ROADBLOCKS AND RESOLUTIONS: DEVELOPMENT OF PRO INSTRUMENTS FOR MULTINATIONAL TRIALS

Bruce Crawford, Syneos Health, Tokyo, Japan
Hyunchung Ray Kim, Syneos Health, Tokyo, Japan
Tara Symonds, Clinical Outcomes Solutions, Folkestone, UK
Louise Newton, Clinical Outcomes Solutions, Manchester, UK

ISPOR Barcelona; 10-15 November 2018

Objectives of this workshop

- To provide a better understating of simultaneous development of patient-reported outcome (PROs) instruments (and other clinical outcome assessments (COAs)*)
- To present challenges and issues of simultaneous development
  - Examples from our experiences
- To suggest solutions and share best practice methods

* COAs include PROs, observer-reported outcome (ObsRO) clinician-reported outcome (ClinRO) and performed-based (PerfO) measures
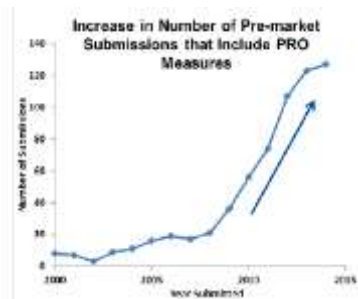
1

Background

3

Expanding use of COA / PRO instruments

- Growing presence of PRO measures in studies
  - In FY2017, 75+% of approved, pivotal original and new study IDEs submitted to the Office of Device Evaluation included a PRO measure
- More multinational studies
- More studies with highly varied cultures

Figure 3. Yearly distribution of ingredients regarding PRO reporting

Increase in Number of Pre-market Submissions that Include PRO Measures

Number of submissions including PRO measures for calendar years 2000 – 2015.

CDRH Strategic Priorities 2016-2017 report
https://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHVisionandMission/UCM588576.pdf
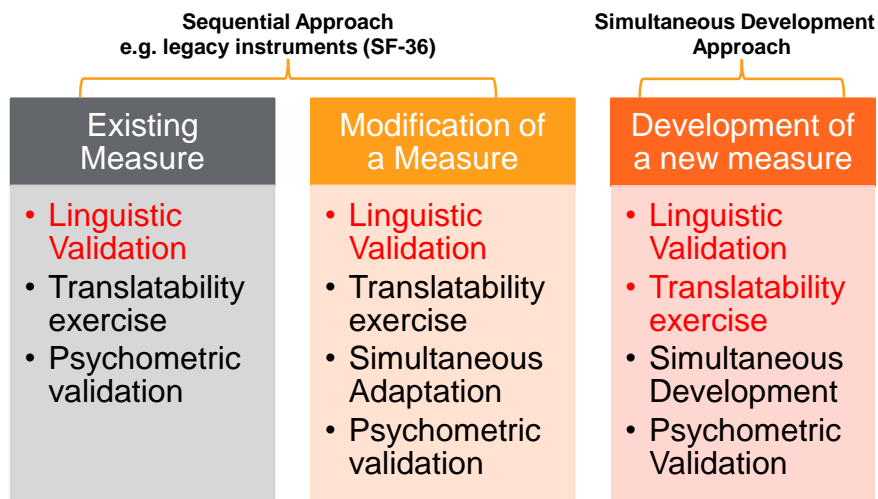
4

Current issues in use of Global COA instruments

- COA measures are commonly used in global clinical trials
- COA measurement development is often restricted to a single language and subsequently translated or culturally validated into additional languages or countries
- Translation and linguistic validation cannot always ensure the **cultural equivalence and relevance** of an adapted COA instrument, necessitating the development of country-specific instruments

Options for development of language versions of COA measures in multi-national trials

Sequential Approach e.g. legacy instruments (SF-36)		Simultaneous Development Approach
Existing Measure	Modification of a Measure	Development of a new measure
• Linguistic Validation • Translatability exercise • Psychometric validation	• Linguistic Validation • Translatability exercise • Simultaneous Adaptation • Psychometric validation	• Linguistic Validation • Translatability exercise • Simultaneous Development • Psychometric Validation

