PROmats

Educational Program on Patient-Reported Outcomes (PRO) in Clinical Trials

European Medicines Agency, London, UK
September 23rd, 2005

Facilitators: Neil Aaronson, PhD; Catherine Acquadro, MD; Olivier Chassany, MD, PhD

Agenda

Time	Type of session	Content	Speakers
9:15 - 9:30	Presentation	**Welcome & Objectives of the day** **Introduction to PROmats**	C Acquadro
9:30 - 10:30	Presentation	**Various issues surrounding the assessment of QoL (or PROs)** ➢ What is QOL / PRO?	N Aaronson
10:30 – 10:45 Coffee Break			
10:45 - 11:45	Interactive groups	**PROmat 3: How is a new questionnaire developed?** Development of items and item reduction ➢ Describe the process of item generation and item reduction	N Aaronson, C Acquadro & O Chassany
11:45 - 12:45	Interactive groups	**PROmat 4: How is a new questionnaire developed?** Psychometric validation and cultural adaptation ➢ Describe the evaluation of psychometric properties: reliability, validity, and responsiveness ➢ Describe the process of cultural adaptation	N Aaronson, C Acquadro & O Chassany

Agenda

12:45 – 14:00 Lunch			
14:00 - 15:15	Interactive groups	**PROmat 5: Choosing an appropriate existing measure** ➢ Explore the process for selecting an appropriate PRO instrument for use in a specific clinical trial scenario ➢ Examine the trade-offs in the selection process ➢ Review the criteria necessary to evaluate a PRO instrument appropriately ➢ Identify and evaluate established questionnaires for use in a specific patient group	N Aaronson, C Acquadro & O Chassany
15:15 - 16:15	Presentation	**Analysis & Interpretation of HRQL data included in clinical trials**	O Chassany
16:15 - 16:30	Presentation	**Concluding Remarks**	N Aaronson, C Acquadro & O Chassany

Course objectives

Lecture

- Have an overview on concepts of Patient Reported Outcomes (PRO) and Health-related Quality of Life (HRQL)
- Have an overview of the analysis and interpretation of HRQL data included in clinical trials

Course objectives

PROmats modules

- Identify the different steps for the development and validation of a PRO intrument

- Choose an existing measure appropriate to the purpose of the application

Who are we?

Neil Aaronson, PhD

Head, Division of Psychosocial Research & Epidemiology, The Netherlands Cancer Institute, Amsterdam

Professor, Chair in Psychosocial Oncology, Faculty of Medicine, Vrije Universiteit, Amsterdam

Olivier Chassany, MD, PhD

Medical Co-ordinator & Assistant Director of the Clinical Research & Development of Public Assistance Department of Hôpitaux de Paris (AP-HP)

