FORUM

ISPOR TASK FORCES

ISPOR Patient Reported Outcomes Task Forces

PRO Mixed Modes Task Force

ePRO Systems Validation Task Force

FORUM

Considerations for Using Mixed Modes for Patient-reported Outcomes Data Collection in Clinical Trials

and

Validation of Computerized Systems to Capture Outcomes Data

Moderator

Sonya Eremenco MA

Chair, ISPOR PRO Mixed Modes Task Force

ePRO Manager, United BioSource Corporation, Bethesda, MD, USA

PRO: Mixed Modes Task Force Members

Ethan Basch, MD, MSc, Memorial Sloan-Kettering Cancer Center

Antonia V. Bennett, PhD, Memorial Sloan-Kettering Cancer Center

Stephen Joel Coons, PhD, PRO Consortium, Critical Path Institute

Karin Coyne, PhD, United BioSource Corporation

Cindy Gao, PhD, PharMerit

Zhanna Jumadilova, MD, MBA, Pfizer, Inc.

Damian McEntegart, BSc, Perceptive Informatics

Willie Muehlhausen, DVM, ePRO Solutions, ERT

PRO: Mixed Modes Task Force Members

Jean Paty, PhD, Scientific, Quality & Regulatory Affairs invivodata, inc.

Jeff Sloan, PhD, Division of Biostatistics, Mayo Clinic

Tara Symonds, PhD, PRO Center of Excellence, Primary Care, Pfizer Ltd.

Simrandeep K. Tiwana – Sidhu, MBA, PhD, Health Partners Research Foundation

Hwee Lin Wee, PhD, Faculty of Science, National University of Singapore

Kathleen Wyrwich, PhD, United BioSource Corporation

Vladimir Zah, PhD(c), CEO, ZRX Outcomes Research, Inc.

Task Force Objective

Develop a Good Research Practices report to address the use of more than one mode of data collection or administration 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

Background

- ISPOR ePRO GRP Task Force Report (2009)
  - Migrating from paper to electronic data capture
  - 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)

Terminology: Methods, modes, modalities?

- FDA Guidance definitions:
  - Methods of data collection (paper, electronic)
  - Modes of administration (self vs. interviewer)
- PRO measurement field
  - "Mixed methods" refers to mixing qualitative and quantitative data collection
  - Modes often refers to electronic as well as self vs. interviewer in literature
- Task Force Report covers *ALL* modes: administration and data capture

Modes and Sources of Variability

Mode of administration	Method of data capture	Sources of variability between methods	Sources of variability between modes
1. Self-administered *Direct patient report considered PRO*	Paper Handheld Tablet /Netbook IVRS Web via computer Web via phone	Variation due to: -items being seen or heard; -how they appear on page or screen; -number of items visible on page or screen at one time, -how responses are presented, and -how patients are to input answer	Patient may alter response due to presence of interviewer (e.g. social desirability); and variation across interviewers (e.g. age, gender, personality)
2. Interviewer administered Considered PRO if items read verbatim and patient answer recorded without interpretation	In person – paper In person – tablet Over the phone – paper Computer-Assisted Telephone Interview (CATI)	Variation due to direct or indirect presence of interviewer; and variation across interviewers (e.g. age, gender, personality)	

Key Concerns

- Technology makes mixed modes data collection feasible operationally
- If we do mix:
  - Will data integrity and reliability be affected?
  - Will there be better compliance (more data)?
- If we don't mix:
  - Will there be more missing data?
  - Will data that are collected be better quality?
- Is losing data from patients a bigger disservice to the trial than collecting possibly compromised data using different modes?

Task Force Report

- Issues with mixed modes of data capture
  - Potential risks
    - Potential measurement error may impact treatment effect
    - Randomization process doesn't solve the problem
  - Potential benefits
    - Reduced missing data
    - Increased compliance with PRO data capture
- Strategies for appropriate use of mixed modes
  - Work with instrument developers to standardize formats
  - Build upon previous Task Force report
    - Equivalence studies for moderate changes
- Analytical approaches for evaluating mixed modes data from clinical trials

Presenters

Jean Paty, PhD, Founder & Senior Vice President, Scientific, Quality & Regulatory Affairs, invivodata, inc., Pittsburgh, PA, USA

Stephen Joel Coons PhD, Director, Patient-Reported Outcome (PRO) Consortium, Critical Path Institute, Tucson, AZ, USA

Antonia Bennett PhD, Investigator & Postdoctoral Research Scholar, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Special thanks to **Ethan Basch MD, MSc, Damian McEntegart, Karin Coyne PhD**, and **J. Jason Lundy PhD**, for their contribution to this presentation as members of the ISPOR PRO Mixed Modes Task Force

