Literature Review to Determine Empirical Basis for Response Scale Selection in Patient-reported Outcome Measure Development

Katharine Gries, PharmD, PhD
Senior Research Associate, Outcomes Research, Evidera

SIXTH ANNUAL
PATIENT-REPORTED OUTCOME CONSORTIUM WORKSHOP

April 29–30, 2015 ■ Silver Spring, MD
The views and opinions expressed in the following slides are those of the individual presenters and should not be attributed to their respective organizations/companies, the U.S. Food and Drug Administration, the Critical Path Institute, the PRO Consortium, or the ePRO Consortium.

These slides are the intellectual property of the individual presenters and are protected under the copyright laws of the United States of America and other countries. Used by permission. All rights reserved. All trademarks are the property of their respective owners.
Session Objective

• Present results of a literature review project sponsored by the C-Path patient-reported outcome (PRO) Consortium
  – Review and summarize empirical evidence for common response scale types to enhance response scale selection for newly developed PRO measures
Literature Review to Determine Empirical Basis for Response Scale Selection in Patient-reported Outcome Measure Development

Prepared by:

• Margaret Vernon, PhD
• Katharine Gries, PharmD, PhD
• Shima Safikhani, MPH
Project Background and Rationale

• Response scale selection for patient-reported outcome (PRO) measures is a critical part of:
  – Content validity
  – Psychometric properties
  – Scoring
  – Interpretation of the score
  – Cross-cultural relevance and comparability
  – Ease of administration/operational usability

• Despite the importance of response scale selection for new measures, there seems to be little empirical basis for type of response scale selected
  – Response scale selection is often based on therapeutic convention or preference of the developer
Project Objectives

• Conduct a comprehensive review of the scientific literature:
  – Identify response scale option types
  – Review and summarize available empirical evidence for each type of scale by context of use
    • Psychometric properties and sensitivity/responsiveness
    • Therapeutic area
    • Study population
    • Format of PRO measure
    • Recall period of PRO measure
    • Acceptability from regulatory authorities
  – Enhance response scale selection for newly developed PRO measures
Specific Literature Search Objectives

- **Search 1**: Formal guidelines / review articles
  - Recommendations on the selection of response scale types
- **Search 2**: Direct and indirect empirical evidence for response scale types
  - Psychology and survey methodology literature
- Articles with direct or indirect empirical evidence for PRO response scale selection:
  - **Search 3**: based on recall period
  - **Search 4**: based on the format of the PRO measure or response scale
  - **Search 5**: based on scoring and psychometric property implications
  - **Search 6**: based on the intended study population
  - **Search 7**: based on therapeutic indications (of interest)
- **Search 8**: Response scales used in PRO measures achieving successful regulatory review by the United States (US) Food and Drug Administration (FDA) and European Medicines Agency (EMA) and included in the approved product label
• Translatability and electronic implementation assessment of the identified response scale types
  – Feedback about format, words, or phrases which are structurally or culturally problematic when translated into different languages.
  – Implementation issues for electronic modes of data collection
    • Web-based, smart phone, tablet, interactive voice response (IVRS)
Summary of Literature Search

• Across the literature we identified and reviewed:
  – Almost 7,000 abstracts
  – Approximately 400 full text articles

• Final count: 196 unique references

• The FDA and EMA searches (search objective 8)
  – From 2006–2014
    • 36 unique PRO measures with claims in FDA labels
    • 37 unique PRO measures with claims in EMA labels
Common Response Scales Types Identified in the Literature Search
<table>
<thead>
<tr>
<th>Response Scale Type</th>
<th>Frequency Count Across Searches 3–7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Analog Scale (VAS)</td>
<td>68</td>
</tr>
<tr>
<td>Verbal Rating Scale (VRS)</td>
<td>59</td>
</tr>
<tr>
<td>Numeric Response Scale (NRS)</td>
<td>48</td>
</tr>
<tr>
<td>Faces</td>
<td>16</td>
</tr>
<tr>
<td>Other Graphical</td>
<td>13</td>
</tr>
<tr>
<td>Binary</td>
<td>5</td>
</tr>
<tr>
<td>Likert</td>
<td>3</td>
</tr>
</tbody>
</table>

Many studies included multiple scales, frequency counts do not represent unique studies.
Types of Response Scales

VRS

NRS

VAS (not to scale)
Types of Response Scales

Likert Scales

Please indicate how much you agree or disagree with each of these statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Somewhat disagree</th>
<th>Neither agree nor disagree</th>
<th>Somewhat agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The president is doing a good job.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The Congress is doing a good job.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The Secretary of Defense is doing a good job.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Compared with 3 months ago, how would you rate your health in general now?