3

Advantages of simultaneous development

- Simultaneous Development (SD) of a COA instrument in the languages and cultural groups relevant to the research program is the most efficient approach to ensure valid COA data and interpretable results[1,2].
  - The likelihood of conceptual equivalence and relevance is increased
- SD of COA instruments offers advantages over traditional linguistic validation methods[3]
  - It provides local language versions earlier: no need to translate later
  - It helps ensure translation challenges are identified early and avoided
  - Enhances pooling of data and comparability across countries
  - Saves time in global studies

Example: Translation of the MusiQOL

*"The initial **simultaneous development** of the MusiQoL in 14 countries **minimized the challenges** faced during the translation **in the 57 additional languages**" Poster presented at ISPOR Milan by Boucher et al (2015) doi.org/10.1016/j.jval.2015.09.2491*

1. Wild et al. Principles of good practice for the translation and cultural adaptation process for patient- reported outcomes (PRO) measures: report of the ISPOR task force for translating adaptation. Value Health 2005;2:94-104
2. Patrick et al. Content validity - Establishing and reporting the evidence in newly-developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1 - Eliciting concepts for a new PRO instrument. Value Health 2011;14; 967-977
3. Conway & Patrick Translatability assessment. In: International Workshop on Comparative Survey Design & Implementation Available from: https://www.csdiworkshop.org/images/Conway.pdf (accessed 31st Oct 2018)

Sometimes Concepts are Understood but Not Relevant

Physical Function	ALWAYS TRUE	USUALLY TRUE	SOMETIMES TRUE	RARELY TRUE	NEVER TRUE
Because of my weight I have trouble tying my shoes.	5	4	3	2	1

- Many Japanese do not tie their shoes or wear them loosely tied – there is a need to take shoes off when entering homes
- Japanese is more direct – "because I'm fat" not "because of my weight"

Transferability Of Patient Reported Outcomes In Diabetes: Cultural Differences In Japan

Crawford et al. ISPOR (September 7-9, 2008, Seoul, Korea)

Going Back To Look for Conceptual Equivalence Can Be a Daunting Process

- NPSI was being used in multinational trials and US FDA asked for evidence that concepts being used in various language versions was equivalent
- 132 patients; 5 countries compared to the US to establish conceptual equivalence
  - Pre-focus group questionnaires, focus groups
  - Lots of discussions around meaning: Chinese: "heart stabbing", "needle through heart"
    - More likely to relate extreme pain with the heart because they believe the heart is the most critical and sensitive part of the body

Crawford et al. *Health and Quality of Life Outcomes* 2008; 6 (62): 1-8

Achieving agreement on difficult to translate terms

- A major advantage of simultaneous cross-cultural development is that it facilitates early identification of difficult to translate words and phrases
- These can be discussed among the linguists, instrument development specialists and clinicians at the item generation meeting to reach a consensus
- Sometimes challenges of translating into a specific language can lead to adjustments in the text in the reference language

Direct patient comments should be used in item development
But needs to be culturally relevant. . . to everyone

- Asian patients' statements while discussing jaw pain: "I avoid eating dried squid" while talking about their eating habits

- How many Australians said this??
- We decided to use "chewy food" if we were to list an example so the item can have universal examples

What is simultaneous development?

What is simultaneous development?

- Development of a COA measure in several languages / cultures simultaneously
- Different ways of conducting SD:
  1. Reference Language Approach
  2. Parallel Language Approach

