Multinational Trials - What translations are required, what methodology should be used and what methodology supports pooling the data?

Patient-Reported Outcomes Task Force: Changing Culture or Language
FORUM
PRO Task Force

Background:

- Task force initiated in March 2007
- Representatives from academia, translation and linguistic validation providers and industry
- Representatives from Europe and North America
Members

- **Chair:** Diane Wild MSc - Oxford Outcomes Ltd
- **Isabelle Mear MA** – MAPI Research Institute
- **Ingela Wiklund PhD** - GlaxoSmithKline
- **Lawrence Cohen PharmD, BCPP, FASHP, FCCP** – Washington State University
- **Sonya Eremenco MA** – United BioSource Corporation
- **Mary Gawlicki MBA** – Corporate Translations
- **Caroline Houchin MA** – Oxford Outcomes Ltd
- **Lee Yee Chong BSc** – Royal College of Surgeons of England
- **Mona Martin RN, MPH** – Health Research Associates
- **Asha Hareendran PhD** – Pfizer Ltd
- **Robyn von Maltzahn MSc** – Oxford Outcomes Ltd
Presenters:

Isabelle Mear MA, Director, MAPI Research Institute

Sonya Eremenco MA, ePRO Manager, UBC

Diane Wild MSc, Director, Oxford Outcomes Ltd
Goal

This work is a continuation and expansion of

*Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation.*

Issues to Address

1. Selection of languages required for translation for specific countries

2. Translation methods for ‘same language’ versions used in multiple countries

3. Issues and concerns around pooling of trial data across countries
Methods

- Monthly teleconferences
- Goals identified and divided among 3 subgroups
- Development of report
- Dissemination for review to:
  - ISPOR Patient-Reported Outcomes Special Interest Group & Task Force membership
  - ISPOR PRO CCL TF Forum
Plans

- Integration of comments into Task Force report
  - May 2008

- Submission of manuscript to Value in Health
  - June 2008
FORUM Overview

- Selection of languages required for translation for specific countries

- Translation methods for ‘same language’ versions used in multiple countries

- Issues and concerns around pooling of trial data across countries

- Discussion
Sub-group 1

Selecting the number and specific languages required for translation in a country

Sub-group Chair: Isabelle Mear
Ingela Wiklund
Lawrence J.Cohen
Objectives

Identify a process for selecting the number of languages required for adequate coverage of global trial populations
Participants

- Academics
- Providers of linguistic validation services
- Industry
Methods

- Literature review
- Design draft tool
- Update of the tool

Tested by 4 linguistic validation companies

- Assess consistency of information
- Assess ease of use
- Develop instructions

Decision-making Tree
### 1. Analysis of the Population

**Goal**: determine the nature of the population

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin of natives</td>
<td></td>
</tr>
<tr>
<td>Immigration in % of population</td>
<td></td>
</tr>
<tr>
<td>Origin of immigration</td>
<td></td>
</tr>
<tr>
<td>Location of immigrants</td>
<td></td>
</tr>
<tr>
<td>Literacy rate in country</td>
<td></td>
</tr>
</tbody>
</table>
## Analysis of the languages

**Goal:** determine the likelihood of the need for a language according to the criteria mentioned below

<table>
<thead>
<tr>
<th>“Name of Country”</th>
<th>Location of the sites</th>
<th>Associated predominant language</th>
<th>Targeted number of patients per site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“city or region”</td>
<td>“language A”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“language B”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Main languages in “Name of Country”</th>
<th>“language A”</th>
<th>“language B”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Official language (Y/N)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral/Written language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of mother language</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location of speakers (indicate any specific location + whether they live in urban or rural areas)</th>
<th></th>
</tr>
</thead>
</table>

**Conclusion:** language needed or not needed

---

Draft Tool II
Results

- Use of reliable source of information and documentation
- Mention both percentage (%) and number of speakers Epidemiology
- List languages from most common to the least common
- Number of languages?
Addition of a Decision-making Tree

- **Country**
  - **Official national language(s)**
    - Percentage of speakers > 3%
      - To be considered
    - Percentage of speakers < 3%
      - No
  - **Other language(s)**
    - Location of sites
      - Predominant language(s)
        - Percentage of speakers > 3%
          - To be considered
        - Percentage of speakers < 3%
          - To be considered
      - Other language(s)
        - Percentage of speakers > 3%
          - To be considered
        - Percentage of speakers < 3%
          - No