- Much Better
- Somewhat Better
- About the Same
- Somewhat Worse
- Much Worse

Faces Scales

Wong-Baker FACES® Pain Rating Scale

- 0: No Hurt
- 2: Hurts Little Bit
- 4: Hurts Little More
- 6: Hurts Even More
- 8: Hurts Whole Lot
- 10: Hurts Worst

*Licensure required for healthcare facility use. Contact: WongBakerFACES.org or WongBakerFACES@gmail.com
Response Scale Selection for Use in PRO Measures
Response Scale Selection Based on Sensitivity/Responsiveness (n=13)

- Inconclusive to which response scale demonstrates greater responsiveness, however there was a slight trend in favor of NRS over VAS
- Dependent on the construct being measured
- 6- or 7-point VRS more responsive than 4- or 5-point VRS
- 6- or 7-point VRS may be as responsive as NRS or VAS
Response Scale Selection Based on Therapeutic Area

- **Asthma** (n=3)
  - VRS typically used
  - For ages 4–11 years old: fewer response options or VRS assisted by graphics

- **Pain** (n=62)
  - 11-point NRS likely the optimal response scale
    - Adult patients without cognitive impairment

- **Fatigue** (n=9)
  - VAS or NRS suitable
    - The NRS was more responsiveness than the VAS in exercise capacity study

- **Oncology** (n=14)
  - VAS or NRS typically used
    - 11-point NRS favored over VAS for ease of implementation, scoring, and good psychometric properties
  - Faces scales appropriate for pediatric populations
Response Scale Selection Based on Pediatric Population (n=31)

- Visual orientation of scales, emotions expressed in graphical scales, and word choice for verbal anchors can produce unexpected biases due to immaturity in abstract thinking skills of respondents
  - Child might choose a numerical response based on favorite number rather than representation of experience.
    - NRS and VAS suitable for children 7 and older
  - Faces scales appear to be well understood and produce reliable and valid estimates for children 4–12 years
  - VRS may be most suitable for older children (10–12 years and older) who are confident readers and have a higher level of reading comprehension
    - 5-point VRS may be appropriate for children as young as 7 years old
    - Likert scale appropriate for children 7 or 8 years of age and older
Response Scale Selection Based on Geriatric Populations (n=7)

• Evidence is somewhat inconsistent
  – VRS or NRS recommended over VAS or graphical scale
  – VAS preferred over 5-point VRS or a 4-faces scale
  – 7-point VRS more widely accepted and understood than a faces or NRS
  – NRS preferred by surgical patients over VAS or 6-point VRS
  – VAS and 5-point VRS yielded similar reliability and responsiveness, however the VRS was found to have superior interpretability
  – 4- and 5-point VRS demonstrated more stability than an 11-point NRS because of difficulty to complete as cognitive impairment increased
Translatability and Cultural Adaptation

• Choice of verbal anchors needs to be considered

• NRS
  – For the 0 end of the scale, common anchors include None, Not at all, Never, No pain, No [symptom], and Absent
    • *Absent* is particularly challenging to translate because it is not a simple term in the medical context and is not patient-friendly language

• VRS
  – Need to achieve similar intervals between the verbal descriptors to match the source language
  – Amounts of time “How much of the time...?” is difficult to translate
    • Simple frequency scales using terms such as Never, Rarely, Sometimes, Often, Always are recommended

• Faces Scales
  – There is little empirical evidence available to demonstrate cultural comparability across languages/cultures
Electronic Feasibility

- **NRS**
  - Easily implemented via Interactive voice response systems (IVRS) and other electronic modes
    - Formatting needs to be considered so that anchors are associated with intended number

- **VAS**
  - Impossible to implement in IVRS
  - Difficult to implement in other electronic modes due to need for specified length (100 mm)

- **VRS**
  - Number of response options and length of verbal descriptors should be carefully selected so as to lighten cognitive load (for IVRS)
  - Equidistant formatting on 1 screen for handheld/tablet devices
Discussion and Recommendations
VAS widely used in the PRO literature reviewed and generally psychometric properties and responsiveness are strong.