Reviewer at the French Medicines Agency (AFSSAPS) and at the EMEA

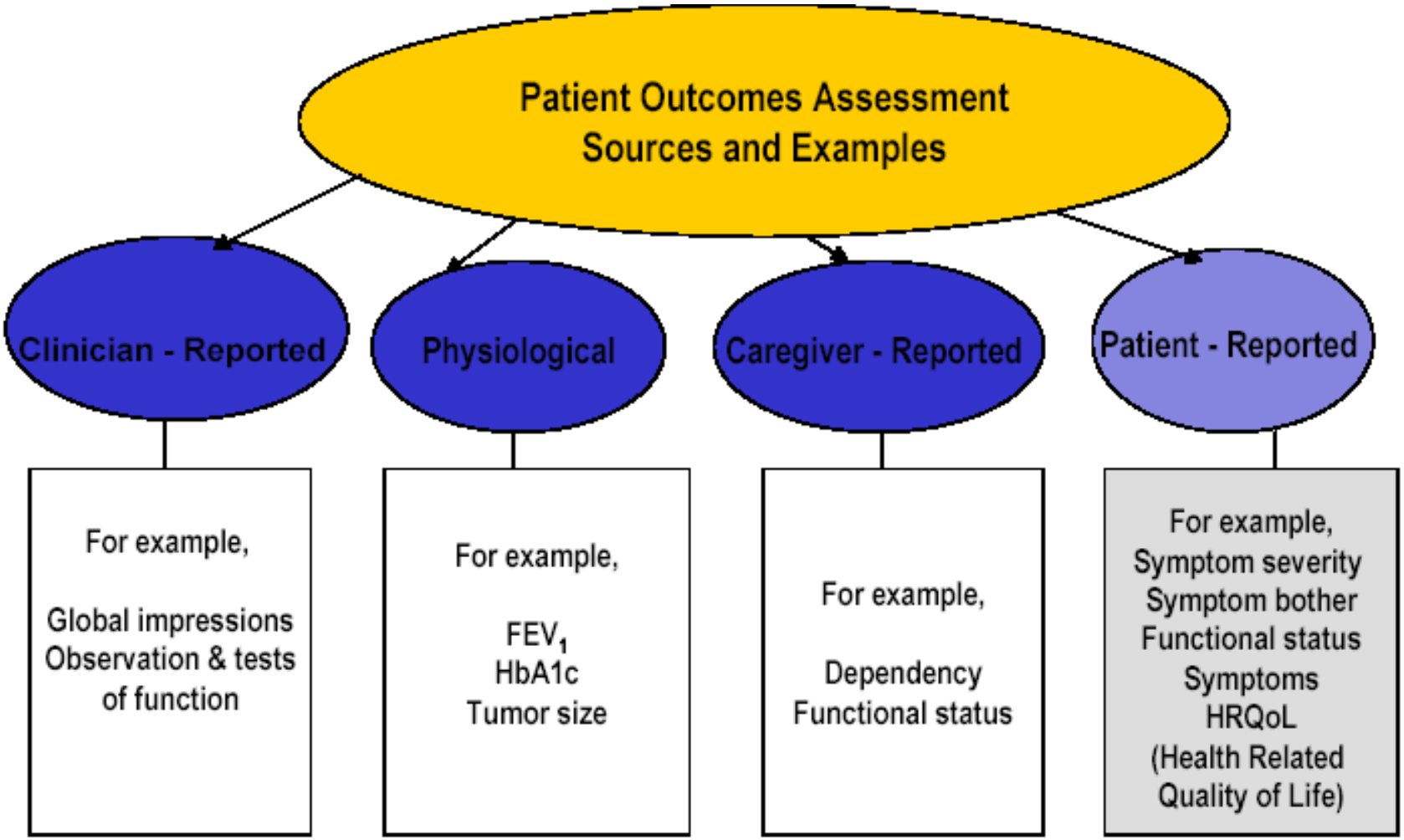
Catherine Acquadro, MD

Scientific Director of the MAPI Research Institute & Mapi Research Trust (Lyon, France)

Coordinator of the ERIQA Group and the Patient-Reported Outcomes (PRO) Harmonization Group

Co-convenor of Cochrane Collaboration Patient-Reported Outcomes Methods Group

Lecture 1



PROmats Introduction

Background

- **PROmats is an interactive educational program on PRO consisting of 7 modules**
- **1st version** developed by Adelphi and Mapi Values in 1995 on model from airline industry
- **Adaptation** and development for a Program on HRQL/PRO in Clinical Trials in 2002, collaboration between:
  - **Mapi Research Institute/Trust, Lyon, France**
    Juliette Longin, PhD
  - **The ERIQA Group**
    Olivier Chassany, MD, PhD
  - **The Cochrane HRQL Methods Group**
    Catherine Acquadro, MD
    Gordon Guyatt, PhD
    Donald L. Patrick, MSPH, PhD

Programme Objectives

- **To help** pharmaceutical companies, reviewers, and investigators of clinical trials acquire the skills needed to assess PRO included in regulatory files and publications

- **To facilitate decisions** made by health authorities and health-care providers

- **To facilitate dialogue** between regulators, members of pharmaceutical companies, and health-care providers through the same training

Interactive Learning Method

- Participants divided into ***small working groups***

- Assignments completed through group discussion

- Group answers discussed by all the members of the group to reach a consensus

PROmats description

Each PROmat is a Large Worksheet consisting of three parts:

- **Part 1:** A brief **Background** introducing the theme
- **Part 2:** The **Objectives** of the assignments to achieve
- **Part 3:** **Assignments** to be completed through group discussions enabling participants to arrive at a joint conclusion.

Take-away material

Each participant receives some documents in order to have additional background information on PRO

- Participant's Workbook
- A copy of the slide presentation
- Articles and copies of key PRO instruments

Content

PROmat	Content
1	How do disease and treatment impact upon a patient – from the patient's perspective?
2	Deciding which PRO to assess the impact of disease and treatment
3	How is a new PRO questionnaire developed? Development of items and item reduction
4	How is a new PRO questionnaire developed? Psychometric validation and cultural adaptation
5	Choosing an appropriate existing measure
6	Analysis of PRO data
7	Presentation and interpretation of PRO included in clinical trials

How is a new PRO questionnaire developed? Development of items and item reduction

Objective

- To describe the process of item generation and item reduction

Development of a PRO Instrument: Stages

Hypothetical basis for PRO as outcome → **Elicitation of concepts to be measured**
Literature review
Patients interviews

