Testing the measurement equivalence of electronically migrated patient-reported outcome (PRO) instruments: Challenges and Solutions

Sonya Eremenco, Evidera, Inc., Bethesda, MD, USA
Diane Wild, ICON, Oxford, UK
Karl McEvoy, CRF Health, Hammersmith, UK
Jason Lundy, Critical Path Institute, Tucson, AZ, USA

Overview

- Introduction
- Qualitative Assessment of Conceptual Equivalence
  - Challenges
  - Study design considerations
  - Study design options
- Quantitative Assessment of Statistical Equivalence
  - Challenges
  - Study design considerations
  - Study design options
- Diaries and Setting of Assessment
- Q&A

ePRO Consortium Background

- The ePRO Consortium was established on April 1, 2011
- Mission: To advance the quality, practicality, and acceptability of electronic data capture (EDC) methods used in clinical trials for PRO endpoint assessment
- Provides a non-competitive, neutral environment for ePRO providers to collaborate on issues related to capturing patient data electronically

ePRO CONSORTIUM

CRITICAL PATH INSTITUTE

ALMAC

Bracket™

CRF HEALTH

REAL PATIENT DATA 24/7

ERT®

Getting It Done. Right.

exco intouch

Healthcare at your fingertips

ICON

A Symbol of Excellence

medidata®

PHT

Introduction

- Assessing equivalence of migrated instruments is necessary
- Level of equivalence evidence is dependent on the extent that the changes or modifications are likely to have affected the subjects' interpretation and responses to the items in the instrument.
- Lack of consensus around study design considerations for both qualitative and quantitative methods
- ePRO Consortium establishing good practices

Levels of Equivalence Evaluation: ISPOR ePRO Task Force Report

Table I PRO to ePRO measurement equivalence: instrument modification and supporting evidence

Level of modification	Rationale	Examples	Level of evidence
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].

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(4):419-429.

http://www.ispor.org/TaskForces/ePROTF.asp

Introduction: Qualitative

- Purpose of cognitive interviews

Type of interview	Exploration	Confirmation
Concept elicitation interview	x	
Cognitive interview for content validity	x	x
Cognitive interview for migration equivalence		x

- In migration studies, cognitive interviews are not intended to revisit the content validity of the original instrument.

Qualitative Assessment of Conceptual Equivalence

- Challenges
- Study design considerations
- Study design options

Qualitative - Challenges

- Goal: confirm that the interpretation and response to the items has not changed due to the migration
- How to assess this using qualitative methods rather than statistical methods?
- Incorporate both cognitive interview and usability testing in the same interview session

Qualitative – Study Design Considerations

- Sample size
  - Smaller numbers than content validity because saturation is not as difficult to achieve
  - Range: 5 to 20
    - Is 5 sufficient to have confidence in the results?
    - 10 seems optimal to detect issues
    - 20 may be necessary if subgroups are involved

Qualitative – Study Design Considerations

- Conducting multiple rounds of interviews
  - PRO: if sufficient number of participants, allows for identification of issues, revisions and then further testing in a later round
    - Allows for multiple sites, geographic and other diversity
  - CON: if changes are made based on a small sample to begin with, later groups may contradict those findings
    - Multiple rounds more time consuming
  - Option: conduct 1 round, but pause after first 5 interviews to look for issues, make changes if necessary and then continue with the interviews
- Sufficient sample size needed for multiple rounds

Qualitative – Study Design Considerations

- Whether to have patients answer both versions of the instrument or only the electronic version
- Is the comparison between formats helpful or necessary?
- How to assess response if only one version is answered?

Study Design Option 1

- Subjects complete instrument on both modes
- Determine if responses for any items differ between modes
- Interview focuses on those items individually to determine if different responses were random (could go either way) or systematic due to differences in meaning or interpretation between modes
  - This approach is not quantitative, responses are not analyzed but discussed with the subject qualitatively to identify reason for difference
- If latter occurs with a substantial number, the migration of those items should be revisited
- If changes are made, additional cognitive interviews are recommended
- If further discrepancies are found, quantitative equivalence study should be considered

Study Design Option 2

- Subjects complete instrument on both modes
- Ask subjects if they think their responses for any items differed between modes
- Interview focuses on those items individually to determine whether the subject feels that they responded differently because of the mode of administration

Study Design Option 3

- Subjects complete new mode only
- Ask how they interpret what each item is asking them
  - Repeat item in their own words (paraphrasing)
  - Think-aloud
- Subject's interpretation is compared to an item definition document to see if there is concordance
  - Assumes this document is available from instrument developer or easy to create
- More closely parallels cognitive interview during instrument development
- Responses are not examined

Study Design Option 4

- Ask only about instructions and/or items that were modified during migration
- Enables a more focused investigation of potential impact of the changes
- Subjects asked to read both versions on the two modes and identify any perceived differences in the task or in the interpretation/ meaning of modified items.