Developing Alternatives:
Before We Get to Multiple Modes

Jean Paty, PhD
Chief Scientific Advisor
PRO Consulting, invivodata, inc.
Pittsburgh, PA

Overview

- Need to develop alternative modes of data capture before mixing
  - The most common path is migrating from paper to electronic.
- Need to consider following issues:
  - Regulatory
  - Scientific
  - Operational
  - Legal

Paper to ePRO Migration

- The primary issue is that patients comprehend questions the same way regardless of mode of data capture.
- It is important to demonstrate this comprehension by hearing from patients and/or demonstrating equivalence in responses.
- It is important that the migration does not introduces changes to the measurement properties.
  - Reliability, validity, ability to detect change

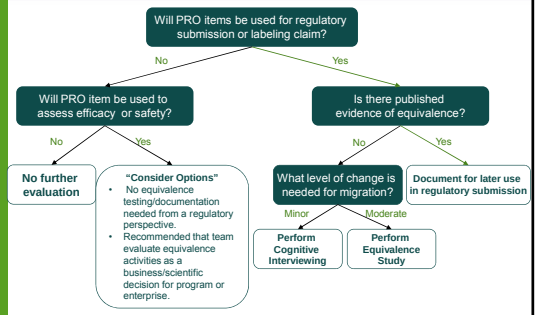
Key Considerations

- Regulatory
  - Will the PRO endpoint be used to support labeling language?
- Scientific
  - Has equivalence between paper and this particular electronic solution been previously demonstrated?
  - What level of effort is needed to demonstrate equivalence?

Key Considerations

- Operational
  - How can a 'faithful' migration be conducted that minimizes the differences between paper and the electronic platform?
- Legal
  - Has appropriate interaction occurred with instrument developer (e.g., license)?

Instrument Implementation: Paper to Electronic Equivalence Decision Tree



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

Regulatory Considerations

- Is data from PRO endpoint going to be used to support product labeling (claims)?
- If going to support labeling, the PRO Guidance is applicable
- If not supporting labeling, then a business/scientific decision
  - e.g., studies for publication only
- This can potentially alleviate significant effort.

19

Scientific Considerations

- Has the equivalence between paper and this electronic format been previously demonstrated?
- It is important that the current and published format are similar.
- What type of equivalence work as published?
  - Is it adequate to defend to regulators?
- If no published work, then need to evaluate level of equivalence work needed.

20

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

21

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 interreliability. 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. [42].

Coons et al., 2009

22

Conclusions

- Paper to electronic is a good example of issues involved in developing an alternative mode of data capture.
  - In near future, electronic to paper will be the migration direction.
- Need to develop alternative modes of administration before mixing them.
  - Level of change is key driver when developing the alternative modes of data capture.

Mixing PRO Data Capture Modes in Clinical Trials: *Issues to Consider*

Stephen Joel Coons, PhD

Patient-Reported Outcome (PRO) Consortium

Critical Path Institute

Tucson, Arizona, USA

Focus on PRO Endpoints

- The FDA's PRO Guidance has focused increased attention on the scientifically sound measurement of PRO endpoints in clinical trials.
- As the focus on PRO measures as efficacy endpoints has increased, the use of electronic data capture devices/systems has expanded dramatically as well.
- This has led to the need to assure measurement equivalence across and among the various methods and modes of PRO measure administration.

Regulatory Perspective

- As evidenced by the quote cited earlier by Sonya, it is clear from the FDA's PRO Guidance that the mixing of data capture modes is anticipated to occur within clinical trials.
- However, the PRO Guidance does not discuss ways for clinical trial designs to ensure the comparability of the data when mixed modes are used.

Why Not Mix Data Capture Modes?

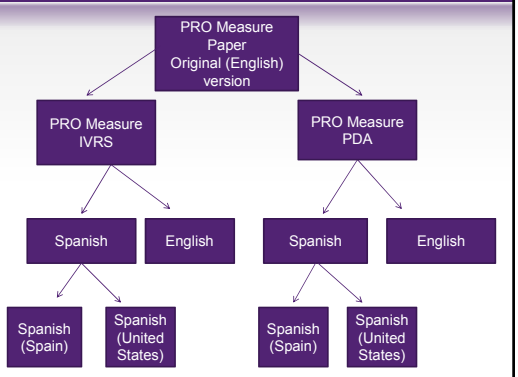
It is important to consider the reasons why you may not want to vary PRO data capture modes within a single clinical trial or between trials that seek to provide comparable data.

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

Why Not Mix Data Capture Modes?

