Check for updates

Expert Commentary

DIA

Therapeutic Innovation & Regulatory Science
1-5
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2168479018793369
tirs.sagepub.com

Measurement Equivalence of Patient-Reported Outcome Measures Migrated to Electronic Formats: A Review of Evidence and Recommendations for Clinical Trials and Bring Your Own Device

Bill Byrom, PhD[1], Chad Gwaltney, PhD[2], Ashley Slagle, PhD[3], Ari Gnanasakthy, MBA[4], and Willie Muehlhausen, DVM[5]

Abstract

A growing number of clinical trials employ electronic media, in particular smartphones and tablets, to collect patient-reported outcome data. This is driven by the ubiquity of the technology, and an increased awareness of associated improvements in data integrity, quality and timeliness. Despite this, there remains a lingering question relating to the measurement equivalence of an instrument when migrated from paper to a screen-based format. As a result, researchers often must provide evidence demonstrating the measurement equivalence of paper and electronic versions, such as that recommended by the ISPOR ePRO Good Research Practices Task Force. In the last decade, a considerable body of work has emerged that overwhelmingly supports the measurement equivalence of instruments using screen-based electronic formats. Our review of key works derives recommendations on evidence needed to support electronic implementation. We recommend application of best practice recommendations is sufficient to conclude measurement equivalence with paper PROMs. In addition, we recommend that previous usability evidence in a representative group is sufficient, as opposed to per-study testing. Further, we conclude that this also applies to studies using multiple screen-based devices, including bring-your-own-device, if a minimum device specification can be ensured and the instrument is composed of standard response scale types.

Keywords

patient-reported outcomes; ePRO, BYOD, measurement equivalence

Introduction

The collection of patient-reported outcome (PRO) data is an increasingly important component of today's clinical trials and patient care, and an increasing number of studies use electronic screen-based formats, such as smartphones and tablets, to collect this data in field-based and in-clinic settings. The ubiquity, low cost, and robustness of modern mobile devices has helped to drive this increased adoption, along with the need to demonstrate the integrity, quality, and timeliness of data collected. Further, due to the desire to make trial participation more convenient and patient-centric, there is increased interest in using the patients' own mobile devices (ie, BYOD [bring your own device]) to collect PRO data. This may also have potential additional benefits such as improved patient-reported outcome measure (PROM) completion compliance and reduced implementation costs.[1]

Because most PROMs were originally developed and validated in paper form, care is needed when migrating to electronic formats to ensure the instrument measurement properties are unaffected. The most common changes applied when migrating an instrument from paper to an electronic screen-based format are minor in nature. These include small formatting changes such as presenting only a single question per screen; wording changes such as changing question response instructions from "tick" or "circle" on pen and paper to "select" on an electronic implementation; and display orientation changes such as transposing response options from a horizontal listing on paper to a vertical listing on a mobile device screen.

[1] CRF Health, London, United Kingdom
[2] Gwaltney Consulting, Westerly, RI, USA
[3] Aspen Consulting, LLC, Philadelphia, PA, USA
[4] RTI Health Solutions, Research Triangle Park, NC, USA
[5] Muehlhausen Ltd., Tipperary, Ireland

Submitted 26-Apr-2018; accepted 12-Jul-2018

Corresponding Author:
Bill Byrom, PhD, CRF Health, 3rd Floor Brook House, 229-243 Shepherds Bush Road, Hammersmith, London, W6 7AN, United Kingdom.
Email: bill.byrom@crfhealth.com