Qualitative Study Design

- Combination of approaches is possible
  - Option 1 with asking about all items
  - Option 1 focusing only on items that have changed

- Tradeoffs
  - Time to complete both versions in Options 1 and 2
  - Lack of assessment of response in Options 3 and 4
  - Difficulty identifying reasons for differences in Option 1

Quantitative Assessment of Statistical Equivalence

- Challenges
- Study design considerations
- Study design options

Quantitative - Challenges

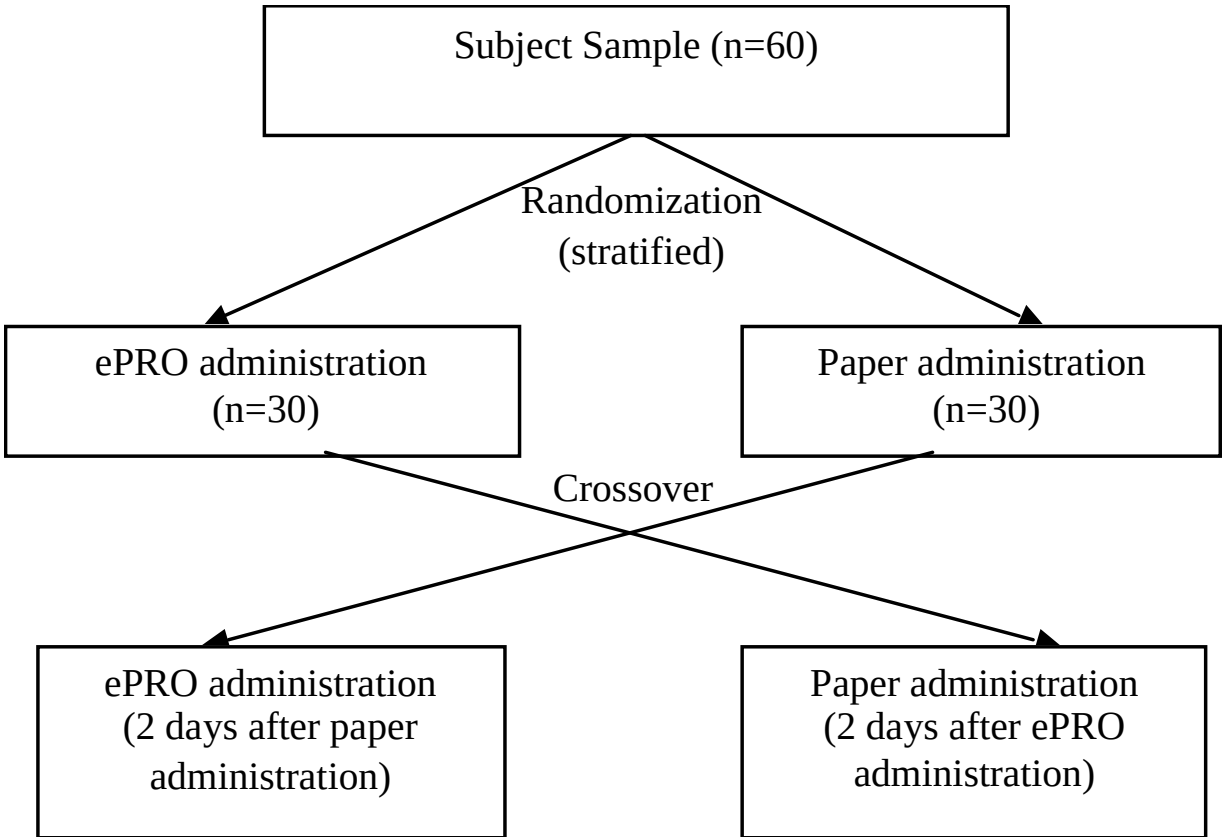
- Goal: to ensure that migration to the electronic mode did not introduce systematic bias to the scores obtained from the instrument
- Study Design Challenges
  - Sample size and statistical methods
  - Time between administrations (for crossover designs)
  - Setting and population for study completion