↓

Choice of a questionnaire format → **Item generation**

↓

Pilot test
Face & content validity

↓

Item reduction validation
Translations & Psychometric validation

↓

Communication & publication

Instrument ready to be used in a trial ← **Item reduction validation**

Involvement of a steering committee *(recommended)*

Patient/Expert Interviews

- **Identify/confirm relevant HRQL domains**
- **Test/confirm drug hypotheses:**
  - drug benefit meaningful for patients
  - domains on which the focus should be put
  - how to best measure those domains
- **Identify possible cross-cultural issues**

ASSIGNMENT 1

OSTEOARTHRITIS PATIENTS INTERVIEWS(collection)

"When I'm in acute pain, I can't move . I put my mattress on the floor and lie down. If I have to go somewhere in the house, I crawl around on all fours. The pain wakes me up at night; once or twice a night, often more. I often can't sleep at night — I try to lie on my back, on the right, the left..try everything. Sometimes you get very depressed — you can get angry Your moods change easily and I can be aggressive and bad tempered . I avoid anything that involves my fingers — can't do a lot of jobs to help with the children, changing diapers, carrying them. Before I was very active. I went dancing and bowling. Now I struggle to do the garden. It's also had an influence on my sex life. I'm often in a bad mood and moaning , and sometimes just can't be bothered My children have difficulty in understanding too — sometimes I'm unbearable, don't want to speak , feel like doing nothing. t affects me emotionally as I'd like to go out and do things but I can't I used to see my friends from the bowling club. Now I don't see them any more. You don't feel that you are any good in company. You don't feel like socializing when you're in pain. Sometimes I forget to take my treatment because it upsets my stomach. I know I can't do the job I used to do — it hurts me too much — I can't lift heavy things or stand up for too long...in the future I suppose I'll have to find some other work. Sometimes I think I should be taking another treatment. Not being able to work affects the family financially , but the most important thing is to fight and get on with your life. I don't want to be a burden on other people and I feel frustrated by my lack of independence, but I'm determined not to give in to it and let it ruin my life. I guess it's something you just have to learn to live with, there's no treatment ."

Table 1: Types of PRO

Type of PRO	Concepts	Attributes
Symptoms	Reports of physical and psychological symptoms or sensations not directly observable, e.g., energy and fatigue, nausea, irritability	Frequency, severity, bothersomeness
Functional Status		Frequency, difficulty, severity, ability, with help
Physical	Functional limitations and activity restrictions, e.g., self-care, walking, mobility, sleep, sexual	
Psychological	Positive or negative affect and cognitive, e.g., anger, alertness, self-esteem, sense of well-being, distress	
Social	Limitations in work or school, participation in community	
Health-Related Quality of Life (HRQL)	HRQL is **multidimensional** and represents the patient's evaluation of a health condition and its treatment on daily life: physical function, psychological function, social function, role, function, emotional function, well-being, vitality, etc.	Frequency, impact, intensity, severity, bothersomeness
Perceptions		Frequency, severity/intensity, satisfaction
Global	General ratings of health and quality of life, e.g., satisfaction or overall well-being	
Worries and Concerns	About health, finances, future	
Advantage/Opportunity	Perceptions of stigma or reports of discrimination because of health condition	Frequency, impact
Treatment Satisfaction	Evaluations of treatments	Frequency, intensity
Treatment adherence	Reports or observations of actual use of treatments	Frequency

Adapted from: Patrick and Erickson, 1993; Patrick and Chiang, 2000

The ERIQA Group

THE COCHRANE COLLABORATION®

HRQL Methods Group

3- HOW IS A NEW PRO QUESTIONNAIRE DEVELOPED?
Development of items and item reduction

BACKGROUND

The development of an instrument is a long process, involving: item generation; item scaling, item aggregation into domains (e.g. HRQL); item reduction; and demonstration of acceptability, reliability, validity, and responsiveness.

Patients play a key role in an instrument's development, especially at the item generation phase.

This Workmat explores some initial steps of the Patient-Reported Outcomes (PRO) questionnaire development, i.e.: item generation, item scaling, and item reduction.

ASSIGNMENTS

Assignment 1:
Among the PRO identified in Workmat 2, choose 1 PRO and devise 2 or 3 questions that you think would help to evaluate the subjective opinion of patients with osteoarthritis.

Assignment 2:
A number of scoring systems have been used to measure PRO in questionnaires; some examples are shown in this assignment. Compare how the different scales can be used to measure pain, and discuss which of the scales is the most accurate and responsive.

Assignment 3:
A colleague has just developed a new questionnaire to measure osteoarthritis pain and stiffness symptoms for a clinical trial. She asks for your expert review.
Identify as many problems as you can with the content and quality of the questionnaire.

Assignment 4:
A sample of patients have filled the draft questionnaire. Your colleagues ask you to review their responses. Using the data presented in this assignment, keep the most relevant items and explain your decisions.