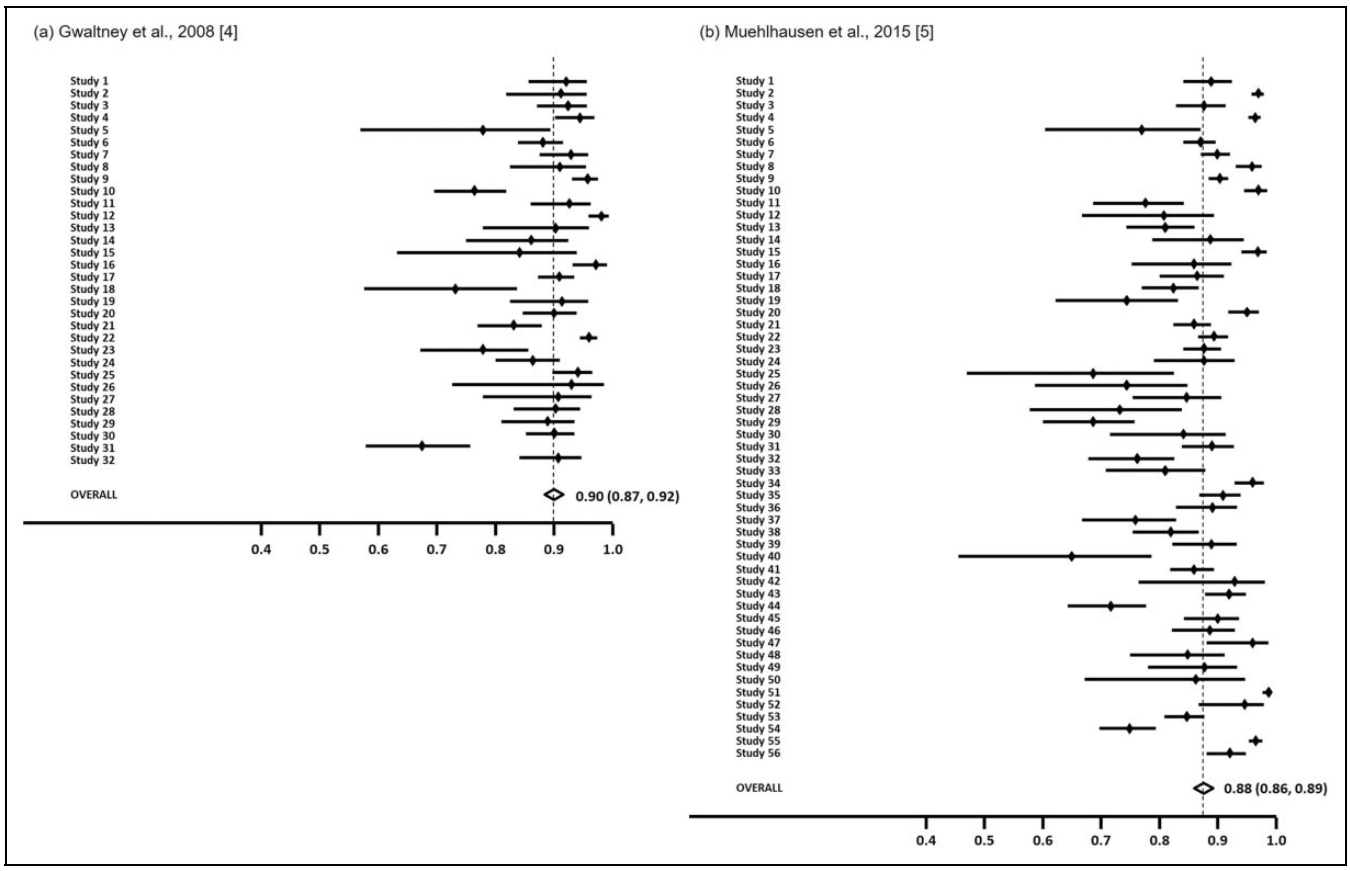

Figure 1. Results of correlation analyses from published meta-analyses of quantitative equivalence studies (correlation and 95% confidence intervals).

Outcomes researchers in the biopharmaceutical industry perform these simple modifications on a frequent basis, but despite this, unanswered questions about equivalence create uncertainty about the evidence needed to support appropriate migration. As a result, researchers often must provide evidence demonstrating the equivalence of the original paper version and the electronic version before administering the electronic version in a clinical trial. For example, the ISPOR ePRO Good Research Practices Task Force (2009) recommend this type of modification requires de novo evidence prior to administration in a trial—specifically the conduct of a cognitive interview and usability study in the target patient population.[2] These recommendations have been largely adopted by the industry.[3]

In the last decade, a large number of quantitative and qualitative equivalence studies have been conducted, and this provides the opportunity to derive evidence-based recommendations. This commentary summarises the main works in this area, and makes recommendations for work needed to support minor instrument changes due to migration between formats including the use of bring your own device (BYOD) for ePRO. We limit our consideration of electronic formats to screen-based modalities such as PC, tablet computer, and smartphone/handheld device and do not extend our findings to other modalities such as voice (interactive voice response [IVR] and voice assistants).