Reference Language Approach

- **Step 1: Select the reference language or "core" language version of the COA instrument**
  - This approach enables the instrument's conceptual framework to evolve in a **consistent** manner across different language versions at each stage.
  - Ensures final measurement concepts and instrument content are **equivalent** regardless of which language version of the COA is being used.
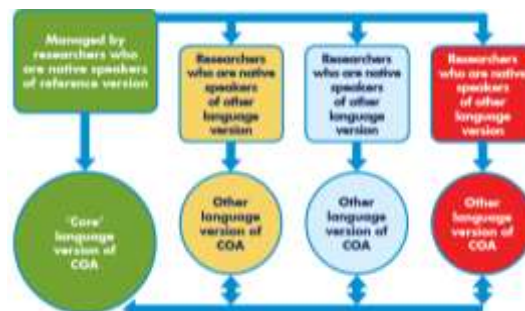
Taken from: Simultaneous item development (SID) for clinical outcomes assessments (COAs) Humphrey, L. et al. ViH 2014; 17 (3): A207 - A208

Reference Language Approach

- **Step 2: Translate and adapt fieldwork materials**
  - Develop interview guide in core language and then translate to other relevant languages once key concepts of interest are defined
  - Use a **Question-to-Concept Map** ensure same conceptual topics are discussed despite difference in actual wording on guide
  - Forward translation usually sufficient for other materials (such as consent form) but *confirm cultural appropriateness of content*, for example country-specific disease criteria, severity categorisation

Question-to-Concept Map (QCM): example				
Section and No.	Interview guide question	Underlying concept	Definition of concept	Additional comments
Symptom experiences: Q7	Tell me about your most recent "attack" [subject's word for panic attack] experience	Panic attacks	Period of intense fear or apprehension of sudden onset	Focus on description of the severity and sensation of experience. Frequency/duration covered later

Syneos Health

15

Reference Language Approach

- **Step 3: Interview population of interest across different languages and cultures**
  - Interview methods follow best practice[1]. Methods must be robust and pragmatic to overcome the **inherent complexities** of multinational/cultural research.
  - Important that the interview guide is **well translated**, and that all interviewers are extensively briefed and clear on the aims of the interview – use role-play.
  - Interviewers:
    - should be native speakers of the language the interview will be conducted in
    - ideally also fluent in the reference language (often English)
    - if fluent in a common language, all interviewers should attend a central briefing to ensure they take a consistent approach
  - There should be opportunity for the interviewers to ask questions and highlight if any aspects of the guide are **inappropriate for their language/culture**
    - For example, in some cultures it may be inappropriate to ask questions about sex life
  - **Review** as interviews conducted:
    - Ideally core research team reviews first few interviews for each language to ensure the interviewer has correctly understood the aims of the interview and is probing appropriately. This can be a time consuming approach exercise.

Syneos Health

16

1.   Patrick DL, et al. Value Health 2011;14; 967-977

Reference Language Approach

- **Step 4: Analyse cross-cultural qualitative data**
  - Ensure differences in experiences between different languages or cultures are adequately captured in the content of the newly developed COAs
    - Be aware of ***Linguistic vs Experiential*** differences
  - "Core" research team develops codebook and codes reference language transcripts
    - Other language transcripts are then forward translated into the core language and are analysed simultaneously
  - Qualitative analysis completed in advance of the item generation meeting (IGM) and the coded content from the different languages combined centrally
    - May develop draft item content for reference language prior to IGM to facilitate discussion/ development
    - Consider the targeted **mode of administration** (pen/paper, hand held electronic device, Interactive Voice Recognition System [IVRS], web-based), **recall period** and format for **response scales**
      - These factors can impact on item content, and the impact can differ by language
      - E.g. German words are typically much longer than other languages, and so require much more screen space if using a device such as tablet or smartphone

1.   Patrick DL, et al. Value Health 2011;14; 967-977

Reference Language Approach

- **Step 5: Generate content for COA suitable for different language and cultural groups**
  - Create a **conceptual model** of COA instrument, ideally one that is applicable across all languages/cultures
    - Highlight in model any differences identified between languages/cultures
  - Generate items for each item in conceptual model
    - Item development should be undertaken by ***COA experts, disease experts*** (to ensure concepts are clinically relevant) and ***interviewers/ linguists*** in collaboration (ideally at face-to-face IGM)
    - Involving interviewers:
      - Able to reflect on the interviews they conducted, ensure the concept/content reflects the natural language used by interview participants and facilitate accurate simultaneous translation of COA content as it evolves during discussion
  - Use an **Item Tracking Matrix**
    - Concept name and definition of concept
    - Response scale and options
    - Core language content and translations including rationale for any modifications between versions