ASSIGNMENT 1 — Defining relevant PRO and generating scale items

Choose 1 PRO and devise 2 or 3 questions that you think would help to evaluate the subjective opinion of patients with osteoarthritis. Discuss the clarity, conceptual content, and appropriate language, (absent of medical jargon) of each.

PRO selected :
Question 1
Question 2
Question 3

ASSIGNMENT 2 — Developing response choices

a) The table below shows possible scoring systems using a typical Visual Analogue Scale (VAS) (a), a Likert scale (d), an alternative format of VAS (numerical) (e), and other severity scales (b and c). Now imagine that you are suffering from a condition that causes chronic pain (such as a headache or lower back pain). Mark a score on each of the scales below to indicate how you feel. Then record the score for each scale and enter it in the box. Compare the results and discuss the different findings within the group.

Score

a) No pain at all — 0 — 100 — Worst pain imaginable

b) Do you have pain right now? Yes ❶ No ❷

c) Indicate the severity of your pain: Mild ❶ Moderate ❷ Severe ❸

d) Indicate the severity of your pain:

None	Very Mild	Mild	Moderate	Somewhat Severe	Severe	Extremely Severe
0	1	2	3	4	5	6

e) No pain at all — 1 2 3 4 5 6 7 8 9 10 — Worst pain imaginable

b) Which of these scales is likely to be most accurate and sensitive to change over time? Explain your reasoning.

_____

ASSIGNMENT 3 — How do you ensure the content and quality of the questionnaire?

a) Review the items in the questionnaire below.
Discuss within the group and write down any possible problems with the questionnaire and suggest changes that would improve the items or response choices.

_____

_____

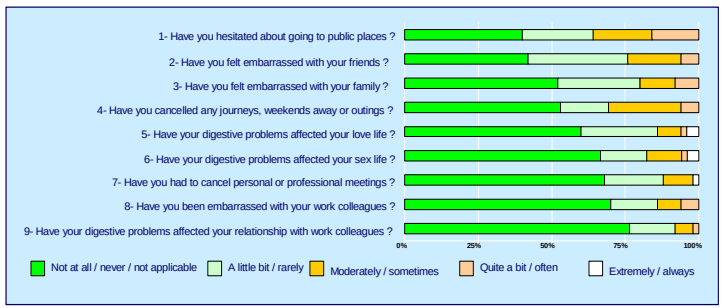
_____

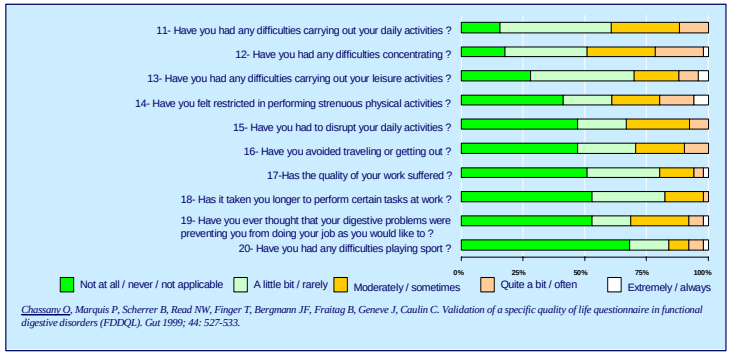
Osteoarthritis questionnaire

ITEMS	PROBLEMS
1. During the past month, how much of the time has severe pain been a problem for you? 1 All of the time — 4 Some of the time 2 Most of the time — 5 A little of the time 3 A good bit of the time — 6 None of the time	
2. During the past month, how much of the time has epigastralgia been a problem for you? 1 All of the time — 4 Some of the time 2 Most of the time — 5 A little of the time 3 A good bit of the time — 6 None of the time	
3. During the past month, how distressed have you been by morning stiffness? 1 Not at all — 4 Moderately 2 A little — 5 Extremely 3 Somewhat	
4. Could you easily scratch your low-back with hand? 1 All days — 4 Few days 2 Most days — 5 No days 3 Some days	
5. During the past month, how distressed have you been with problems moving around? 1 Not at all — 4 Moderately 2 A little — 5 Extremely 3 Somewhat	

ASSIGNMENT 4 — Item reduction (Keeping the most relevant items)

During the development of a new PRO questionnaire in irritable bowel syndrome, patients were asked to answer these items ranging from "not at all" to "extremely." Results are presented as a percentage of patients. Study the data below and decide which items should be deleted. Give reasons and criticize the method.





Chassany O, Marquis P, Scherrer B, Read NW, Finger T, Bergmann JF, Fraitag B, Geneve J, Caulin C. Validation of a specific quality of life questionnaire in functional digestive disorders (FDDQL). Gut 1999; 44: 527-533.

