The First Validated Innovation in Scale Design Since the Migration to Screens: Results of an Equivalence Study, Testing a New Graphical Version of the Widely Used Verbal Rating Scale

Muehlhausen W[1], Doll HA[2], Hellman B[3], James B[3], Nagrani G[1], O'Neill B[1], Ward T[4]
1 SAFIRA Clinical Research Ltd, Ireland; 2 Clinical Outcomes Solutions, UK; 3 uMotif, UK; 4 Dublin City University, Ireland

PCR49
Presented at ISPOR 2023, 8-May, Boston, MA, USA

Abstract

Objective

While the use of Patient Reported Outcomes Measures (PROMs) in clinical research continues to grow and user interface design evolves steadily, newly developed PROMs use mostly standard scale types such as Numeric Rating Scale (NRS), Verbal Response Scale (VRS), Visual Analog Scale (VAS) in a paper-based original design (Figure 1). This research validated a modern design for the VRS and tested it against the standard VRS-design of the EQ-5D-5L to encourage instrument developers to utilise modern design approaches supported by new technology

Methods

A randomised equivalence study with 55 participants in two groups (Group A=28, Group B=27) and a cross-over design using the standard EQ-5D-5L VRS vs. the new VRS design was conducted. A short training of the two VRS versions was provided to the participants. A period with an active distraction task between the applications was observed and the Intraclass Correlation Coefficients (ICCs) for the EQ-5D-5L were calculated to show measurement equivalence between the two VRS versions. Participants of both groups also completed the identical EQ-5D-5L vertical scale.

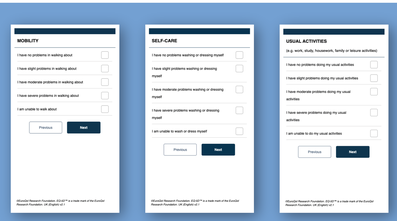
Results

Participants in both groups reported no difficulty with the new petal-like (Motif) VRS widget and all participants successfully completed both versions of the VRS. The ICC of 0.875 and the lower bound 95%-confidence interval of 0.796 are strong indications that these two representations of the VRS are equivalent.

Conclusions

Existing and newly developed PROMs utilise only a limited number of well established scale types. These were developed for use on paper and then, almost unchanged, migrated to electronic platforms. This project shows that it is possible to develop modern graphical user interfaces of existing standard scale types (i.e. VRS) without changing psychometric properties of the PROM. The Motif VRS can be used to replace the standard VRS design in existing PROMs and in the development of new PROMs.

Figure 1



Introduction

The ISPOR ePRO Taskforce in 2009 published recommendations on evidence needed to support measurement equivalence between electronic and paper PRO instruments[1]. Most PROMs use only three scale types or widgets: Visual Analog Scale (VAS), Verbal Response Scale (VRS), and the Numeric Rating Scale (NRS). Other widgets are more generic in nature and not research-specific. They include numbers, date and time pickers, and multiple-choice answer options.

Further research by B. Byrom, W. Muehlhausen et al. has shown that it is possible to establish equivalence across platforms with less arduous processes, when the eCOA Consortium Best Practices have been utilized during the migration process[2]. This includes the use of standard representations of the three widgets.

New PROMs, developed with only these standard widgets and which adhere to the recommendations by the eCOA Consortium's best practices can often be relieved of the recommended equivalence testing, cognitive debriefing and user acceptance testing[3]. Hence most new instruments only use the standard widgets.

uMotif has developed a new graphical user interface (Motif) for data capture, in the form of flower-like data input petals. By swiping in or out, each petal is used to enter data with a corresponding answer on the display (Figure 2).

Figure 2

Patient feedback about the innovative VRS has been very positive, but we needed confirmation that the new form factor was equivalent when tested against a standard VRS. When a new widget or a new representation of an existing widget is part of a PROM, the ISPOR Taskforce recommends a quantitative equivalence study to establish equivalence.

