FORUM

Considerations for Using Mixed Modes for Patient-reported Outcomes Data Collection in Clinical Trials and Validation of Computerized Systems to Capture Outcomes Data

FORUM

Sonya Eremenco MA
Chair, ISPOR PRO Mixed Modes Task Force
ePRO Manager, United BioSource Corporation, Bethesda, MD, USA

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

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.

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**

<table>
<thead>
<tr>
<th>Source of Variability</th>
<th>Method of Administration</th>
<th>Source of Variability</th>
<th>Method of Capture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variations due to items being seen or heard, how they appear on page or screen, number of items visible</td>
<td>Self-administered (paper)</td>
<td>Variations due to direct or indirect presence of interviewer and variation among interviewers (e.g., age, gender, personality)</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
</tr>
<tr>
<td>Patient may alter response due to presence of interviewer (e.g., social desirability)</td>
<td>Interviewer-administered (paper, IVRS, Web)</td>
<td>Patient may alter response due to presence of interviewer (e.g., social desirability)</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
</tr>
<tr>
<td>Patient may alter response due to direct or indirect presence of interviewer</td>
<td>In person – paper</td>
<td>In person – paper</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
</tr>
<tr>
<td>Patient may alter response due to presence of interviewer</td>
<td>Paper</td>
<td>Paper</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
</tr>
<tr>
<td>Patient may alter response due to direct or indirect presence of interviewer</td>
<td>Tablet</td>
<td>Tablet</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
</tr>
<tr>
<td>Patient may alter response due to direct or indirect presence of interviewer</td>
<td>In person – tablet</td>
<td>In person – tablet</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
</tr>
<tr>
<td>Patient may alter response due to direct or indirect presence of interviewer</td>
<td>Over the phone – paper</td>
<td>Over the phone – paper</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
</tr>
<tr>
<td>Patient may alter response due to direct or indirect presence of interviewer</td>
<td>Over the phone – tablet</td>
<td>Over the phone – tablet</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
</tr>
<tr>
<td>Patient may alter response due to direct or indirect presence of interviewer</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
<td>Computer-Assisted Telephone Interview (CATI)</td>
</tr>
</tbody>
</table>

**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 introduce 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?

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.

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.


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

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
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.

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.

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.)

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.

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.
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.

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

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

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.
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
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.

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)
  - 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

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Power for 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;1.00&quot;</td>
<td>80%</td>
</tr>
<tr>
<td>0.99</td>
<td>80%</td>
</tr>
<tr>
<td>0.90</td>
<td>75%</td>
</tr>
<tr>
<td>0.80</td>
<td>71%</td>
</tr>
<tr>
<td>0.70</td>
<td>65%</td>
</tr>
<tr>
<td>0.60</td>
<td>58%</td>
</tr>
</tbody>
</table>

Power w/reliability estimate:

$$\sqrt{\frac{r n d^2}{2 \sigma^2}} \cdot Z_{\alpha} = Z_{1-\beta}$$

Note: Power for two-sample difference of means with equal variances

Reliability and sample size

<table>
<thead>
<tr>
<th>Reliability</th>
<th>N for 80% power and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;1.00&quot;</td>
<td>393</td>
</tr>
<tr>
<td>0.99</td>
<td>357</td>
</tr>
<tr>
<td>0.90</td>
<td>437</td>
</tr>
<tr>
<td>0.80</td>
<td>511</td>
</tr>
<tr>
<td>0.70</td>
<td>581</td>
</tr>
<tr>
<td>0.60</td>
<td>655</td>
</tr>
</tbody>
</table>

Effect size = 0.2 Effect size = 0.5

Note: Power for two-sample difference of means with equal variances
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.

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.

<table>
<thead>
<tr>
<th>Effect Size</th>
<th>Power for 95% CI &amp; N=63</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50 &quot;True&quot;</td>
<td>80%</td>
</tr>
<tr>
<td>0.45</td>
<td>71%</td>
</tr>
<tr>
<td>0.40</td>
<td>61%</td>
</tr>
<tr>
<td>0.35</td>
<td>50%</td>
</tr>
<tr>
<td>0.30</td>
<td>39%</td>
</tr>
<tr>
<td>0.25</td>
<td>29%</td>
</tr>
<tr>
<td>0.20</td>
<td>20%</td>
</tr>
</tbody>
</table>

Note: Effect Size = d/σ

2. Power:

\[ \sqrt{\frac{n \cdot d^2}{2 \cdot \sigma^2}} \cdot Z_{\alpha/2} = Z_{\beta} \]

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.

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

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

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.
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 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
eolsen@ispor.org