ASSIGNMENT 2

Developing response choices

Score

a) No pain at all 0 ——— 100 Worst pain imaginable ☐

b) Do you have pain right now? Yes ❶ No ❷ ☐

c) Indicate the severity of your pain Mild ❶ Moderate ❷ Severe ❸ ☐

d) Indicate the severity of your pain: ☐

None	Very Mild	Mild	Moderate	Somewhat Severe	Severe	Extremely Severe
0	1	2	3	4	5	6

e) No pain at all 1 2 3 4 5 6 7 8 9 10 Worst pain imaginable ☐

ASSIGNMENT 3

How do you ensure the content of the questionnaire?

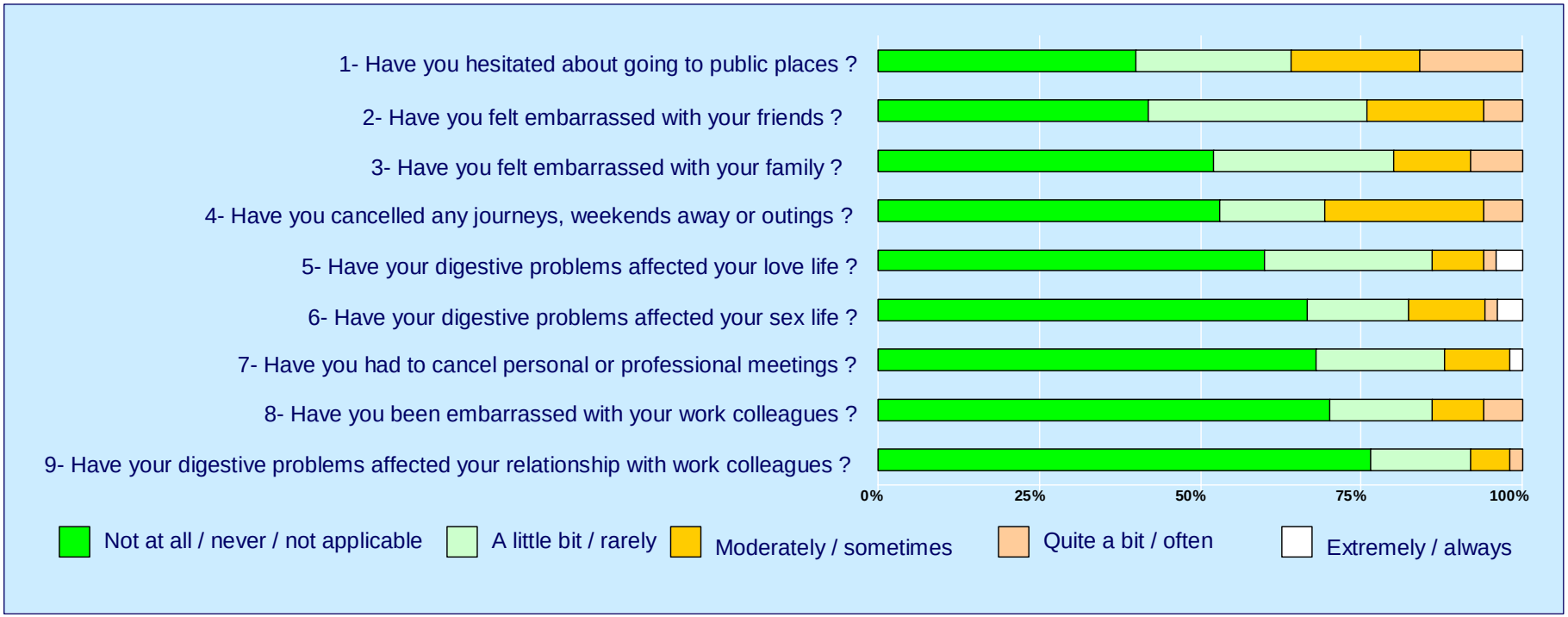
Osteoarthritis questionnaire

ITEMS	PROBLEMS
1. During the past month, how much of the time has severe pain been a problem for you? 1 All of the time 2 Most of the time 3 A good bit of the time 4 Some of the time 5 A little of the time 6 None of the time	
2. During the past month, how much of the time has epigastralgia been a problem for you? 1 All of the time 2 Most of the time 3 A good bit of the time 4 Some of the time 5 A little of the time 6 None of the time	
3. During the past month, how distressed have you been by morning stiffness? 1 Not at all 2 A little 3 Somewhat 4 Moderately 5 Extremely	
4. Could you easily scratch your low-back with hand? 1 All days 2 Most days 3 Some days 4 Few days 5 No days	
5. During the past month, how distressed have you been with problems moving around? 1 Not at all 2 A little 3 Somewhat 4 Moderately 5 Extremely	