Reference Language Approach

- **Step 5: Generate content for COA suitable for different language and cultural groups (cont)**

  Item Generation Meeting (IGM)

  - Aim to ensure all instructions, items and response scales are *conceptually equivalent* in all languages
    - *Remember Linguistic vs Experiential* differences
  - Will not always be literal, word for word translations, what is important is that the **conceptual meaning** is translated and is the same
  - Where proposed wording does not translate easily, IGM participants should discuss to see if the original can be adjusted – otherwise the aim is to find a conceptually equivalent phrase
  - This can require quite a lot of discussion backwards and forwards
  - Critical to keep referring back to the qualitative findings, and also to get input from expert clinician(s) – to ensure resulting items are **relevant to the population ('Context of Use')**, **clinically relevant** as well as being **conceptually equivalent across cultures**

Reference Language Approach

- **Step 6: Assess the face and content validity of COA for cross-cultural use**
  - Interview methods follow best practice regardless of languages conducted in[1]
  - 'Core' research team lead and manage the process
  - Analysis of cognitive debriefing data is first completed by the 'core' research team and then complete the analysis of the non-core language transcripts (once translated).
  - Analysis of language sub-groups to explore differences between patients from the different countries

- **Step 7: Harmonize and Finalize**
  - Use Item Tracking Matrix to document and justify revisions
  - If content validation has only be conducted in the reference language, conduct a translatability assessment to optimize measure for country adaptation or if other languages required beyond those involved in Step 6

1. Patrick DL et al. Value Health 2011;14; 967-977

Parallel Language Approach

- Main difference between reference approach and this approach relates to **Step 4 and Step 6**
  - *Rather than the non-core transcripts translated into reference language and analysed by core research team, native speakers in each language conduct qualitative analysis using source language transcripts*

  - **Benefits**: If any errors or mistranslations occur from the source language transcript to the translated transcript, the research team may never pick this up as they are only analysing the translation. In contrast, if native-speaking analysts code data in source language, these nuances/subtleties will be captured during coding and not "missed" due to translation mistakes.

  - **Disadvantages**: Often the same kinds of experiences can be coded with different concept names between research teams adding many redundant and inconsistent codes to the codebook. Process of harmonization becomes very lengthy and complex. Delays data analysis and interpretation.
- Also, if using parallel approach, different language content is amalgamated and combined centrally and draft versions of all COA versions created before IGM
  - Each linguist is responsible for developing the language version for their own language, and highlighting if any phrases or concepts are difficult to translate during the meeting

Challenges surrounding simultaneous development

Major challenges identified during simultaneous development

1. Lack of well-established regulatory guidelines
2. Linguistic and cultural differences not easily understood by all stakeholders
3. Technical challenges in coding and analyzing qualitative data in multiple languages
4. Item generation and determining conceptual saturation across languages/cultures
5. Making decisions about whether revisions necessary for comprehension in one language/culture should be implemented in all languages
6. Requires longer timelines and incurs greater cost in the early stages of development

Linguistic and cultural differences

- Some linguistic and cultural differences are not easily understood by all stakeholders

- Determining the relevance of a concept across different cultures and language and matching conceptually equivalent items

- Ensuring each item and the full translated sentence sounds natural
i.e "right now my item literally reads like this in English… what do you think?" and revise if needed

Examples – Linguistic differences

- The word order can highly influence the nuance of the questions.
- Inappropriate word order translation can lead to reduced understanding of the question (e.g. Japanese vs English word order)

[Sentence in the core language; English]
I feel confident about my appearance even without wearing mascara

||

[The literal English translation of the Japanese version]
Without wearing mascara, my eyelashes make me feel confident about my appearance

