ISPOR FORUM

ISPOR PRO Mixed Modes Task Force

(www.ispor.org/sigs/mixedmodes.asp)

ISPOR FORUM

Mixing Modes of Patient-Reported Outcomes Data Collection in Clinical Trials: Recommendations

Moderator: **Sonya Eremenco, MA**
United BioSource Corporation, Bethesda, MD

Presenters

Stephen Joel Coons, PhD – Critical Path Institute, Tucson, AZ
Jean Paty, PhD – ERT (formerly invivodata, inc.), Pittsburgh, PA

ISPOR Patient Reported Outcomes Task Forces
PRO Mixed Modes Task Force

- Working Group formed in March 2010
- Task Force approved by ISPOR Board March 2011
- Leadership team: 20 members
- Acknowledgments:

Special thanks to Antonia Bennett, PhD, Ethan Basch, MD, MSc, Damian McEntegart, Kathy Wyrwich, PhD, and Karin Coyne, PhD for their contributions as members of the ISPOR PRO Mixed Modes Task Force

Task Force Objective

- Develop a Good Research Practices report to address the use of more than one mode of data collection for a specific instrument in the clinical trial setting
  - Provide recommendations to ensure the quality and comparability of the resulting PRO data
  - Review analytical approaches for evaluating and pooling mixed modes data

4

Current Status of Report

- Outline
  - Background
  - Modes and Mode Selection
  - Migration
  - Equivalence
  - Mixing Modes Considerations
  - Recommendations
- Review process
  - Distributed for first round of review 10 May 2013
  - Full distribution to the PRO SIG Review Group June 2013

5

Data Collection Modes and Migration between Modes

Stephen Joel Coons, PhD

Background

- ISPOR ePRO Task Force Report (Coons et al. 2009)
  - Migrating from paper to electronic data collection
  - Mixing modes not explicitly addressed
- 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)
- In this presentation, "mode" refers to all means of administration and methods of data collection
- Mixing modes is most challenging when one of the modes is paper

Issues to Consider

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

- Clinical trial designs should avoid as many sources of error variance in the PRO data as possible.
- Measurement error can be introduced into the trial design by different PRO data collection modes that are not providing comparable data (i.e., the modes lack sufficient measurement equivalence).
- 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.

Mixing Modes: Avoidable Variability in Clinical Trials

- Mixing modes may be **an avoidable source of measurement error** in multinational clinical trials
- Unavoidable sources of measurement error:
  - Translation and cultural adaptation
  - Cultural biases due to differing experiences of the condition
  - Variability in patient's ability to reflect and provide a response
- Recommendation
  - Data collection modes *should not be varied within a single clinical trial or between trials that seek to pool or compare the data* without prior evidence of sufficient measurement equivalence between the modes
  - Avoid this source of additional measurement error in trials

Mixing Modes: Need To Be Pragmatic

- Mixing modes does occur in clinical trials and has to be addressed pragmatically
- Some evidence in literature of measurement equivalence across modes, but more needed
  - Literature not definitive, limited by publication bias (i.e., positive findings more likely to be published than those with inconclusive or negative results
- Report objective – Provide recommendations on how to avoid potential measurement error when mixing is inevitable
  - "Faithful" migration
  - In-depth exploration of assessing measurement equivalence
  - Operational and statistical considerations for the clinical trial setting

Modes of Data Collection (1 of 2)

- **Paper-and-Pencil**
- **Digital Pen**
  - Specially printed paper questionnaire
- **Handheld Devices**
  - Small, handheld computers/smartphones with touchscreens
- **Tablet Computers**
  - Larger mobile devices that have integrated touchscreens
- **Desktop or Laptop Computers**
  - Larger screen size than tablets; not intended to be mobile;
  - Do not have touchscreens, so keyboard or mouse needed to enter responses
- **Interactive Voice Response Systems**
  - Pre-recorded voice question and response option script

Modes of Data Collection (2 of 2)

