Ensuring Equivalence of Electronic and Paper Administration of Patient-Reported Outcome Measures

Moderated by: Bill Byrom (ICON)
Presented by:

Paul O’Donohoe (CRF Health)  Willie Muehlhausen (ICON)  Bill Byrom (ICON)
Biographies

Bill Byrom, PhD, is Senior Director of Product Innovation at ICON, UK. He has worked in the Pharmaceutical industry for over 25 years in a variety of roles, specializing in eClinical technology. Bill has authored over 60 publications including an industry textbook on electronic Patient Reported Outcomes. Bill is the incoming Vice Director of the ePRO consortium.

Willie Muehlhausen, D.V.M., is ICON’s Head of Innovation, based in Ireland. He has been in the Clinical Research Industry since 1996 and held various roles in CROs and Technology Providers. Willie served as the inaugural Vice-Director of the ePRO consortium from 2011-2013 and was a member of PharmaVoice100 in 2015.

Paul O’Donohoe is Director of Health Outcomes at CRF Health. He is responsible for developing the company’s internal health outcomes expertise and supporting clients across the range of scientific issues that can arise during the course of a clinical trial. He is passionate about developing the field of eCOA through research and active involvement in industry consortia.
Please use Q&A feature to submit questions to presenter

- If in full screen mode, select following:

- If not in full-screen mode, the Q&A box is open to your right.

- When asking questions, be sure to select “All Panelists”
The Electronic Patient-Reported Outcome (ePRO) Consortium was established by the Critical Path Institute (C-Path) in 2010. Along with C-Path, the members of the ePRO Consortium are firms that provide electronic data collection technologies and services for capturing patient-reported outcome (PRO) and other clinical outcome assessment (COA) data in clinical trials.

The mission of the ePRO Consortium is to advance the science of clinical trial endpoint assessment by collaboratively supporting and conducting research, designing and delivering educational opportunities, and developing and disseminating best practice recommendations for electronic collection of clinical outcome data.
ePRO Consortium members
Agenda

1. Brief overview of scientific and regulatory considerations when migrating a paper PRO measure to ePRO

2. Studies to ensure equivalence between alternate modes of PRO administration

3. Equivalence considerations when using mixed modes of administration; existing evidence
Electronic Patient-Reported Outcomes (ePRO)

Data collected directly, electronically from the patient

Variety of Input Methods
- Dedicated handheld devices (provisioned)
- Tablets
- Smart phones
- Interactive Voice Response System (IVRS)
- Internet/web-based
Advantages of ePRO

Electronic methods help both patient and researchers
- Only valid, in-range entries can be made
- Time stamping and time windows
- Automatic validation and navigation
- Reminders and feedback enhance compliance
- Missing data can be reduced or eliminated
- Data available for prompt review
- Easy to use and generally preferred to paper
Differences Between Electronic and Paper

There are typically *some* differences in wording, even if appearances are closely matched

- Selecting/tapping rather than ticking or circling a choice
- Navigation

These differences could have an impact

- Instrument migrated from paper to any electronic format, and validation data from paper version is used to support electronic version
- Need to use data from paper and electronic modes interchangeably
A “faithful migration” refers to the development of alternative modes of data collection that do not introduce response bias that results from changes in the way the instrument is presented/formatted or how the subject interacts with it.

- Retain look of question as much as possible
- Retain exact wording where possible
- Keep question and responses together
- Evaluate need for instructions on same screen or different screens
- Do not delete instructions because they ‘seem’ intuitive
Presented at DIA 2016 52nd Annual Meeting
June 28, 2016, Philadelphia, PA
When do we need to show equivalence?

Guidance for Industry¹
Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

December 2009
Clinical/Medical

“The adequacy of an instrument’s development and testing is specific to its intended application in terms of population, condition, and other aspects of the measurement context for which the instrument was developed. **When a PRO instrument is modified, sponsors generally should provide evidence to confirm the new instrument’s adequacy.** That is not to say that every small change in application or format necessitates extensive studies to document the final version’s measurement properties. **Additional qualitative work may be adequate depending on the type of modification made.**” (p. 20)
When do we need to show equivalence?

“Examples of changes that can alter the way that patients respond to the same set of questions include:

- Changing an instrument from paper to electronic format“ (p. 20)
When do we need to show equivalence?

- Increasingly, instruments are developed from the start in electronic form, and all validation data are obtained using the electronic modality.
- In this case, paper to electronic equivalence is not needed.
- However, now we may need to show equivalence between the original electronic version and the paper version.
When do we need to show equivalence?


