419–29.

[2] 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.

[3] 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(2):322–33.

[4] Williams EJ. Experimental designs balanced for the estimation of residual effects of treatments. Aust J Sci Res 1949;Ser. A 2:149-68.

[5] Tombaugh TN. A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). Arch Clin Neurophysiol 2006;21(1):53–76.

[6] Android Developer Best Practices. Supporting Multiple Screens. Available from: http://developer.android.com/guide/practices/screens_support.html. [Accessed February 1, 2017].

[7] Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey: reliability and validity in a patient population. Med Care 1988;26(7):724–35.

[8] Hays RD, Sherbourne CD, Mazel RM. User's manual for the Medical Outcomes Study (MOS) core measures of health-related quality of life. Santa Monica, CA: Rand Health, 1995.

[9] 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].

[10] 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/epro/#section-5648. [Accessed February 1, 2017].

[11] McEntegart DJ. Equivalence testing: validation and supporting evidence when using modified pro instruments. In: Byrom B, Tiplady B, eds. ePRO: Electronic solutions for patient-reported data. London, UK: Gower, 2010:185–210

[12] Tiplady B. The use of a mobile phone to administer visual analogue scales. Conference presentation. Low Wood, Windermere, UK: Psychobiology Section of the British Psychological Society, 2005.