Stand-Alone vs. Web-based Systems

- Stand-Alone: self-contained, all software and functionality located on the device
- Web-based: device is vehicle to access system through Web browser
- Substantial difference in control of the presentation of PRO instrument between these systems
  - Standalone: consistent presentation
  - Web-based: variability in devices and interfaces resulting in variable presentation
    - Considered mixed mode due to potential variability
- "Apps" downloaded to users own phone another possibility with Web-based systems

Importance of Mode Selection

- Selection of PRO-based trial endpoint and mode of data collection should *not* be an afterthought
- Selection and evaluation of mode of data collection for clinical trial should occur *early* in clinical trial planning process
- Allows for sufficient time to build a validated system in the optimal mode for the trial that is (21 CFR) Part 11 compliant
  - Minimum of 3 months

Factors to Consider in Selection of Mode(s)

- Patient population
  - clinical trial subjects' sensory and physical abilities
- Location of data collection
  - clinic (investigative site) or field (e.g., home, workplace)
- Characteristics of instrument
  - length of questionnaire or length/format of responses
- Data collection schedule
  - monthly, weekly, daily, multiple times per day
- Feasibility of implementation
  - infrastructure of regions in which clinical trial conducted
- Cost
  - not a scientific consideration, but important nonetheless

Migration to New Data Collection Mode

A "faithful migration" is the development of alternative modes of data collection that do not bias response

- A faithful migration of an instrument does not need to look exactly like the original version, but it needs to capture the same data
- The migration process must ensure that…
  - only necessary changes to the format and instructions are made and that the content of the items and responses has not changed.
  - subjects ***interpret and respond*** to the questions/items the same way regardless of mode

15

Steps to Conduct a "Faithful Migration"

- Contact copyright holder for permission to migrate and determine if there are migration requirements
- Review original mode to identify necessary changes
- Contact copyright holder for approval of wording and formatting changes
- Conduct migration process and generate draft screens or IVR script
- Send draft screens/IVR script to instrument developer for approval
- Finalize version on electronic mode

Migration Recommendations

- Retain exact wording of item where possible
- Retain the order of response options
- Keep question and response options together on the same screen
- Evaluate the need for instructions on the same screen or different screens due to space constraints
- Although space may be available for multiple items, a single item per screen can provide consistency across screen-based migrations of the instrument
- Consider aesthetic elements such as spacing between item stems and responses and equal spacing of response options to reduce bias

17

Usability Considerations during Migration

- Screen-based devices
  - Ensure font size and resolution allow readable text for all ages
  - Balance space constraints of longer translations
  - Easy and intuitive navigation, clear buttons to move forward and back
  - Error messages for items subject to validation rules

18

Mode-Specific Considerations for Migration

- Smartphone/Handheld device
  - Space constraints imposed by smaller screens
- Tablet device
  - A single item per screen can provide consistency across multiple screen-based migrations of the instrument.
- Web-based format
  - Must consider the type of device used to access the Web interface.
- IVR
  - Migrating to a verbal script can lead to significant changes to the wording of item stems and response options

19

Usability vs. Feasibility Testing

- Usability

"Usability testing examines whether respondents…are able to use the software and the device appropriately. This process includes formal documentation of respondents' ability to navigate the electronic platform, follow instructions, and answer questions…as intended." Coons et al. (2009, p 423)

- Focused on respondent's ability to use the system
- May be conducted at investigative site, controlled environment with observation of the subject

20

Usability vs. Feasibility Testing

- Feasibility testing
  - Evaluation of the system (PRO instrument and data collection mode) within a specific study design
  - Need driven by novelty of the study design in which the PRO data collection system is to be implemented
    - Event-driven field-based data collected multiple times per day for a given population would benefit from feasibility testing
  - Testing plan for feasibility testing
    - Recruit subjects similar to trial population
    - Subjects follow the study procedures for a period of time (e.g., answer diary at home for 7 days)
    - Perform debriefing interviews to assess compliance with study procedures and assess usability