Stephen Joel Coons, PhD,1 Chad J. Gwaltney, PhD,2 Ron D. Hays, PhD,3 J. Jason Lundy, PhD,4 Jeff A. Sloan, PhD,5 Dennis A. Revicki, PhD,6 William R. Lenderking, PhD,7 David Cella, PhD,8 Ethan Basch, MD, MSc,9 on behalf of the ISPOR ePRO Task Force
# Modification Matrix

**Table 1** PRO to ePRO measurement equivalence: Instrument modification and supporting evidence

<table>
<thead>
<tr>
<th>Level of modification</th>
<th>Rationale</th>
<th>Examples</th>
<th>Level of evidence</th>
</tr>
</thead>
</table>
| Minor                 | The modification can be justified on the basis of logic and/or existing literature. No change in content or meaning. | 1) Nonsubstantive changes in instructions (e.g., from circling the response to touching the response on a screen).  
2) Minor changes in format (e.g., one item per screen rather than multiple items on a page). | Cognitive debriefing  
Usability testing |
| Moderate              | Based on the current empirical literature, the modification cannot be justified as minor. May change content or meaning. | 1) Changes in item wording or more significant changes in presentation that might alter interpretability.  
2) Change in mode of administration involving different cognitive processes (e.g., paper [visual] to IVR [aural]),. | Equivalence testing  
Usability testing |
| Substantial           | There is no existing empirical support for the equivalence of the modification and the modification clearly changes content or meaning | 1) Substantial changes in item response options  
2) Substantial changes in item wording | Full psychometric testing  
Usability testing |

Adapted from Shields et al. [62].
Examples of Minor Changes

Minor changes in instruction
- Paper: “Check” or “Circle”
- Electronic: “Tap” or “Select”

Minor changes in layout
- Single item per screen vs 10 items per page
Examples of Moderate Changes

Changes in display
- Scrolling/Toggling through answers
- Separate question from answer

Changes in modality (cognitive processes)
- IVRS to paper
Equivalence Decision Tree

Types of Equivalence Evidence

Existing literature
- Meta-analysis of electronic/paper comparisons
- Studies directed at impact of specific changes

Qualitative studies for minor modifications:
- Cognitive interviews
- Investigate how patients comprehend/use instruments

Quantitative studies for moderate modifications:
- Formal comparisons using methods similar to the test-retest reliability
- Crossover within-patients design
- Evaluate mode differences and correlations
Equivalence Study Objectives

Demonstrate that patients comprehend questions the same way regardless of mode of administration

- Demonstrate this comprehension by hearing from patients and/or demonstrating equivalence in responses
- Demonstrate that target population can use the electronic platform
- Ensure that the migration does not introduce changes to the measurement properties
- Reliability, validity, ability to detect change
Case Studies

St. George’s Respiratory Questionnaire (SGRQ)
- Cognitive interviewing

AQLQ
- Equivalence testing
Case Studies

St. George’s Respiratory Questionnaire (SGRQ)
- Cognitive interviewing
## St. George’s Respiratory Questionnaire

**PART 1**

*Please describe how often your respiratory problems have affected you over the past 4 weeks.*

<table>
<thead>
<tr>
<th></th>
<th>almost every day</th>
<th>several days a week</th>
<th>a few days a month</th>
<th>only with respiratory infections</th>
<th>not at all</th>
</tr>
</thead>
</table>

1. Over the past 4 weeks, I have coughed: □ □ □ □ □ □
2. Over the past 4 weeks, I have brought up phlegm (sputum): □ □ □ □ □ □
3. Over the past 4 weeks, I have had shortness of breath: □ □ □ □ □ □
4. Over the past 4 weeks, I have had wheezing attacks: □ □ □ □ □ □
ST. GEORGE’S RESPIRATORY QUESTIONNAIRE (SGRQ)

This questionnaire is designed to help us learn much more about how your breathing is troubling you and how it affects your life. We are using it to find out which aspects of your illness cause you most problems, rather than what the doctors and nurses think your problems are.

Please read the instructions carefully and ask if you do not understand anything. Do not spend too long deciding about your answers.

Before completing the rest of the questionnaire:

Please tap one box to show how you describe your current health:

Very good    Good    Fair    Poor    Very poor
Demographics

COPD Sample (n=10)

- Mean age 62 years, SD 8 (range 45-74)
- 70% women
- 90% Caucasian, 10% Black

Length of time since diagnosis (months)

- Longest: 743
- Shortest: 34
- Average: 160
- SD: 211.2
Data Analysis and Reporting

Side-by-side comparison of electronic and paper versions

Example interview questions:

Q. How would you compare the instructions of the paper diary and the electronic diary?
Q. How would you compare the overall appearance of the paper diary and the electronic diary?
Q. How would you compare the text size of the paper diary and the electronic diary?
Q. How would you compare moving from question to question of the paper diary and the electronic diary?
Q. Do you feel that any of your answers were different due to the layout of the diaries, from paper version to electronic version? If yes, please can you explain why

With few exceptions, patients reported that all items were interpreted and comprehended in the same way on paper and electronic versions. There was some feedback related to respondents experience with technology, particularly older participants.
Results - Usability