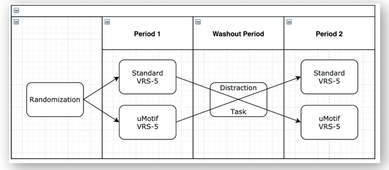
The Motif (Figure 2) differs substantially from a generic VRS and hence we conducted this study to assess if both representations are equivalent and thus that the Motif does not introduce any bias.

Methods

We recruited 55 participants with a chronic or acute pain condition and randomized them into two groups (Group A = 28, Group B = 27). We had a gender distribution of 35 female vs 20 male participants.

Participants used their own devices (BYOD) and we recorded 30 Android and 25 iOS phones.

We used a single cross-over design with Group A answering the EQ-5D-5L first with the standard widget and the Motif widget after the distraction task. Group B did the same in reverse order with the same distraction task between the sessions.



The distraction task was a Spatial Memory Test with increased complexity.
The responses were then transmitted to a central database and analyzed.

The equivalence of responses across the two representations was examined in a mixed model by computing an $ICC_{2,1}$ for Shrout-Fleiss reliability between the two EQ-5D-5L sum scores. Calculations were done overall and by study group. The VAS scale score was excluded from the calculation.

Results

An ISPOR task force has published a guideline recommending the minimum threshold of an ICC value of 0.7 that needs to be exceeded in order to consider any modified instrument versions to be equivalent.

As shown in Table 1, the ICC for equivalence overall between the EQ-5D-5L sum scores of the two VRS representations was 0.875 (95% CI 0.796 to 0.924).

Table 1: Equivalence of the EQ-5D-5L Sum Score[1] Between Standard[2] and Petal[3] completion

Population[4]	N	ICC (95% CI)
All Participants	55	0.875 (0.796 , 0.924)
Group A	28	0.904 (0.808 , 0.953)
Group B	27	0.854 (0.711 , 0.929)

CI = Confidence Interval; ICC = Intra-Class Correlation Coefficient.

[1] The EQ-5D-5L Sum Score is a simple sum of the 5 individual dimension scores where 20=health state equivalent to death and 0=full health.
[2] Standard completion of the EQ-5D-5L.
[3] Petal completion of the EQ-5D-5L.
[4] The population comprises all participants with EQ-5D-5L sum scores from both modalities: 28 in Group A and 27 in Group B.

The equivalence is strong, both overall and by study group, with both the ICC and the lower bound of the 95% CI being above the recommended threshold of 0.7.

Conclusion

This project was conducted to determine if the Motif widget is equivalent to the standard VRS widget and does not introduce bias.

The ICC results are very strong with the lower bound of the 95% CI being well above the standard threshold of 0.7, demonstrating excellent agreement.

It shows that, while most instruments are configured with standard widgets, it is possible to design new widgets and maintain equivalence across the representations of standard PROMs.

We encourage developers and authors to utilize modern user experience methodologies to develop new concepts to capture data from patients directly. We all expect today's apps and websites to have a modern look which supports our efforts to set up more patient-focused clinical trials.

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 in Health 2009; 12, 419-429.

(2) Muehlhausen W, Byrom B, Skerritt B et al. Standards for Instrument Migration When Implementing Paper Patient-Reported Outcome Instruments Electronically: Recommendations from a Qualitative Synthesis of Cognitive Interview and Usability Studies. Value in Health (2018 Jan;21(1):41-48).

(3) C-Path Institute ePRO Consortium (2014). Best Practices for Electronic Implementation of Patient-Reported Outcome Response Scale Options.

Corresponding Author

Willie Muehlhausen, DVM, SAFIRA Ltd, Ireland
Email: Willie.Muehlhausen@SAFIRA.Partners

Thanks to EuroQOL for use of the EQ-5D-5L

SAFIRA
CLINICAL OUTCOMES SOLUTIONS
umotif
Insight SFI RESEARCH CENTRE FOR DATA ANALYTICS