21

Evaluating Measurement Equivalence

Sonya Eremenco, MA

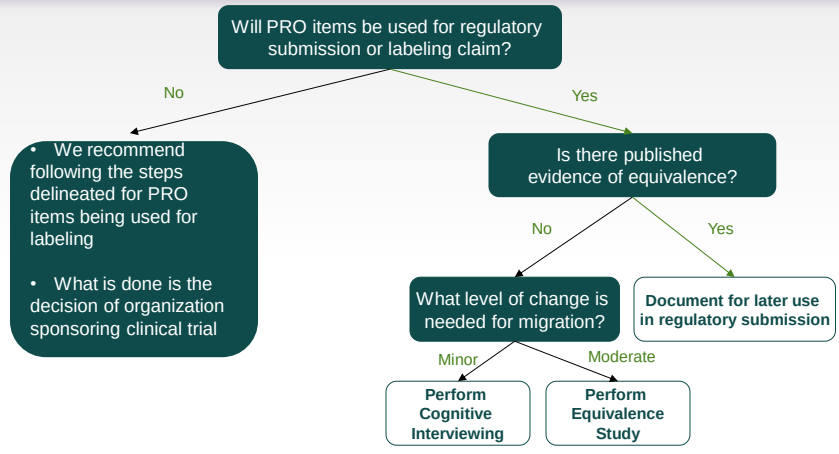
Goal of Evaluating Measurement Equivalence

- Any migration involves some type of modification
- After migration, determine whether the goal of faithful migration was achieved by evaluating measurement equivalence
  - Measurement equivalence vs. “validation”
- Ensure that respondents ***interpret and respond*** to the items the same way between original and migrated modes

Evaluating Measurement Equivalence

- Additional considerations for determining the level of evidence needed to establish measurement equivalence
- Delineate types of measurement equivalence testing and associated study designs

Need to Establish Measurement Equivalence



25

Measurement Equivalence Evaluation

- Level of equivalence evidence is dependent on the extent that the changes or modifications are likely to have had an effect on the subjects' interpretation and responses to the items in the instrument.

26

Measurement Equivalence Evaluation: Coons 2009 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 et al., 2009

27

Types of Changes

- Types of changes due to migration
  - Format: differences in how items/responses are presented
    - Adapting instructions: changing “circle” to “select”
  - Procedural: differences in how modes are implemented in studies
    - Edit or validation checks
    - Ability to skip questions if not relevant
    - Completion windows
    - Compliance with protocol requirements

28

Minor Modifications

Level of Modification	Rationale	Examples	Level of Evidence
Minor	The changes to instrument are *not likely* to have changed interpretation or responses.	Format: 1) Non-substantive 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). Procedural: 1) Implementation of tablet at the site with differences in edit checks, validation rules, branching logic.	Cognitive Interviewing Usability testing

Moderate Format Modifications

Level of Modification	Rationale	Examples	Level of Evidence
Moderate	The changes to instrument *may* have changed interpretation or responses	Format: 1) Changes in item wording or more significant changes in presentation that might alter interpretability. (e.g., splitting an item over two screens, changing the structure of the response options.) 2) Change in mode of administration involving different cognitive processes (e.g., paper [visual] to IVR [aural]). 3) Change in mode of data collection to web-based administration (e.g., variance between screen sizes too great to be considered minor modification.	Equivalence Study Usability testing

Moderate Procedural Modifications

Level of Modification	Rationale	Examples	Level of Evidence
Moderate	The changes to instrument *may* have changed interpretation or responses	Procedural: 1) Migration of paper diary to electronic platform with differences in edit checks, validation rules, branching logic, completion windows, compliance with administration recall period. 2) Differences in the ways that subjects are alerted to complete instruments (e.g., alerts on a handheld device always available vs. email reminders for web that require logging into email are not as proximal to the actual reminder time, and compliance could differ).	Equivalence Study Usability testing

Equivalence Study Designs

- Qualitative: Cognitive Interview
  - Provides qualitative data to evaluate equivalence
  - Associated with minor modifications
- Quantitative: Measurement Equivalence Study
  - Evaluate statistical equivalence of responses
  - Associated with moderate modifications