- All participants (100%) could see all of the text and images.
- All participants (100%) found navigation easy.
- Nine participants (90%) were comfortable with using the touch screen.
- One participant (10%) noted that the touch screen sometimes did not respond.
- All participants (100%) felt that they would have been able to complete the instruments on the device on their own with no help from the interviewer.
- Many positive remarks were expressed for the tablet device including ‘easy’, ‘more convenient’, ‘good size’.
- There were few negative remarks: having to press ‘next’ on the SGRQ; ‘a bit heavy’; and one participant who said that they ‘don’t like technology’.
Case Studies

AQLQ

- Equivalence testing
Asthma Quality of Life Questionnaire with Standardised Activities (AQLQ(S))

SELF-ADMINISTERED
(≥12 years)

The Asthma Quality of Life Questionnaire with Standardised Activities (AQLQ(S)) is copyrighted and all rights are reserved. No part of this questionnaire may be sold, modified or reproduced in any form without the express permission of Elizabeth Juniper on behalf of QOL Technologies Limited.

APRIL 2008

Copyright © 2008 QOL Technologies Limited

Please complete **all** questions by selecting the number that best describes how you have been during the **last 2 weeks as a result of your asthma**.

1. How limited have you been during the last two weeks as a result of your asthma?

**STRENUOUS ACTIVITIES** (such as hurrying, exercising, running up stairs, sports)?

1. Totally Limited
2. Extremely Limited
3. Very Limited
4. Moderate Limitation
5. Some Limitation
6. A Little Limitation
7. Not at all Limited
Asthma Sample

- 60 participants were enrolled in the study. Of these, 9 did not achieve AQLQ(S) or ACQ scores within the required range in the first administration and did not complete the main study. These subjects were not included in the analysis.
- Male n=27, Female, n=24
- Range 18-65
- Mean age - 36.5 years, SD=14.5
- White (n=44), Black (n=3), Other (n=4)
- Hispanic or Latino (n=38), Not Hispanic or Latino (n=13)
Results

In summary, the overall ICC scores (0.91) demonstrate that the electronic version of the Standard Asthma Quality of Life Questionnaire can be considered equivalent to the original paper versions.
Considerations for Mixed Modes

FDA PRO Guidance

“We intend to review the comparability of data obtained when using multiple data collection methods or administration modes within a single clinical trial to determine whether the treatment effect varies by methods or modes.” (FDA, 2009)
Considerations for Mixed Modes

Technology makes mixed modes of data collection feasible operationally, however…

- Clinical trial designs need to consider sources of error variance in the PRO data.
- Only sufficiently tested PRO collection modes should be considered.
- Measurement error reduces statistical power and attenuates the ability of the trial to detect real change (i.e., treatment effect) in the PRO-based trial endpoint.

Considerations for Mixed Modes

Recommendation:

“However, we also strongly discourage the mixing of paper and electronic field-based instruments and suggest that mixing of only electronic modes be considered for clinical trials and only after equivalence has been established.”

For more information:

Area is evolving – as more evidence becomes available, there will be less of a need for additional evidence

Chad J. Gwaltney, PhD,¹,⁵ Alan L. Shields, PhD,²,⁵ Saul Shiffman, PhD¹,⁵
¹Brown University, Providence, RI, USA; ²East Tennessee State University, Johnson City, TN, USA; ³University of Pittsburgh, Pittsburgh, PA, USA; ⁴Invidata, Inc., Pittsburgh, PA, USA; ⁵PRO Consulting, Pittsburgh, PA, USA

ABSTRACT

Objectives: Patient-reported outcomes (PROs; self-report assessments) are increasingly important in evaluating medical care and treatment efficacy. Electronic administration of PROs via computer is becoming widespread. This article reviews the literature addressing whether computer-administered tests are equivalent to their paper-and-pencil forms.

Methods: Meta-analysis was used to synthesize 65 studies that directly assessed the equivalence of computer versus paper versions of PROs used in clinical trials. A total of 46 unique studies, evaluating 278 scales, provided sufficient detail to allow quantitative analysis.

Results: Among 233 direct comparisons, the average mean difference between modes averaged 0.2% of the scale range (e.g., 0.02 points on a 10-point scale), and 93% were within ±5% of the scale range. Among 207 correlation coefficients between paper and computer instruments (typically intraclass correlation coefficients), the average weighted correlation was 0.90; 94% of correlations were at least 0.75. Because the cross-mode correlation (paper vs. computer) is also a test-retest correlation, with potential variation because of retest, we compared it to the within-mode (paper vs. paper) test-retest correlation. In four comparisons that evaluated both, the average cross-mode paper-to-computer correlation was almost identical to the within-mode correlation for readministration of a paper measure (0.90 vs. 0.91).