Quantitative – Study design considerations

- Sample size and statistical methods
  - Two-period, crossover design using ICC and test of mean differences
  - n=50-60 suitable in most cases

  - Alternative approaches use item response theory
  - Differential Item Functioning (DIF)
    - Requires large samples; n=300-500

Crossover Design



Study Interval

- In the equivalence literature, retest intervals range from minutes to weeks apart
- Shorter intervals
  - Benefit: allow the subject to complete both instrument administrations within a single study visit
  - Risk: Memory effect could bias results
- Longer intervals might be better suited for modes that don't require physical hardware deployment

Statistical Analysis

- ICC (3,1) as defined in Shrout & Fleiss (1979)
  - Intraclass Correlations: Uses in Assessing Rater Reliability, Psychological Bulletin, 1979; 86 (2): 420-428.
  - Shrout & Fleiss define six types of ICCs, others have been defined elsewhere
  - Selecting the wrong ICC will impact results
- ICC threshold = to within-subjects ICC, or 0.70

Quantitative – Study design considerations

- Time between administrations
  - Memory effect from repeated administrations is a threat to the validity of the study
- In equivalence literature, retest intervals range from 1 minute to several months
  - Interval needs to be long enough to wash-out any memory effect, but short enough to ensure the subject's condition has not changed

Quantitative – Study design considerations

- Strategies for time between administrations

1. Same day completion
   - Subjects complete two administrations within same visit – minutes to hours apart
     - Efficient, minimize loss to follow-up
     - Use of distraction task can minimize memory effect

Quantitative – Study design considerations

- Strategies for time between administrations

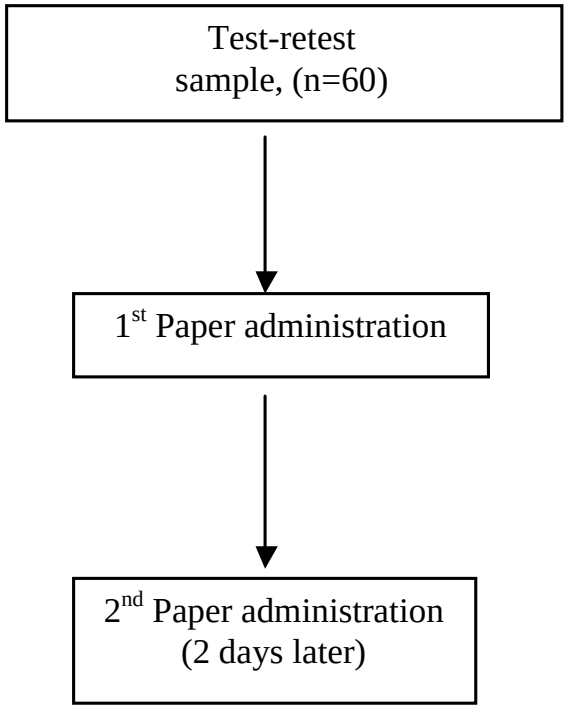
2. Multi-day/take home completion
   - Two administrations completed at least 24-hours apart
     - May reduce memory effect, requires more planning and follow-up