32

Qualitative Study Design: Cognitive Interview

- Purpose: to evaluate if the migration has impacted how subjects *interpret and respond* to the items
  - Not intended to revisit content validity of the original instrument
- Minor modifications to format
- Small sample size: 10 to 15 subjects
- Assess usability of instrument as a secondary goal

Approaches to Cognitive Interviews

- Subjects complete instrument on both modes, items with different responses probed if random or systematic due to mode
  - Revise and retest; if discrepancies persist consider quantitative measurement equivalence study
- Subjects complete new mode only and asked for interpretation of items; compare with item definition to show concordance
- Ask subjects only about instructions or items modified during migration: show both versions ask for perceived differences in interpretation/meaning
- No consensus on optimal approach
- Combination of above approaches is also possible

Measurement Equivalence Studies

- When is a measurement equivalence study recommended?
  - Moderate modifications between the modes
  - Mixing modes, especially in the following scenarios
    - For visual vs. auditory modes
    - Web at subject's homes
    - For paper vs. electronic diary studies
- Greater risks for differences in response between modes and greater need to demonstrate that they provide sufficiently equivalent results

35

Measurement Equivalence Study Designs

- Single visit
  - Evaluate if the migration changed interpretation at a point in time
  - Sufficient when moving away from paper data collection, whether site or field-based assessment
- Multi-visit field evaluation
  - Evaluate if migration changed interpretation and completion in the context of implementation of study design
  - Most useful for evaluating field-based assessments which are intended to be completed on a daily basis over a period of time and scores are typically averaged
  - Needed if intending to mix modes in the future

36

Common Measurement Equivalence Study Designs

Single Visit
Site- or Field-based Assessment
(1 visit; N=60)

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

Multi-Visit Field Evaluation
Field-based Assessment
(3 visits; N=60)

- Randomized to order
- Visit 1: Provide 1st mode, training if ePRO
- 1st mode completed between visit 1 and 2
- Visit 2: Provide 2nd mode
- 2nd mode completed between visit 2 and 3
- Time between visits varies
  - 1 week – 2 weeks
- Results are compared statistically

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

Common Equivalence Study Comparisons

Instrument Type	Study Design Type	Pros	Cons	Limitations
PRO instruments completed at site; Field-based assessments where mixing is not intended	Single Visit – randomized cross-over	Statistical equivalence level between modes can be established	Assesses format differences but not procedural differences	Comparison with original mode test-retest reliability may be limited; doesn't reflect true performance of paper diary in clinical trial setting
Field-based assessments, especially frequent or episodic assessments per day, where mixing is intended although not recommended	Multi-visit field evaluation randomized cross-over	Statistical equivalence level between modes can be established; real world setting for field-based	Studies difficult to operationalize because target concepts are variable, need to control for change; high likelihood that equivalence won't be found	Comparison with original mode test-retest reliability may be limited;

Conclusions to Evaluating Measurement Equivalence

- Qualitative study designs
  - Acceptable for minor migration equivalence
  - Do not show statistical equivalence for mixed modes and insufficient for mixed paper and electronic field-based assessment use
- Equivalence study designs
  - If field-based assessment tested in clinic-based design using one time administration, inconsistent with actual trial use and doesn't reflect true performance
  - Critical that subject population is stable, unchanging, to limit true change in response in equivalence studies
  - Clinical trial use assumes that subject 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 data collection mode only.