**Yes:** if low literacy rate or mainly oral language, only interviewer-administered forms for the language

- **Instructions for completing the grids**
Conclusion

- No magic recipe
- Tool shared by all decision-makers
  - Facilitate decision-making process
  - Help develop a better rationale
Sub-group 2: Same Language, Different Country

Sub-group Chair: Sonya Eremenco
Mary Gawlicki
Caroline Houchin
Lee Yee Chong
Introduction

- After language selection, need to choose methodology for translation


- How to approach creating instrument in same language for multiple countries not addressed
Objectives

- Present possible approaches for same language, different country situation
- Provide options including advantages and disadvantages via scenarios
- Provide decision-tree flowchart
- Explain methods to implement these approaches
Background

- Reasons for same language, different countries
  - Colonialism
  - Boundary changes
  - Immigration

- Greater differences between spoken language versus written language
Methods

- Literature review
  - Few publications addressed issue
  - Recommendations for specific instrument (EORTC, EQ-5D, FACIT)

- More information needed in this area
Possible Approaches

- Country-specific approach
- Linguistic adaptation
- Universal approach

**Combined approach:** begin from universal and result in country-specific or vice versa

**NOT recommended:** assuming translation for one country will work in others with no further evaluation
Scenarios

- In practice, many variations are possible depending on instrument.

- Scenarios intended to illustrate how these approaches are applied in real-life situations, including trade-offs in terms of time and cost.
Scenario 1: Multiple Spanish-speaking countries, no existing translation

- Develop 4 separate translations independently (Spain, Mexico, US, Argentina)

- Translate and debrief for Spain. Second translation for one of the other countries, adapt from second translation to remaining 2 target countries
Scenario 1: Multiple Spanish-speaking countries, no existing translation

- Translate for Spain. Remaining 3 countries adapt it for their country.

- Translate for Spain. Translators from remaining 3 countries work together to develop 1 version.

- Translators from all 4 countries work together to develop 1 version.
Scenario 2: French for 4 countries, France version exists

- Separate new translations for Belgium, Canada, Switzerland
- Translators from each country adapt the French (France) version for their country
Scenario 2: French for 4 countries, France version exists

- Group countries regionally: Europe and Canada. One adaptation from France version for Belgium / Switzerland, separate for Canada

- Translators from 4 countries work together on 1 version
Country-specific

**Advantages**
- More colloquial and idiomatic usage
- Timelines may be shorter

**Disadvantages**
- Variation in translations may introduce bias, reflect stylistic differences
- Costs for CRF printing may be higher
Linguistic Adaptation

**Advantages**
- Similar to country-specific
- Allow for terminology specific to country

**Disadvantages**
- Fewer translators involved so increased influence of personal opinion over linguistic necessity
Universal Approach

Advantages

- Same question asked to all groups, less likelihood of bias from differently worded questions

- Practical and logistical benefits
Universal Approach

- **Disadvantages**
  
  - Wording may not sound as natural
  
  - May be less straightforward if future countries are needed and not in original development
Factors in Deciding Approach I

- No one approach suitable for all
- Need to consider languages/countries involved and content of the measure

  - Cultural similarities and differences
  - Some subject matter cannot be universal
Factors in Deciding Approach II

• Some subject matter cannot be universal:
  - Educational systems
  - Demographics
  - Health Resource Utilization
  - Government institutions and programs
  - Computer terminology
Start

Does Developer/client have a methodology preference?

Yes: Do as he wants.

No: Does a same language version already exist?

Yes: Is the existing version very old/poor?

Yes: Are the countries very similar geographically/culturally/linguistically to the existing version? (eg France/ Belgium or Columbia/ Venezuela)

Yes: Consider linguistic adaptation approach.

No: Does the text contain wording about education/ health systems/ brand names/ IT?

Yes: Linguistic adaptation approach

No: Universal approach

No: Are the countries very similar geographically/culturally/linguistically to existing version? (eg France/ Belgium or Columbia/ Venezuela)

Yes: Produce new, country-specific versions.

No: Use existing version everywhere (cognitively debrief first).

