



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

Stephen Joel Coons, PhD, Chad J. Gwaltney, PhD, Ron D. Hays, PhD, J. Jason Lundy, MS, Jeff A. Sloan, PhD, Dennis A. Revicki, PhD, William R. Lenderking, PhD, David Cella, PhD, Ethan Basch, MD, MSc on behalf of the ISPOR ePRO Task Force

Published *Value in Health*, 2009

1



Table 1 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 sheet).	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 visibly changes content or meaning.	1) Substantial changes in item response options 2) Substantial changes in item wording	Full psychometric testing Usability testing

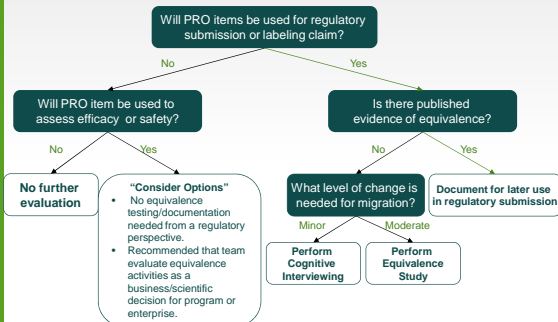
Adapted from Shih et al. [32]

Coons et al., 2009

2



Instrument Implementation: Paper to Electronic Equivalence Decision Tree



Note: The appropriate license must be procured, regardless of equivalence activities

Presented by Jean Paty, ISPOR Forum, May 24, 2011

3



How is migration different from mixing?

- Migration results in new mode that substitutes for the original mode in future studies
 - Reliability is the main concern
 - Electronic to paper may be the migration direction in the future.
- Mixing involves using both old and new modes within a study and pooling the data
 - Reliability is the main concern
 - Increased error possible
 - Levels: Country, Site, between patient, within patient



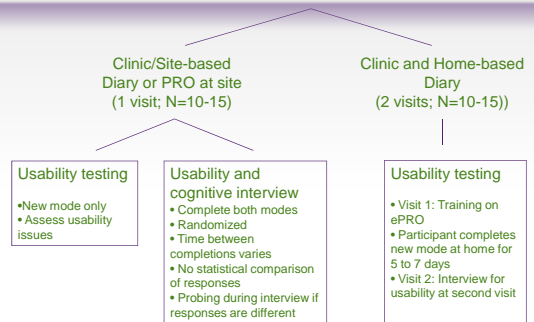
Prerequisites for Mixing Modes

- ⊙ Stakes are higher when mixing
- ⊙ Equivalence between modes should be evaluated before decision to mix
- ⊙ What equivalence study designs can be used?
 - ⊙ Qualitative: Usability/Cognitive Interview
 - ⊙ Quantitative: Equivalence Study

5



Common Qualitative Study Designs



6



Qualitative Study Comparisons

Study design type	Pros	Cons	Recommended for	Limitations
Clinic/site-based – new mode only	Focus on new mode, identify usability issues	No comparison with previous mode; artificial setting for diary	New instruments where cognitive interview used single page format; PRO at site	Insufficient for mixed modes situation; artificial setting for diary – can't test alarms and compliance features
Clinic/site-based – multiple modes	Usability and format comparison assessed	No statistical comparison due to small sample size	Instruments where format not previously used in cognitive interview; PRO at site	Artificial setting; provides evidence of format equivalence not statistical; insufficient for mixed modes using paper diary
Clinic and home-based	Usability over time, alarms and compliance features assessed	Difficult to compare with previous mode	Diary studies, especially frequent or episodic assessments per day	Insufficient for mixed modes using paper diary; no evidence of statistical equivalence



Equivalence Study Designs

- When is an equivalence study recommended for mixed modes?
 - ⊙ Moderate modifications between the modes
 - ⊙ For paper/web vs. IVR studies
 - ⊙ For paper vs. electronic diary studies
- Note: for a new PRO developed for electronic use, if the electronic version is used in the validation study and only the electronic version will be used in the future, then an equivalence study is not necessary

8

Common Equivalence Study Designs

**Clinic/Site-based
Diary or PRO at site
(1 visit; N=60)**

- Randomized to order
- Complete both modes within same session
- Distraction task in between
- Time between completions varies
 - Few minutes – 2 hours
- Results are compared statistically

• Usability interview may be added at the end

**Clinic and Home-based
Diary
(3 visits; N=60)**

- Randomized to mode
- First mode completed between visit 1 and 2
- Second mode completed between visit 2 and 3
- Time between visits varies
 - 1 week – 2 weeks
- Results are compared statistically

• Usability interview may be added at Visit 3

Coons et al. (2009) also mentions randomized parallel groups design as an option.

Experimental Equivalence Study Designs

**Clinic/Site-based
Diary or PRO at site
(1 visit; N=60)**

- "Double cross"
- Randomized to order
- Complete first and second modes within same session
- Complete first mode again
- Distraction task in between
- Time between completions varies
 - Few minutes – 2 hours
- Results are compared statistically between and within mode

• Usability interview may be added at the end

**Clinic and Home-based
Diary
(4 visits; N=60)**

- "Double cross"
- Randomized to mode
- First mode completed between visit 1 and 2
- Cross to second mode and complete between visit 2 and 3
- Cross back to first mode and complete between visit 3 and 4
- Time between visits varies
 - 1 week – 2 weeks
- Results are compared statistically between and within mode

• Usability interview may be added at Visit 4

Common Equivalence Study Comparisons

Study design type	Pros	Cons	Recommended for	Limitations
Clinic-based – randomized cross-over	Statistical equivalence level between modes can be established	Artificial setting for diaries	PRO instruments completed at site	Comparison with original mode test-retest reliability may be limited; doesn't reflect true performance of paper diary
Clinic and home-based – randomized cross-over	Statistical equivalence level between modes can be established; real world setting for diary	Studies takes longer to complete	Diary studies, especially frequent or episodic assessments per day	Comparison with original mode test-retest reliability may be limited,

Experimental Equivalence Study Comparisons

Study design type	Pros	Cons	Recommended for	Limitations
Clinic-based – double cross	Statistical equivalence level between and within modes with same patient sample	Artificial setting for diaries	PRO instruments completed at site	Stable patient sample necessary for test-retest different from changing population in clinical trial; doesn't reflect true performance of paper diary
Clinic and home-based – double cross	Statistical equivalence level between and within modes with same patient sample; real world setting for diary	Most complex study design – more time and resources needed	Diary studies, especially frequent or episodic assessments per day	Stable patient sample necessary for test-retest different from changing population in clinical trial



Conclusions

- Qualitative study designs
 - Acceptable for migration equivalence
 - Do not show statistical equivalence for mixed modes and insufficient for mixed paper and electronic diary use
- Equivalence study designs
 - Diary may be tested in clinic-based design, one time administration, inconsistent with actual trial use and doesn't reflect true performance
 - Critical that patient population is stable, unchanging to limit true change in response in equivalence studies
 - Clinical trial use assumes that patient will change over time due to treatment, may be impossible to distinguish what is driving change in scores
- May conclude that the potential differences between paper and electronic diaries are too great to allow mixing in a clinical trial, and default should be to use the electronic diary only.

13