ASSIGNMENT 4

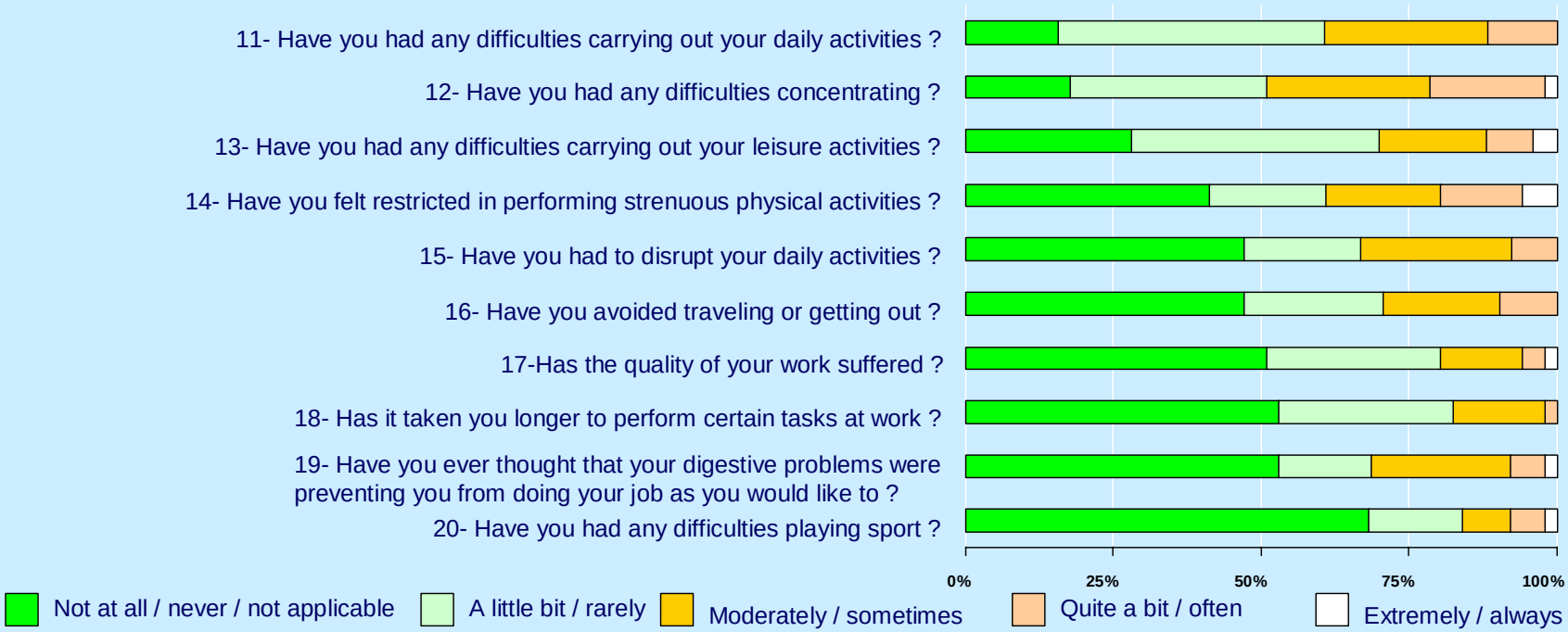
Item reduction (Keeping the most relevant items)

During the development of a new PRO questionnaire in irritable bowel syndrome, patient were asked to answer these items ranging from “not at all” to “extremely.” Results are presented as a percentage of patients. Study the data below and decide which items should be deleted. Give reasons and criticize the method.



ASSIGNMENT 4

Item reduction (Keeping the most relevant items)



Chassany O, Marquis P, Scherrer B, Read NW, Finger T, Bergmann JF, Fraitag B, Geneve J, Caulin C. Validation of a specific quality of life questionnaire in functional digestive disorders (FDDQL). Gut 1999; 44: 527-533.

- **Learning points**
  - Patient plays a key role in the questionnaire development
  - Quality of life experts, clinicians should take important decisions in the questionnaire development process
  - The initial steps determine the quality of the final questionnaire

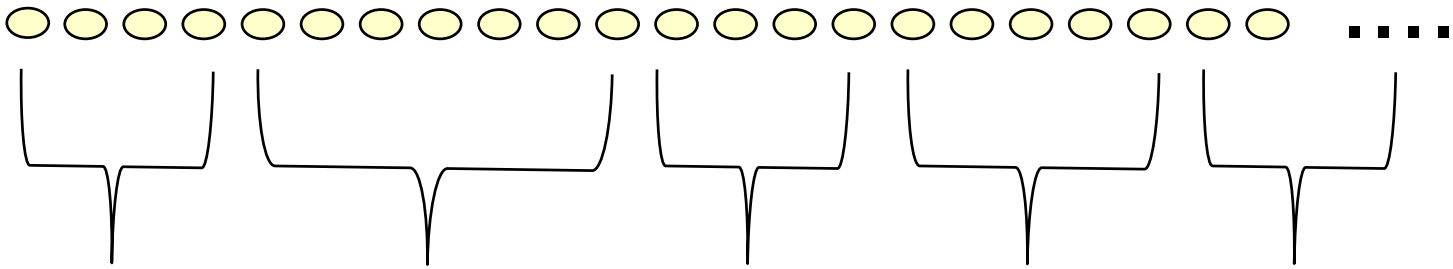
How is a new PRO questionnaire developed? Psychometric validation and cultural adaptation