Research Synthesis of Evidence on Instrument Measurement Equivalence Meta-analyses

Two formal meta-analyses[4,5] and one review of quantitative equivalence studies[6] have been reported in the scientific literature. These quantitative studies considered the equivalence of paper and electronic formats, where ePRO solutions were implemented using common hardware such that each participant used an electronic device, PC, or other solution of the same make, model, and specification in addition to paper.

Gwaltney et al[4] performed a meta-analysis of 46 published equivalence studies reporting 278 instrument comparisons published up to 2006 comparing PROM measurement properties between paper and PC/handheld device implementations. This meta-analysis involved a variety of patient populations, with samples ranging from 10 to 189 subjects. Both the meta-analysis of mean differences and correlations showed high overall levels of agreement between paper and computerized measures (eg, pooled correlation coefficient [weighted summary correlation derived from the individual study correlations reported]: 0.90 [95% CI: 0.87-0.92]; Figure 1a). Estimating the effect of publication bias, this study reported that an additional 95 studies showing a correlation equal to the lowest observed amongst the studies (a correlation of 0.68) would be needed to

be included in the analysis to reduce the pooled correlation to below 0.75.

Campbell et al[6] reported a review of 55 quantitative equivalence studies of 79 PROMs, conducted between 2007 and 2014, comparing paper to PC/handheld device administration. This work explored the methodologies employed by the identified equivalence studies, but did not perform a formal meta-analysis. While the review reported 43 studies (78%) concluding equivalence between paper and ePRO, it identified 2 studies (4%) that failed to report equivalence and a further 10 studies (18%) where the authors' conclusions were not clear. One study reporting nonequivalence identified an intraclass correlation coefficient (ICC) of at least 0.95 as acceptance criterion, despite observing ICCs for all scores above 0.7,[7] which is the typical acceptance threshold for test-retest reliability assessment and often also adopted for migration equivalence assessment.[2] The second study concluding nonequivalence did not use conventional equivalence testing methodology.[8]

The meta-analysis reported by Muehlhausen et al[5] also considered studies conducted since the analysis of Gwaltney (2007-2013) and included 72 studies, 25 of which were included in the review of Campbell, including the two studies in which authors concluded nonequivalence.[7,8] This meta-analysis included 23 different patient populations, and 152 PROM comparisons. Electronic modalities included PC, handheld device, and IVR. Meta-analysis of mean differences and correlations concluded that electronic and paper PROMs, and different modes of electronic administration, produce equivalent scores across a wide range of patient populations and instruments (pooled correlation coefficient: 0.875 [95% CI 0.867-0.884]; Figure 1b). This meta-analysis also explored the potential impact of publication bias affecting the results, and indicated that in the analysis of correlations, a further 123 studies showing a correlation equal to the lowest observed among the studies (a correlation of 0.65) would be needed to reduce the pooled correlation coefficient to below 0.75.

Qualitative Meta-synthesis

More recently, a qualitative synthesis of cognitive interview and usability testing (CI/UT) studies—equivalence testing methodology recommended by the ISPOR task force[2]—has been reported.[3] This synthesis comprised 53 unpublished CI/UT studies involving a total of 101 PROM evaluations. The studies represented the full set of CI/UT evaluations conducted by a particular clinical research organization between 2012 and 2015, and so were robust to potential publication bias effects. Each study required participants to read the PROM on both paper and electronic formats before the conduct of a standardized semistructured interview by an experienced qualitative interviewer to probe whether perceived differences in the self-report task, or aspects of the changes between formats, might impact the potential to answer differently between formats.

In this synthesis, 6 studies (11%) reported significant findings that may affect measurement equivalence from cognitive interview or usability testing, but in all cases these could be eliminated by good product design (eg, the contrast and color used when presenting visual analogue scale lines) or by implementing ePRO design best practice recommendations, such as those reported by the Critical Path Institute's ePRO Consortium.[9,10] Design issues identified in these studies that could result in nonequivalence included the following: needing to scroll where multiple questions were provided on the same screen; difficulties using the mouse for PC-based formats; font display size and reduced paragraph spacing on the electronic format; insufficient thickness of visual analogue scale lines for clear representation on screen; difficulty in the use of up/down arrows to control numeric data entry; difficulty understanding how to enter free text on a mobile device diary; and ensuring appropriate instructional text and term definitions were visible.[3] The authors concluded that issues found would be eliminated when implementing minor changes during instrument migration by (1) applying ePRO design best practice and (2) ensuring product usability has been assessed in a representative patient group. They went on to propose and specify an expert screen review to assess ePRO implementation that could replace CI/UT in many instances.