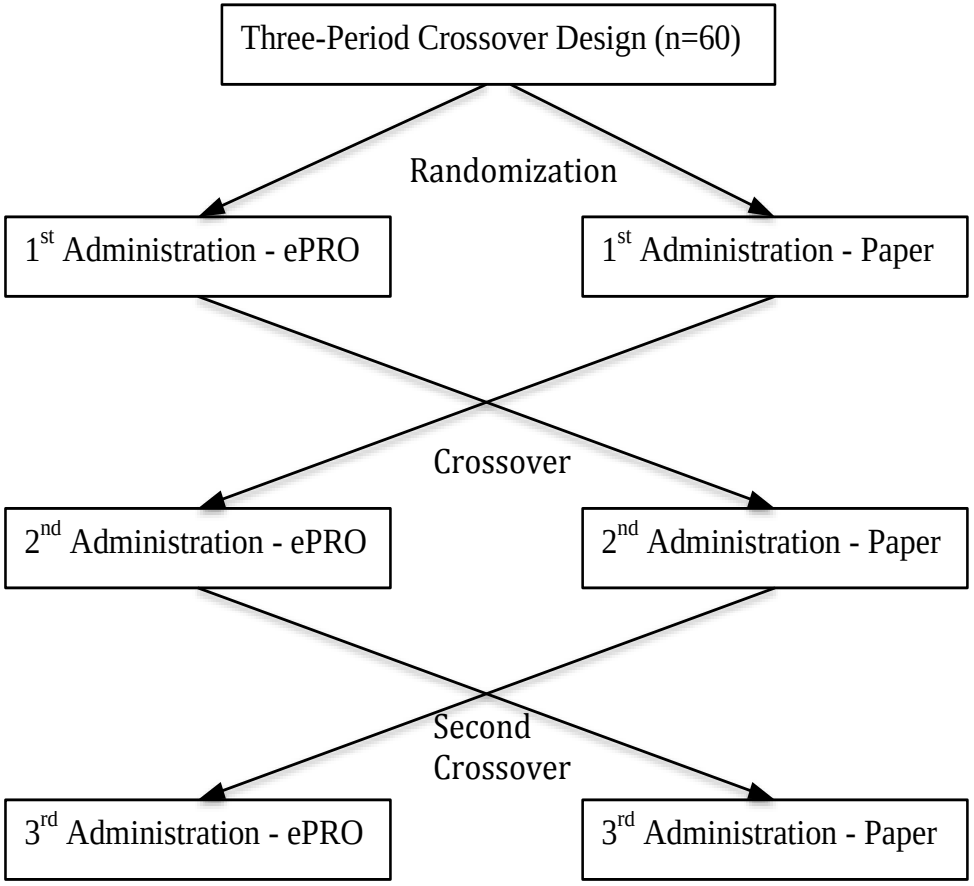
Experimental Study Designs

- Limitations of paper test-retest data
  - Older data from different sample
  - Don't have ICCs to compare nor mean differences from paired sample T-test

- Recommended to generate new source mode retest data
  - Separate single retest arm, extend crossover to a 3 period design

Source Mode Retest Data





Diaries and Setting of Assessment

- Diaries are instruments that are designed to be completed outside the clinic

- For both qualitative and quantitative studies, key question is whether it is necessary to take the diary home, outside of the artificial clinic setting, to collect data as part of the equivalence assessment process

Qualitative and Quantitative

- Diaries are often designed to be completed early in the morning or before bed, or episodically when events occur
- Interview setting is artificial, not done at the time the diary would normally be completed
- Therefore responses are more hypothetical in nature or based on recall, or thinking back to the previous event

Challenge of setting

- Patients completing questionnaires (paper or electronic) in a clinic setting are more likely to focus on the task at hand and one can ensure they complete all required questionnaires
  - This is not possible outside of the clinic setting

- Paper and electronic may be completed very differently outside of the clinic setting resulting in an apparent lack of equivalence
  - Training/technical issues?
  - Expect low ICC

Challenge of setting

- How important is it to replicate the circumstance of use that patients will face in the clinical trial?

- Home doesn't just mean in the comfort of one's own home..
  - Activities of day to day living that cannot be reproduced in a clinical setting.

Clinic Setting

- White coat syndrome?
  - Subjects nervous in the clinical setting?
  - More relaxed in their “home” environment?
- No “daily” data
  - Do you ask patients about a hypothetical event/symptom?
- Possible false sense of the equivalence of paper and electronic given the differences between home and clinic environment

Home Setting

- Having patients take the diary home may extend the timelines of the equivalence study
- Additional burden to patient if they need to provide multiple days worth of data?
  - But this is closer to how it would be used in a trial
- Can we combat “Parking lot syndrome” with regard to paper as we can in the clinic?

Q&A

Thank you!

- Contact details for further questions:
  - Sonya Eremenco: sonya.eremenco@evidera.com
  - Diane Wild: diane.wild@oxfordoutcomes.com
  - Jason Lundy: Jlundy@c-path.org
  - Karl McEvoy: karl.mcevoy@crfhealth.com