© 2018 All rights reserved | **Confidential** | For Syneos Health™ use only   25

Addressing cultural differences

When eliciting key concepts relevant to daily life activities

A RA study, which patients mobility is affected when taking a bath
- Taking a bath in a "bathtub", rather than a shower is a common culture for many Japanese

When asked about disease impact, Japanese patients would respond
"It is inconvenient/ bothersome when I try to get into the bathtub for a bath every night"
This is not a common culture in other countries. We had to take this concept out, when adapting the questionnaire to other countries

Syneos Health   CS
© 2018 All rights reserved | **Confidential** | For Syneos Health™ use only   26

13

Example: Concept of "wide face"

- Concept of "wide face/ big face"
  - Asian patients reporting that "I feel like my face is too big compared to my body"
- Asian cultures vs non Asian cultures
- We decided to debrief both options and eventually kept what was understood by all cultures
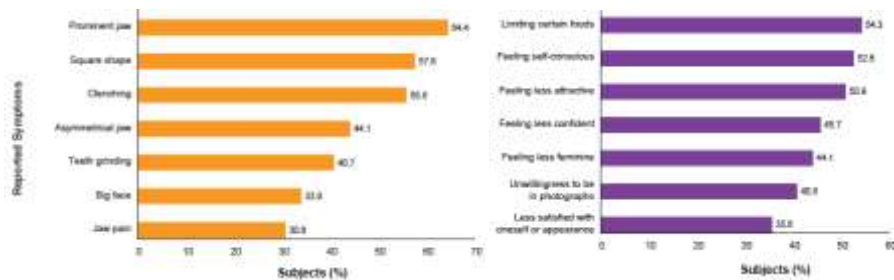
Crawford et al. Value in Health, Vol. 21, S145

Qualitative analysis and coding in original language vs target language (from a technical point of view)

- There are advantages if verbatim transcripts are analysed in the original language
  - "Parallel SD Approach"
  - Helps ensure subtleties of meaning are captured
  - Means original verbatim content is available when developing items, instructions, etc

- However, there are advantages to coding in a reference language (often English)
  - "Reference SD Approach"
  - Can allow coding to be performed centrally, by the same research team, ensuring consistency of coding
  - Facilitates easier combining of results across languages, and allow language differences and comparisons to be made
  - Limits likelihood of technical issues occurring
  - There is not always someone adequately trained in qualitative analysis for all languages

14

When a word (expression) contains various meanings

[Example]

- 困る(komaru)； A common expression used when one is placed in an un-wanting situation, a feeling or a reactive expression against a negative situation
- The meaning/ nuance of the expression subtly changes depending on the context of the situation

Weight gain due to my current diabetes treatment is "KOMARU"

Stressful?

Anger?

Troubled?
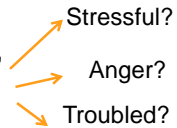
We can relate to this sentiment, but from a clinical point of view, does this sentence give you enough information to extract an item?

Why is the patient mentioning "komaru" because of weight gain?   What concept would you derive out of this?

Item generation and determining conceptual saturation

- Challenges about determining conceptual saturation across languages/cultures and ensuring item wording is conceptually equivalent

- Item generation meetings can be a very long process depending on the number of languages
  - time-consuming for some of the items, instructions, and response options
  - However, no need to translate later

- Lack of documentation on discussions and rationale especially while generating items
  - Why did local researchers use the term XXX instead of YYY that he originally intended
  - Need to ensure everyone documents thoughts / rationale very thoroughly

Example-Onomatopoeia to describe subjective symptoms

◆Pain is considered a difficult therapeutic area to properly treat

◆Doctors have a pain vocabulary sheet to ensure this is well captured.

◆For a monolingual person, it is a huge challenge to express their subjective pain in a different way. Even if we ask them to describe the pain, "throbbing" is "throbbing" and "zukin zukin" is "zukin zukin" ! "my pain is "zukin zukin" after "chiku chiku"

◆So what is zukin zukin? How do we code this to an English concept?!