No: Use linguistic adaptation approach
Country-specific Methods

- Country-specific new instrument should follow consensus guidelines

- Linguistic adaptation of existing translation:
  - Start from existing version
  - 2 native speakers review and comment
  - Compare the comments, look for agreement, discuss discrepancies and resolve them
Country-specific Methods

Linguistic adaptation of existing translation:

- Back-translation from target to English
- Implement changes and proofread
- Cognitive debriefing in new target country
Universal Methods

- Forward translators from multiple locales
- Reconciler familiar with different countries
- After back-translation, additional country reviews
- Translators work together to find solutions acceptable to all
Universal Methods

- Proofreading

- Cognitive debriefing in target countries

- If agreement is not reached or if patients find problems, then country-specific variations based on the universal may be created
Conclusion

- Complex issues to address

- No single approach is right for all situations, all have advantages and disadvantages

- Assess each study case by case with regard to languages, countries, and instrument content

- More empirical research is needed
Sub-group 3: Issues around the aggregation of data from multiple languages and cultures

Sub-group Chair: Diane Wild
Mona Martin
Asha Hareendran
Robyn von Maltzahn
Data is often pooled across multiple languages/cultures

Over the last decade there has been an expansion in the variety of countries included in clinical trials
Challenges I

• Differences in clinical data across countries could result from:
  
  - nuances of clinical practice standards of care
  - site personnel, procedures and so on

• An additional challenge for PROs can be the differences resulting from cultural and linguistic differences in the expression of outcomes.
Challenges II

- Current literature does not address this issue adequately.
- There are currently no established criteria to assure the appropriateness of the aggregation of data derived from multiple languages and cultures.
Many types of equivalence can be considered for pooling PRO data from multiple sources.

- Literacy levels
- Cognitive equivalence of the concepts
- Equivalence in the translation/adaptation methods
- Measurement equivalence
Measurement
Equivalence
Definition

- The degree to which an instrument can yield reliable and valid data about some phenomenon of interest across different populations (Horn and McArdle 1992, Mullen 1995)

- The extent to which individuals in different cultural groups respond to similar items in similar ways (Bullinger et al 1993)
Methods for Demonstrating Measurement Equivalence

- A diverse range of methods are employed in order to assess equivalence – including:
  - Classical test theory
  - Factor Analysis
  - SEM
  - DIF
Classical Test Theory

- A body of psychometric theory concerned with understanding and improving the reliability of tests

- An observed score is the sum of the true score and a measurement error term
Commonly Used CTT Methods

- Internal consistency (using Cronbach Alpha)

- Components can have different means and variances, but covariances should be equal – implying membership of a single factor in factor analysis.
Commonly used CTT methods

- Equivalent forms using Pearson correlation coefficient
- Test-retest
- Item to Scale correlation
Factor Analysis

- A statistical technique used to explain variability among observed variables in terms of a smaller number of unobserved variables – termed factors.

- Factors may represent subscales or domains.
Factor Analysis

- Factor loadings of instrument items can be examined cross-culturally.

- The relationship between factor scores with other outcome measures can be assessed in each population (construct validity).
A French translation of the Strauss and Carpenter revised outcome criteria scale (2004) was assessed using factor analysis. The similarity of factor loadings was taken as support of the instrument structure.
Structural Equation Modelling (SEM)

- A technique for testing and estimating causal relationships using a combination of statistics and theoretical assumptions

- It is hypothesis-driven. In this instance, that instruments function in a similar fashion for different populations.
Structural Equation Modelling (SEM)

- Model fit is examined using a number of indices (dependent upon the methodology employed).

- Models are modified to produce the best explanation of the data.
Circadian Rhythm Questionnaire (Horne & Ostberg) demonstrates sound CTT properties, but poor SEM model fit (Smith et al 1991).

Potential problems associated with using classical measurement theory to assess between-population differences.
Differential Item Functioning (DIF)

- Originated in education test settings to determine whether individual test items are unfair to particular groups (e.g. females or ethnic minorities) even after allowing for individual ability

- Can be conducted in various ways (contingency table, logistic regression, item response theory techniques)
Use of DIF in QoL of cancer patients