Objectives

- To describe the evaluation of psychometric properties: reliability, validity, and responsiveness
- To describe the process of cultural adaptation

Questionnaire Structure

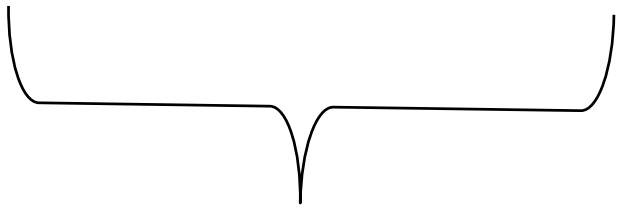
Items



Profile
(dimensions)

Activities **Emotions** **Sleep** **Relationships** **...**

Expectations **Performance** **Conformity** **Satisfaction**



Index

(Global score)

Psychometric Analysis

Acceptability	missing data descriptive statistics
Scaling assumption	item-item correlations item-scale correlations factorial/multitrait analysis
Reliability	reproducibility internal consistency (Cronbach's alpha)
Validity	content validity criterion/construct validity clinical validity
Sensitivity	discriminative power responsiveness over time

Psychometric Analysis

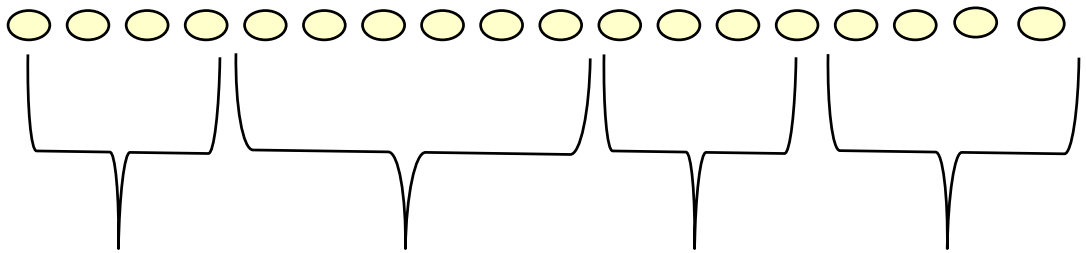
- **Acceptability**
  - missing data
  - descriptive statistics
- ***Scaling assumption***
  - item-item correlations
  - item-scale correlations
  - factorial/multitrait analysis
- **Reliability**
  - reproducibility
  - internal consistency (Cronbach's alpha)
- **Validity**
  - content validity
  - criterion/construct validity
  - clinical validity
- **Sensitivity**
  - discriminative power
  - responsiveness over time

Scaling Assumptions

- Hypotheses of item aggregation into dimensions
  - Index - profile (or both)
- Factor structure - Multitrait analysis - (Rasch model)
- Weighting
  - Implicit or explicit (population - experts - sample - individual)
- Units - Anchors - Direction of changes
- "zero" only has a relative meaning
  - Except for preference-based measures

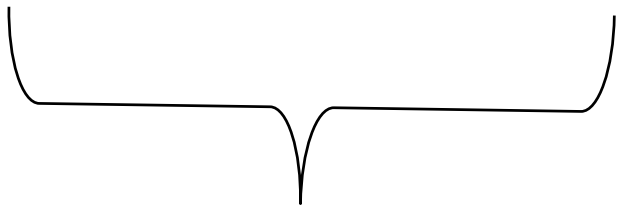
Scoring Method: Questionnaire Structure

Items



Profile
(dimensions)

Daily Activities **Emotions** **Sleep** **Relationships**



Index

(Global score)

Psychometric Analysis

- **Acceptability**
  - missing data
  - descriptive statistics
- **Scaling assumption**
  - item-item correlations
  - item-scale correlations
  - factorial/multitrait analysis
- ***Reliability***
  - reproducibility
  - internal consistency (Cronbach's alpha)
- **Validity**
  - content validity
  - criterion/construct validity
  - clinical validity
- **Sensitivity**
  - discriminative power
  - responsiveness over time