◆How can we ask questions in a better way during cognitive debriefings for people from different cultures to understand the concepts and determine its conceptual equivalence

◆Phrase questions in a contextual way. "When you say "zukin zukin" is it a similar pain when you experience a paper cut? Or like a toothache? When do you usually experience this type of pain?

◆Constant verification with bilinguals is a must ! Further validation with experts is highly recommended

Wrap up

• Simultaneous development can help filter out non-universal and match conceptually equivalent items from different cultures and languages

• Core language codebook with subcategories that can potentially capture similar concepts across different cultures – Better documentation

• Preferable for the coding analysts, interviewers and project manager to be bilingual and familiar with various cultures

• Simultaneous development offers several advantages; however, bigger investment upfront is required

16

Industry Perspective

Three industry case studies

1. Development of a PRO measure for use with women taking estrogen plus progestin therapies
2. Development of a PRO measure for use with Fibromyalgia patients
3. Development of PRO measure to assess sexual function in women

Development of a PRO measure for use with Fibromyalgia patients

- Conducted qualitative interviews with 40 fibromyalgia patients
  - 20 in US
  - 10 in France
  - 10 in Germany
- Saturation analysis was conducted
  - 10/10 for US and then compared the 10 from France and Germany to US data

Results overview

- Spontaneous reports of symptoms
  - Pain most commonly reported (78%)
  - Fatigue (43%)
  - Sleeping problems (18%)
- Greatest bother 89% (31/40) reported pain
- Fatigue also mentioned as bothersome, however
  - Differed across country (11/17 US, 2/10 Germany, 5/8 France)
- In a collage exercise…
  - pain-related concepts were depicted by 45% of participants
  - Fatigue/tiredness shown by 25% of participants, none of whom were from Germany

Summary of Fatigue Data

Table 2 Summary of the characteristics that define FM fatigue

FM fatigue characteristics	US (n = 20)	France (n = 10)	Germany (n = 10)	Total (n = 40)	Example Quote

- US were most focussed on fatigue as key symptom of their condition, Germany the least but also more limited in France
  - This could be because in Germany and France there is only one word for both fatigue and tiredness?
  - True cultural difference in willingness to report fatigue?

Syneos Health

37

Menopausal Symptoms

- Study conducted to explore cross-cultural experience of women taking estrogen plus progestin therapies
  - Side effect of bleeding/spotting and breast tenderness is issue for adherence
- Interviews with women conducted in 4 countries by trained qualitative interviewers:
  - US (n=14)
  - Italy (n=15)
  - Mexico (n=15)
  - China (n=15)
- Native speakers but also spoke English (except the Chinese interviewer)
  - Important for item generation meeting
- Centralised briefing on the aims of the study to aid consistency of interviews

Syneos Health

38

Saturation

- Saturation analysis was conducted in the US initially
  - Set 1 n=5
  - Set 2 n=5
  - Set 3, n=4
- The Mexican, Italian and Chinese transcripts were then compared to the US concepts to ensure all concepts of relevance in these countries were also captured.

Summary of pain descriptors used

- Many descriptors used with the majority using the term pain. Tenderness was not utilised by women ou...
  de...
- Ot...
  th...
  En...
  -
- Co...
  consistency

  - US version pain/tenderness item
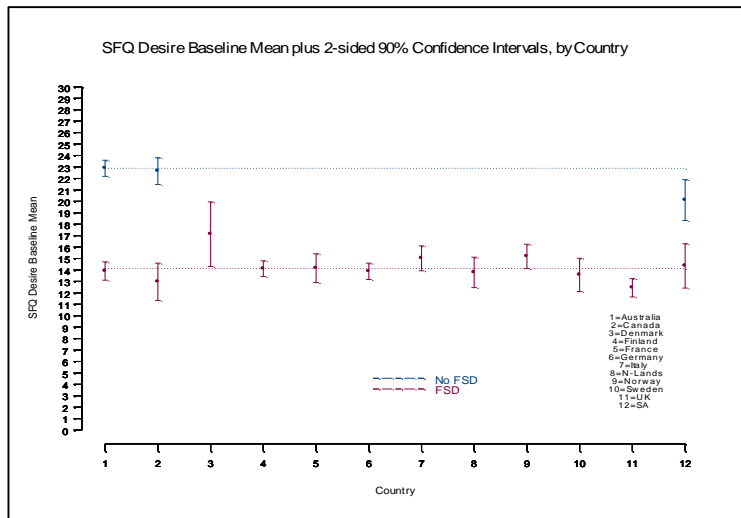 - Mexican and Italian pain only