BYOD Quantitative Equivalence Study

To date, one formal quantitative equivalence study investigating PROM measurement properties using patient's own devices (BYOD) has been published.[11] The completed study compared the equivalence of measures collected using a PROM composed of typical response scale types (visual analog, verbal response, and numeric response scales) between BYOD (the device brought to clinic by the subject), paper, and a provisioned device administration. Evidence for equivalence was defined as the lower bound of the 95% confidence interval for the ICC exceeding 0.7.

Provisioned devices were selected from a predefined set of 5 devices provided to the clinic, enabling additional equivalence assessment across a range of standard device sizes and the two main platforms (Android and iOS). The study was conducted in 155 subjects, ages 19-69 years, suffering from chronic health conditions resulting in daily pain or discomfort. All administrations were conducted in clinic, within a single visit. The devices brought by subjects were iOS or Android smartphones or tablets of various sizes (79 [51%] had 3- to 5-inch screens, 52 [34%] had 5- to 7-inch screens, and 24 [15%] had >7-inch screen sizes). Aside from requiring an iOS or Android device with an active app store account, there were no other limitations imposed on the devices that subjects could use for the BYOD administration. However, 16 subjects (10%) were unable to complete the PROM on their own device as a result of forgotten app store credentials (8/16), app incompatibility issues (4/16), inability to locate the downloaded app (Android only, 3/16), and insufficient

storage space (1/16). Aside from app incompatibility, these issues could be mitigated in conventional study designs where the screening period may provide sufficient time to retrieve access credentials, or make space on the device.

The randomized, crossover study uniquely assessed each PROM item separately, as opposed to looking only at the instrument total or subscores. The authors felt that this enabled greater generalizability of findings to other instruments containing different questions but using the same response scale types. Strong evidence of measurement equivalence between paper, BYOD, and the provisioned device was reported: all ICCs in the overall test exceeded 0.816, with the lower bound of the 95% confidence interval exceeding 0.77 in all cases.

A further study sponsored by the Critical Path Institute's PRO and ePRO Consortia is under way.[12] This will be a valuable addition to the body of evidence as it is designed to capture field-based equivalence in an uncontrolled setting, which goes beyond the simple assessment of measurement equivalence, as reported by the study above.

Recommendations

The body of work we have reviewed provides strong evidence to support the measurement equivalence of PROMs when migrated from paper to electronic formats, and between electronic formats. Average equivalence coefficients from quantitative studies are well above the threshold for test-retest reliability. Beyond this, there are very few individual analyses out of the hundreds that have been conducted, suggesting non-equivalence between paper PROs and electronic PROs. Although file-drawer effects (the effects that negative unpublished work may have on review findings[13]) are always a source of bias, the number of studies that would be required to reach a conclusion of non-equivalence is very large. Based on this, we conclude that the risk of introducing measurement error or bias when using electronic platforms to administer PROMs is small. We propose alternative recommendations on the evidence needed to demonstrate measurement equivalence when migrating PROMs from paper to screen-based electronic formats in clinical research studies where a provisioned handheld device of the same make and model will be provided to all participants, and when using BYOD.

Instrument Migration to a Standard Device

Based on the 2 meta-analyses,[4,5] and the synthesis of CI/UT studies,[3] there is sufficient strength of evidence to conclude that if ePRO design best practice[9,10] is used, and the platform has previously demonstrated usability (eg, assessing the ease of app access and usage, use of navigational controls and touchscreen, and operation without assistance) in a representative group, there is no need to perform additional quantitative or qualitative equivalence studies where response scale types employ commonly used types (visual analog, verbal, and numeric response scales), and standard objects (eg, date pickers). In these cases, a structured expert screen review, as detailed by Muehlhausen et al,[3] is sufficient. To effectively represent most patient populations, representative groups should include subjects with a range of characteristics that may impact their ability to use and operate an ePRO device, such as device-literacy, manual dexterity, eyesight, and cognitive capabilities.