• May be appropriate for use in pediatric populations over the age of 7 or 8
  – BUT may not be as well understood in geriatric populations or populations with cognitive impairment

• Impossible to implement in an oral format

• Difficult to migrate to electronic visual formats due to requirements that the scale have a specific length

• Might be more difficult to implement (administer and score consistently) than other scale types
NRSs were widely utilized in PRO literature reviewed and generally psychometric properties and responsiveness are strong for this scale

- Preferred scale for pain measurement
- Pediatric Population: does not always produce comparable scores
  - NRS may not be advisable for use in younger age groups (4 to 11+ years)
- Suitable for translation and cross-cultural adaptation
- Easy to implement in various modes of administration including IVRS and electronic visual formats
VRSs were widely utilized in the PRO literature reviewed and generally psychometric properties and responsiveness are strong for this scale type

- Scales with more response options (5 to 7) have better performance and responsiveness than scales with fewer response options (3 to 4)

- Selection of verbal anchors is important as acceptability and understanding is dependent on literacy levels of the sample

- Scale performance dependent on assumptions of equidistant intervals

- More difficult to translate and cross-cultural adaptation

- Typically used in multi-item scales
  - PRO measures to support FDA and EMA labeling claims
Faces and Other Graphical Scales

Less widely utilized in the PRO literature reviewed
— More research is needed regarding performance of these scales in specific contexts of use

• Faces scales work well in young pediatric samples, in adult samples, and in cross-cultural contexts
  — Pre-literate participants can more readily abstract their experience to the emotion expressed in the faces as compared to a verbal or numerical scale
  — More research is needed regarding cross-cultural comparability of facial cues presented in the faces response options

• Cannot be administered orally (IVRS) but could be administered via electronic visual modes
Discussion and Recommendations

• About 1/3 of articles identified were in the pain therapeutic area
  – Optimal response scale choice for pain may not always be generalizable to other contexts of use
  – Pain assessments typically consist of single-item questions which use either a VAS or an NRS response scale
    • Multi-item questionnaires also use these response scale types but more commonly use a VRS
  – Pain assessment may be different than other measurement concepts which require multiple items

• Response scale selection needs to take into account multiple factors
  – Target population
  – Study design
  – Concept of interest
  – Recall period
  – Data collection mode
  – Scale responsiveness
Discussion / Questions
Key References

• Example Scale Sources
Key References

- **Therapeutic area**
Key References

- **Study Population**
Backup Slide
## Search Criteria: Empirical Evidence For PRO Measure Response Scale Selection

<table>
<thead>
<tr>
<th>Type</th>
<th>Search Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response scale terms</strong></td>
<td>'response scale':ab,ti OR 'response scales':ab,ti OR 'likert scale':ab,ti OR 'likert scale'/exp OR 'visual analog scale':ab,ti OR 'visual analog scales':ab,ti OR 'visual analogue scale':ab,ti OR 'visual analog scale'/exp OR 'numerical rating scale':ab,ti OR 'numerical rating scales':ab,ti OR 'verbal rating scale':ab,ti OR 'verbal rating scales':ab,ti OR 'competence scale':ab,ti OR 'competence scales':ab,ti OR 'frequency scale':ab,ti OR 'frequency scales':ab,ti OR 'extent scale':ab,ti OR 'extent scales':ab,ti OR 'comparison scale':ab,ti OR 'comparison scales':ab,ti OR 'performance scale':ab,ti OR 'performance scales':ab,ti OR 'developmental scale':ab,ti OR 'developmental scales':ab,ti OR 'qualitative scale':ab,ti OR 'qualitative scales':ab,ti OR 'agreement scale':ab,ti OR 'agreement scales':ab,ti OR 'categorical scale':ab,ti OR 'categorical scales':ab,ti</td>
</tr>
<tr>
<td><strong>PRO terms</strong></td>
<td>'patient satisfaction'/exp OR (patient* NEAR/2 satisfaction):ab,ti OR (patient* NEAR/2 reported):ab,ti OR 'self report'/exp OR (self NEAR/1 report*):ab,ti OR 'patient preference'/exp OR (patient* NEAR/2 preference*):ab,ti OR (patient* NEAR/1 assess*):ab,ti OR 'self evaluation':ab,ti OR 'self evaluations':ab,ti OR (patient* NEAR/2 rating):ab,ti OR (patient* NEAR/2 rated):ab,ti OR 'self-completed':ab,ti OR 'self-administered':ab,ti OR (self NEAR/1 assessment*):ab,ti OR 'self-rated':ab,ti OR 'patient based outcome':ab,ti OR 'self evaluation'/exp OR experience*:ab,ti</td>
</tr>
</tbody>
</table>
Literature Review to Determine Empirical Basis for Response Scale Selection in Patient-reported Outcome Measure Development

Prepared by:
• Margaret Vernon, PhD
• Katharine Gries, PharmD, PhD
• Shima Safikhani, MPH