39

Considerations Prior to Mixing

- When planning equivalence testing process, determine purpose of migration
  - Move permanently to new mode
    - New modes replace old in future studies
    - Demonstrate equivalence for prospective use
    - Qualitative study may be sufficient
  - Migration with intention to mix
    - Both used in future studies and pooling of data for analysis
    - Demonstrate equivalence for concurrent use
  - When mixing within a trial, essential to demonstrate equivalence between modes in a quantitative equivalence test

40

Mixing Modes

Jean Paty, PhD

Prerequisites for Mixing Modes

- Report focuses on mixing modes within a trial in which PRO endpoints are intended to support label claims
- Mixing refers to administration of same instrument via different data collection modes in a trial
- Nonequivalence between modes could be the difference between success and failure for that endpoint in the trial
  - Treatment effects attenuated by mode differences
- Equivalence between modes should be evaluated before decision to mix

42

How Mixing Occurs

- Between product development programs
- Between clinical trials within a program
- Within a single clinical trial, such as
  - countries within a trial
  - sites within a country
  - subjects within a site
  - within a subject
  - timepoints within a trial (e.g., start with one mode and change to another mode)

43

Risk Assessment by Level

Level of Mixing	Risk to Equivalence
Between product development programs	Varies
Clinical trials within a program	Varies
Countries within a clinical trial	High
Sites within a trial	High
Subjects within a site	Very high
Within a subject	Extremely high

44

Mixing Modes: Planning vs. Defaulting

- Plan ahead to the extent possible
- Allow time for measurement equivalence evaluation to provide support for mixing if quantitative evidence is not available
- Defaulting to mixed modes because of problems with new mode in trial
  - Risk because measurement equivalence has not been evaluated

45

Mixing Paper and Electronic

- Most risky combination
  - Subjects not restricted on paper as in electronic modes
- Avoid as much as possible
- Mixing site-based instruments
  - Less risk, if equivalence previously demonstrated
- Mixing paper and electronic field-based assessments
  - Significant potential equivalence issues
  - Significant procedural change
  - “Don’t do it” (FDA guidance discourages paper diaries)

Mixing Electronic Modes

- Mixing visual only methods
  - Less risky
  - Potentially easier to demonstrate equivalence and implement consistently across modes
  - Web: proceed with caution
- Mixing visual and auditory modes
  - Potentially more challenging to demonstrate measurement equivalence because of moderate differences between modes
  - Quantitative equivalence needs to be demonstrated
  - Implementation challenges

Operational and Statistical Considerations for Mixing Modes

- Pre-trial Preparation
  - Evaluate measurement equivalence using study design appropriate for future mixed modes implementation
  - Assess risks of certain types of mixing
  - Power study according to results of equivalence evaluation:
    - Adjust the presumed measurement error in sample size calculation
  - Appropriate training for both modes needed
  - Criteria for which countries, regions, sites or subjects are permitted to mix needs to be documented and conveyed to investigative sites

48

Operational and Statistical Considerations for Mixing Modes

- Trial Implementation
  - Minimize site issues such as training or infrastructure that lead to defaulting to paper
  - If mixing is pre-planned
    - Manage where and when each mode is used
    - Fewer challenges mixing across countries, regions or sites, than within site or patient
  - Avoid ad hoc mixing by having contingency in case of technology failure
    - Consider options other than paper as a backup in diary studies
  - Develop SAP to address analysis of mixed modes *a priori* to evaluate if treatment effect differs by mode

49

Mixing and Statistical Considerations

- Post-Trial
  - Compare results by mode using techniques similar to testing translations for poolability
  - Assess mode as a variable for analysis, similar to site comparisons
  - Consider conducting sensitivity analysis to evaluate impact on data and treatment effect of including or excluding alternate mode data
    - Especially in case of ad hoc mixing where small number of subjects or sites use non-standard mode
  - Work with biostatistician to determine appropriate statistical techniques

50

Steps for Appropriate Migration and Mixing

1. Select appropriate mode(s) for trial
2. Perform a “faithful migration” (“migrate before you mix”)
3. Evaluate equivalence between the modes migrated and/or to be mixed
   - Use appropriate study design
4. If above conditions are met, implement the mode or modes in the trial
   - Avoid mixing paper and electronic diaries; assess risks of other combinations
   - If deciding to mix other modes
     - Plan and implement carefully; mix at country level or higher
     - Assess statistical issues and poolability of data

Questions?

- Thank you!