Sexual Function Questionnaire (SFQ)

• Culturally diverse condition
• Initial cross-cultural development
  – US, UK, Australia, Netherlands, Denmark, France, Italy
• Resulting items were assessed for potential translation issues
  – Wording changed on a number of items
  – 4-items removed
• Psychometrics completed using data from 4 studies
  – 2 European clinical trials (n=577; n=204)
  – 2 Non-FSD matched controls (n=108; n=97)

Numbers Per Country

Country	FSD	No FSD
Australia	105	132
Canada	25	49
Denmark	8	
Finland	132	
France	42	
Germany	135	
Italy	54	
N-Lands	48	
Norway	60	
Sweden	32	
UK	99	
SA	17	20

SFQ Desire Baseline Mean plus 2-sided 90% Confidence Intervals, by Country

SFQ Desire Baseline Mean plus 2-sided 90% Confidence Intervals, by Country

1=Australia
2=Canada
3=Denmark
4=Finland
5=France
6=Germany
7=Italy
8=N-Lands
9=Norway
10=Sweden
11=UK
12=SA

No FSD
FSD

Country

SFQ Desire Baseline Mean

Example of Discriminate Validity using Desire Domain of Sexual Function Questionnaire

Compared baseline values for Desire between FSD and non-FSD subjects

Score range is 5-31

   High scores indicate 'normal' desire, lower scores indicate lower desire

Ignoring country effect gives 22.5 for non-FSD and 13.9 for FSD

Adding country as an effect in ANOVA gives 'adjusted means' of 22.9 for non-FSD and 14.1 for FSD

Interaction test non-significant (p ~ 0.5)

Note small sample size for Denmark and SA

Available Approaches to Assess Cross-Cultural Validity

Pool data ignoring country effect psychometrically
  Longitudinally adjust for country

Perform separate country analyses

Better approach is to be more systematic using IRT or Classical Test methodologies
  Uses whole dataset at once
  Separate results can also be obtained per country and then compared back to the reference language

Item reduction use stratified factor analysis or Differential Item Function analysis

Test / Re-test validation can use stratified Kappa statistic and IRT
  Assess time x language interaction between the two time points

Known-groups validity
  Can use Analysis of Variance (ANOVA) model at the score-level with post-hoc comparisons to the reference language when a main effect of language is indicated in the ANOVA

Summary of Industry Perspective

- Simultaneous development of language versions will help with some additional precision in measurement
  - This especially so with conditions/concepts that have more cultural diversity
- However, is the additional cost & time to use this approach worth it?
- Limited funds/time or a rare disease?
  - Then a translatability exercise is highly recommended as a viable alternative prior to use in a trial
  - Psychometric validation can be conducted using a multi-national study to inform any country effects

Q&A Session

Syneos
Health

Contact details:

Bruce Crawford, MA, MPH
Vice President

Hyunchung Ray Kim, MPP
Senior Research Manager

Syneos Health
5F, Urbannet Nihonbashi 2-chome
Building
2-1-3 Nihonbashi, Chuo-ku
Tokyo 103-0027 Japan
Mobile +81 80 4110 8475
bruce.crawford@syneoshealth.com
ray.kim@syneoshealth.com

Tara Symonds, PhD
Chief Science Officer

Louise Newton, MSc
Senior Director

Clinical Outcomes Solutions
Unit 68
Folkestone Enterprise Centre
Shearway Business Park
Shearway Road
Folkestone
Kent
CT19 4RH