- 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.
- Measurement error introduced by a single alternative mode of PRO measure administration may be small, but most clinical trials have multiple sources of potential response bias.
- The compounding of these sources of variance can add up and have a cumulative impact on statistical power.

Compounding Measurement Error



Randomization (1)

- Randomization helps to avoid possible bias in the selection and allocation of subjects
- It is intended to produce groups (treatment and control) in which the distributions of prognostic factors (e.g., age, severity, co-morbidities) are similar.
- During analysis of trial data, it provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects.
- Hence, as long as the pattern of mixing modes is the same in the treatment and control groups, any potential measurement error introduced by the mixed modes will be comparable across the two groups.

Randomization (2)

However....

- Even the balanced introduction of measurement error across treatment arms has the potential to put the trial at risk of not showing a treatment effect if the signal to noise ratio is decreased.
- Any change during the trial (after randomization) that leads to different data capture mode patterns across the treatment and control patients (or within treatment or control patients) has the potential to differentially introduce measurement error.

Conclusions

- Multiple sources of measurement error exist in multinational trials that could cumulatively impact the ability of the PRO data to show a treatment effect.
- To the extent possible, avoid mixing modes.
- Seriously consider all potential sources of measurement error in your trial and minimize the potential impact by maximizing measurement equivalence across data capture modes.
- Randomization in clinical trials is essential, but it does not protect against overwhelming the PRO-based treatment effect (signal) by measurement error (noise) introduced in both treatment arms.

Potential Benefits of Mixing Modes in Clinical Research

Antonia Bennett, PhD

Kathy Panageas, PhD

Ethan Basch, MD, MSc

Memorial Sloan-Kettering Cancer Center
New York, USA

Overview

- Benefits of mixing modes
  - Less missing data / increased compliance
  - More representative patient sample
- How do we measure equivalence?
- How might analysis be affected by decision to mix modes?
  - Reliability between modes vs. test-retest reliability of a single mode
  - Effect of reliability on power
  - Effect of systematic missing data on power

Benefits of mixed modes (1)

- Multiple modes within patient can reduce systematic missing data, especially data that is not missing at random (NMAR).
  - Examples of missing PRO data:
    - Patient does not have access to web or phone (e.g. traveling, staying with relative, or short-term hospital stay) .
    - Patient becomes too ill to complete PRO measure at computer but is able to use phone in bed.
- Allows for back-up data collection, e.g., phone call by study staff if patient misses web report

Benefits of mixed modes (2)

- Allows for broader and more-representative patient sample within a study
  - Across cultures or regions: Can include populations without widespread web or phone use
  - Across patients: Allows for patient preference and/or needs (e.g. hearing impaired or non-computer literate)
- Allows for comparison of results across studies that did not use same mode for PRO data capture

Slide 31

A7 ab suggested adding an example
Author, 4/29/2011

Slide 33

A6 streamline language since we decided to use modes as the main term
Author, 4/29/2011

Pervasiveness of mixed mode transactions

- In daily life, people accustomed to having choice of mode for transactions (e.g. banking, purchases, customer service, and voting).
- Many patients have busy lives or limited energy.
- In order to reach patients, some accommodation for their day-to-day life is necessary and typical.

Other sources of measurement error

- How does mixed modes compare to other accepted sources of error?
  - Differences in item text or item meaning due to language translation
  - Variation in responses due to personal and cultural attitudes (social desirability, stoicism, propensity for extreme scores)
  - Test-retest reliability of instrument in single mode

How can equivalence be tested?

- Equivalence testing can be included in validation study (e.g. instrument development or Phase II).
- Tests of equivalence
  - Correlation between modes (ICC) ***Reliability***
  - Comparison of mean scores by mode
  - Comparison of scores by sub-group or at particular range of scores: DIF, Bland-Altman, Regression
- Compare the mixed-mode equivalence with test-retest equivalence of the instrument.

Between-mode versus test-retest reliability

- If the between-mode reliability is at least as high as the test-retest reliability, then there is no loss of power by randomly mixing modes.
- In four comparisons that evaluated both, the average between-mode paper-to-computer correlation was almost identical to the test-retest correlation of the paper measure (0.88 vs. 0.91) (Gwaltney, 2008 *VIH*).