Bring Your Own Device

With the additional evidence of the recent BYOD equivalence study,[11] we propose that there is no need to perform measurement equivalence studies across different device sizes when planning for BYOD ePRO if a minimum device specification can be defined and identified, and the instrument is composed of standard response scale types (visual analog, verbal, and numeric response scales) and standard objects (eg, date pickers). Ensuring a minimum device specification is a sensible precaution while the impact of scrolling on PROM measurement properties is unstudied, and screen review should be conducted on a device in common with the minimum device specification permitted for the proposed study. However, we acknowledge that it is becoming less likely that trial participants will present with devices of insufficient size due to the current trend towards manufacture of larger smartphone devices. When reported, the ongoing field-based BYOD equivalence assessment study[12] will provide additional important understanding of the equivalence of BYOD in an uncontrolled field-based setting.

Conclusions

In the last decade, a compelling accumulation of evidence supporting measurement equivalence following PROM migration has been published. In the light of this, we believe there is now enough data to make evidence-based recommendations on the evidence needed to support measurement equivalence following minor changes due to instrument migration, in the ways described above. This also applies to the evidence needed to support measurement equivalence when using BYOD with a defined minimum device specification.

BYOD promises to provide greater convenience for trial participants, enabling subjects to record PROM data on the device they refer to regularly and are familiar with. This may lead to increased PROM compliance and reductions in missing data. BYOD may also simplify trial logistics if device provisioning is not required, and may lower the associated costs of collection of these data.

While the evidence we have reviewed supports the use of BYOD from the viewpoint of ensuring that PROM measurement properties are maintained when used across devices of varying platforms and sizes, it is important to recognize that BYOD is associated with additional technical and practical considerations that need to be addressed by solution providers and those designing and managing clinical trials. These include, for example, the ability for subjects to turn off in-app notifications, remove the study app, change their

device or plan, and run out of data credit or device storage; interruptions due to other activities on the device—such as receiving a phone call; or additional security and privacy concerns when collecting data through a less controlled platform. These are not the focus of this commentary, but should be addressed through appropriate technical and practical working practices.

Declaration of Conflicting Interests

Dr Gwaltney reports personal fees from ERT, Inc, outside the submitted work.

Funding

No financial support of the research, authorship, and/or publication of this article was declared.

ORCID iD


Bill Byrom, PhD http://orcid.org/0000-0002-9186-4271

References

1. Gwaltney C, Coons SJ, O'Donohoe P, et al. "Bring your own device" (BYOD): the future of field-based patient-reported outcome data collection in clinical trials? *Therapeutic Innovation & Regulatory Science* 2015;49:783-791.
2. 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;12:419-429.
3. 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 Health* 2018;21:41-48.
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: 322-333.
5. 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;13:167-187.
6. Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. *Qual Life Res* 2015; 24: 1949-1961.
7. Juniper EF, Langlands JM, Juniper BA. Patients may respond differently to paper and electronic versions of the same questionnaires. *Respir Med* 2009;103(6):932-934.
8. Swartz RJ, Moor CD, Cook KF, et al. Mode effects in the Center for Epidemiological Studies Depression (CES-D) scale: personal digital assistant versus paper and pencil administration. *Qual Life Res* 2007;16(5):803-813.
9. Critical Path Institute ePRO Consortium. Best practices for electronic implementation of patient-reported outcome response scale options. https://c-path.org//wp-content/uploads/2014/05/BestPracticesForElectronicImplementationOfPROResponseScaleOptions.pdf. Accessed December 8, 2017.
10. Critical Path Institute ePRO Consortium. Best practices for migrating existing patient-reported outcome instruments to a new data collection mode. https://c-path.org//wp-content/uploads/2014/05/BestPracticesForMigratingExistingPROInstrumentstoaNewDataCollectionMode.pdf. Accessed December 8, 2017.
11. Byrom B, Doll H, Muehlhausen W, et al. 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. *Value Health* 2018;21(5):581-589.
12. Critical Path Institute Patient-Reported Outcome Consortium. Request for proposal. https://c-path.org/wp-content/uploads/2016/01/20160115_rfp-byod_study_and_ext_study_rev_eprosc.pdf. Accessed December 8, 2017.
13. Rosenthal R. The "file drawer problem" and tolerance for null results. *Psychol Bull* 1979;86:638-641.