- IRT methods used to assess DIF in QoL (measured by the EORTC QLQ-C30) of cancer patients across 4 ethnic groups (Caucasian, Filipino, Hawaiian and Japanese).
- DIF found on several items even though overall QoL was the same (Pagano and Gotay, 2005).
<table>
<thead>
<tr>
<th>Item</th>
<th>Content</th>
<th>Standardized Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Caucasian</td>
</tr>
<tr>
<td>01</td>
<td>Physical</td>
<td>1.09</td>
</tr>
<tr>
<td>02</td>
<td>Physical</td>
<td>0.72</td>
</tr>
<tr>
<td>03</td>
<td>Physical</td>
<td>-1.51</td>
</tr>
<tr>
<td>04</td>
<td>Physical</td>
<td>-1.12</td>
</tr>
<tr>
<td>05</td>
<td>Physical</td>
<td>-2.15</td>
</tr>
<tr>
<td>06</td>
<td>Role</td>
<td>0.81</td>
</tr>
<tr>
<td>07</td>
<td>Role</td>
<td>-1.79</td>
</tr>
<tr>
<td>08</td>
<td>Dyspnea</td>
<td>-0.15</td>
</tr>
<tr>
<td>09</td>
<td>Pain</td>
<td>0.90</td>
</tr>
<tr>
<td>10</td>
<td>Fatigue</td>
<td>1.50</td>
</tr>
<tr>
<td>11</td>
<td>Insomnia</td>
<td>1.05</td>
</tr>
<tr>
<td>12</td>
<td>Fatigue</td>
<td>0.23</td>
</tr>
<tr>
<td>13</td>
<td>Appetite Loss</td>
<td>-0.11</td>
</tr>
<tr>
<td>14</td>
<td>Nausea</td>
<td>-0.96</td>
</tr>
<tr>
<td>15</td>
<td>Nausea</td>
<td>-2.23</td>
</tr>
<tr>
<td>16</td>
<td>Constipation</td>
<td>-0.34</td>
</tr>
<tr>
<td>17</td>
<td>Diarrhea</td>
<td>-0.44</td>
</tr>
<tr>
<td>18</td>
<td>Fatigue</td>
<td>1.26</td>
</tr>
<tr>
<td>19</td>
<td>Pain</td>
<td>-0.35</td>
</tr>
<tr>
<td>20</td>
<td>Cognitive</td>
<td>-0.42</td>
</tr>
<tr>
<td>21</td>
<td>Emotional</td>
<td>0.44</td>
</tr>
<tr>
<td>22</td>
<td>Emotional</td>
<td>1.10</td>
</tr>
<tr>
<td>23</td>
<td>Emotional</td>
<td>0.49</td>
</tr>
<tr>
<td>24</td>
<td>Emotional</td>
<td>0.47</td>
</tr>
<tr>
<td>25</td>
<td>Cognitive</td>
<td>0.54</td>
</tr>
<tr>
<td>26</td>
<td>Social</td>
<td>0.07</td>
</tr>
<tr>
<td>27</td>
<td>Social</td>
<td>0.71</td>
</tr>
<tr>
<td>28</td>
<td>Financial</td>
<td>0.26</td>
</tr>
<tr>
<td>29</td>
<td>Physical</td>
<td>0.08</td>
</tr>
<tr>
<td>30</td>
<td>Overall</td>
<td>-0.22</td>
</tr>
</tbody>
</table>

Values shown in bold are significantly greater than values shown in italics. Values in plain type do not vary significantly from other values. \( \chi^2(3) \) column provides the chi-square value (with 3 DF) for the test of difference across all four ethnic groups. The \( p \) value is for the chi-square. An asterisk (*) denotes that \( p \) value is less than the alpha value which was set to 0.0017 (0.05/30).
Addressing DIF

- Effect of ethnicity could be partialed out using regression to make all ethnicities equal on that item

- A system of weighting could be employed to counteract differences
What degree of measurement equivalence is needed to support data pooling?

It may be possible to verify basic measurement properties (distribution, internal consistency etc) in a few languages;

Or in a pivotal study supporting a label claim evaluate basic measurement characteristics such as means, standard deviations etc when the samples are large enough (Patrick et al 2007).
If measurement equivalence is lacking what steps could be taken to investigate and resolve?

- Investigation: qualitative research, analysis of existing trial datasets, consultation with in-country KOLs
- Resolutions: culturally specific weightings, remove country from analysis
Conclusions

- Data pooling across languages/cultures is common practice.
- Lack of guidance on how to determine whether pooling is appropriate.
- Lack of guidance on level of measurement equivalence to be met in order to pool data.
- Need for practical steps to be taken in order to investigate/resolve lack of measurement equivalence.
- Need for more research and practical guidance.
Questions for the panel?