Effect of reliability on power

Reliability	Power for 95% CI
"1.00"	80%
0.99	80%
0.90	76%
0.80	71%
0.70	65%
0.60	58%

Power w/reliability estimate:

$$\sqrt{\frac{r\,n\,d^2}{2\,\sigma^2}} - Z_{\alpha/2} = Z_{\beta}$$

Note: Power for two-sample difference of means with equal variances
Reference: Fleiss JL. The Design and Analysis of Clinical Experiments. Wiley & Sons, 1986

Reliability and sample size

Reliability	N for 80% power and 95% CI	
	Effect size = 0.2	Effect size = 0.5
"1.00"	393	63
0.99	397	64
0.90	437	70
0.80	491	79
0.70	561	90
0.60	655	105

Effect of systematic missing data (1)

- Example: Patient is unable to complete PRO via the computer at a measurement time-point because they are especially ill, but would have used phone.
  - This missing data would be *not missing at random* (NMAR) because it is dependent on their illness.
- Missing the scores of the most ill patients will deflate the difference between treatment and control study arms.
  - The observed effect size will be smaller than true effect size.

Effect of systematic missing data (2)

Loss of power due to artificially small effect size	
Effect Size	**Power for 95% CI & N=63**
0.50 "True"	**80%**
0.45	71%
0.40	61%
0.35	50%
0.30	39%
0.25	29%
0.20	20%

Note: Effect Size = d/σ

2. Power:

$$\sqrt{\frac{n\,d^2}{2\,\sigma^2}} - Z_{\alpha/2} = Z_{\beta}$$

Optimal study design

Equivalence of modes **vs.** **Systematic missing data**

Summary

- Mixing modes accommodates regional differences as well as patient needs.
  - Less missing data / increased compliance
  - More representative patient sample
- We have approaches for testing and examining equivalence of modes.
- Mixed modes with good equivalence are similar to the test-retest reliability of a single mode.
- Mixing modes with high reliability may preserve power via reduced systematic missing data.

Conclusion

- Develop a Good Research Practices report
- Degree of modification reflects the risk of increased random error due to lack of measurement equivalence
  - Potential measurement error may reduce power to detect a treatment effect
  - Randomization process doesn't solve the problem
- Potential benefits - reduced missing data
- Strategies for mixing modes appropriately
- Analytical methods to assess impact on treatment effect

For more information

For more information on the ISPOR PRO Mixed Modes Task Force or to join our Review Group, please visit our webpage:

http://www.ispor.org/sigs/mixedmodes.asp

or email:

ISPOR PRO Task Force Liaison,
Elizabeth Molsen
emolsen@ispor.org

FORUM

Recommendations for Evaluating the Validation of Computerized Systems that Capture Outcomes Data in Clinical Trials

ePRO Systems Validation Task Force Members

Industry

- **Arthur Zbrozek, RPh, MSc, MBA**
  ISPOR ePRO Systems Validation Task Force Chair
  Senior Director. Global Health Economics, CSL Behring, King of Prussia, PA
- **Kenneth Grice**, Associate Director, ePRO Operations, Global Electronic Data Capture, Bayer HealthCare Pharmaceuticals Inc., Montville, NJ, USA
- **Scottie Kern**, Assoc Director, Pfizer, Ltd, Taplow, England, UK

ePRO Systems Validation Task Force Members

Information Technology

- **Gretchen Craig**, Manager of Quality Assurance, Invivodata, inc., Pittsburgh, PA
- **Gregory Gogates**, VP Regulatory Affairs, CRF Inc., Lansdale, PA
- **Joy Hebert**, Chief Operating Officer, assisTek, Scottsdale, AZ
- **Rod Thorell**, Director, Quality & Compliance, PHT Corp, Charlestown, MA

Academia

- **Raymond Panas PhD, BS, MPH**, Assistant Professor George Washington University, Washington, DC

Background

- ISPOR ePRO GRP Task Force Report (2009)
  - Substitution of paper with electronic data collection
    - Guidance for evaluating quality/validation of ePRO data collection systems was not provided.
- FDA PRO Guidance
  - *The data element collected by the ePRO device must be maintained once the element leaves the device* (FDA, 2009).

Task Force Objective

Develop a *non-technical* guide for clinical trial sponsor use (a practical source of information for study teams) on the requirements and documentation needed from a data collection systems manufacturer to demonstrate systems validation.

User Objective

To ensure that ePRO providers are using system validation and implementation processes that will ensure the systems and services:

•operate reliably when in practical use
•produce accurate and complete data and data files
•support management control
•improve sponsor confidence comply with any existing regulations.

Content Overview

- Basic validation principles
- Minimum system validation elements in context of clinical trial risk
- Background description of process quality
- Glossary of terms
- Current best practices references

For more information

For more information on the ISPOR ePRO Systems Validation Task Force or to join our Review Group, please visit our webpage:

http://www.ispor.org/sigs/ePROsystemvalidationsg.asp

or email:

ISPOR PRO Task Force Liaison,
Elizabeth Molsen
emolsen@ispor.org