Conclusions: Extensive evidence indicates that paper- and computer-administered PROs are equivalent.

Keywords: computer, electronic, equivalence, meta-analysis, paper and pencil, patient-reported outcomes.
Abstract

Objective: To conduct a systematic review and meta-analysis of the equivalence between electronic and paper administration of patient reported outcome measures (PROMs) in studies conducted subsequent to those included in Gwaltney et al's 2008 review.

Methods: A total of 1,997 records were extracted, in which differences (standardized mean difference) and mean differences were identified. Correlations varied by study design, with a pooled coefficient of 0.28 (95% CI 0.08 to 0.90) and were still highly variable (I² = 92.1). Similarly, average platform-specific ICCs from 39 studies (42 estimates) had a pooled estimate of 0.90 (95% CI 0.88 to 0.92) with an I² of 91.5. After excluding 20 studies with outlying correlation coefficients (≥3SD from the mean), the I² was 54.4, with the equivalence still high, the overall pooled correlation coefficient being 0.88 (95% CI 0.87 to 0.88). Agreement was found to be greater in more recent studies (p < 0.001), in randomized studies compared with non-randomised studies (p < 0.001), in studies with a shorter interval (<1 day) (p < 0.001), and in respondents of mean age 28 to 55 compared with those either younger or older (p < 0.001). In terms of mode/platform, paper vs Interactive Voice Response System (IVRS) comparisons had the lowest pooled agreement and paper vs tablet/touch screen the highest (p < 0.001).

(Continued on next page)
**Abstract**

**Purpose** Technological advances in recent decades have led to the availability of new modes to administer patient-reported outcomes (PROs). To aid selecting optimal modes of administration (MOA), we undertook a systematic review to determine whether (and direction) exist.

**Methods** We searched 2004 (date of last complete search) to April 2014, cross-referenced the records of all studies that compared paper versus electronic PRO measurement. In all, 5100 papers were screened, 222 were potentially relevant and 56 met eligibility criteria. No evidence of bias was found for: (1) paper versus electronic self-complete; and (2) self-complete versus assisted MOA. Heterogeneity for paper versus electronic comparison was explained by type of construct (i.e. physical vs. psychological). Heterogeneity for self-completion versus assisted MOA was in part explained by setting (clinic vs. home).

**Results** Of 5100 papers screened, 222 were considered potentially relevant and 56 met eligibility criteria. No evidence of bias was found for: (1) paper versus electronic self-completion at the domain level (effect size = 0.01; Table 3). Figure 2 presents the forest plot of the paper versus self-completion comparison. We found evidence of assisted completion was by self-complete in clinic and electronic MOA was by self-complete at home. Results support the use of mixed MOAs within a research study, which may be a useful strategy for reducing missing PRO data.

**Keywords** Systematic review · Patient-reported outcome · Mode of administration · Bias
Abstract

Objectives: To test the impact of the method of administration (MOA) on score level, reliability, and validity of scales developed in the Patient Reported Outcomes Measurement Information System (PROMIS).

Study Design and Setting: Two nonoverlapping parallel forms each containing eight items from each of three PROMIS item banks (Physical Function, Fatigue, and Depression) were completed by 923 adults with chronic obstructive pulmonary disease, depression, or rheumatoid arthritis. In a randomized crossover design, subjects answered one form by interactive voice response (IVR) technology, paper questionnaire (PQ), personal digital assistant (PDA), or personal computer (PC) and a second form by PC, in the same administration. Method equivalence was evaluated through analyses of difference scores, intraclass correlations (ICCs), and convergent/discriminant validity.

Results: In difference score analyses, no significant mode differences were found and all confidence intervals were within the prespecified minimal important difference of 0.2 standard deviation. Parallel-forms reliabilities were very high (ICC = 0.85–0.93). Only one across-mode ICC was significantly lower than the same-mode ICC. Tests of validity showed no differential effect by MOA. Participants preferred screen interface over PQ and IVR.

Conclusion: We found no statistically or clinically significant differences in score levels or psychometric properties of IVR, PQ, or PDA administration compared with PC. © 2014 Elsevier Inc. All rights reserved.

Keywords: Patient-reported outcomes; Quality of life; Questionnaire; Mode of administration; Method of administration; Item response theory
We know this is equivalent
...so why not this?
Conclusions

- Migrating from paper to electronic platform is considered a modification of the original instrument
- Case by case evaluation of the extent of the modification, regulatory strategy, and the nature of evidence needed to establish the equivalence of the electronic measure
- Typically requires small-sample cognitive interview study
- Large amount of existing equivalence evidence – how much more needed?

http://c-path.org/programs/ePRO
Questions?

http://c-path.org/programs/ePRO
Thank you for attending the ePRO Consortium Webinar

Ensuring Equivalence of Electronic and Paper Administration of Patient-Reported Outcome Measures