Reliability

- Measurement error
  - Ratio of true variance / variance due to instrument
- Inter-rater - Internal consistency - Test-retest
- Tests (reported most of the time)
  - ICC (Intra-class Correlation Coefficient) or Kappa coefficient
  - Cronbach alpha
  - Targeted coefficient >.70

Psychometric Analysis

- **Acceptability**
  - missing data
  - descriptive statistics
- **Scaling assumption**
  - item-item correlations
  - item-scale correlations
  - factorial/multitrait analysis
- **Reliability**
  - reproducibility
  - internal consistency (Cronbach's alpha)
- ***Validity***
  - content validity
  - criterion/construct validity
  - clinical validity
- **Sensitivity**
  - discriminative power
  - responsiveness over time

Validity

- To demonstrate scientific usefulness of a measure
  - Questionnaire measures what it is supposed to measure and does not measure what it is not supposed to measure
- Hypotheses of relationship with defined criteria
  - Clinical status most of the time (clinical validity)
  - Other instrument (concurrent validity)
  - Mainly correlations and group comparisons
- Association with an event that may occur in the future (predictive validity)
  - hospitalization, doctor's visit, ...

Psychometric Analysis

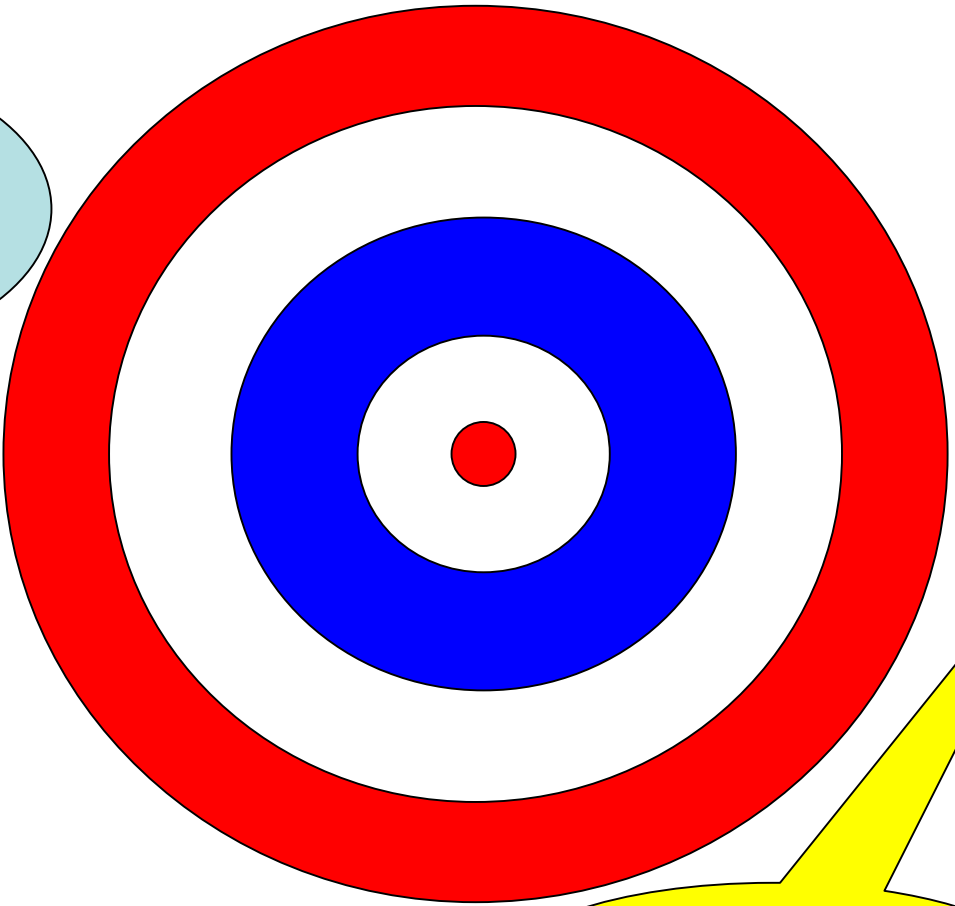
Acceptability	missing data descriptive statistics
Scaling assumption	item-item correlations item-scale correlations factorial/multitrait analysis
Reliability	reproducibility internal consistency (Cronbach's alpha)
Validity	content validity criterion/construct validity clinical validity
Sensitivity	discriminative power responsiveness over time

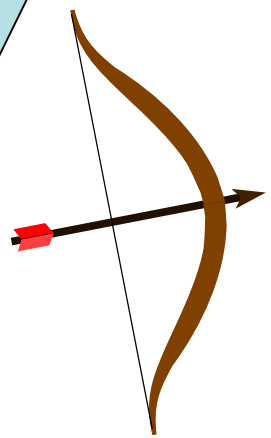
Responsiveness over time

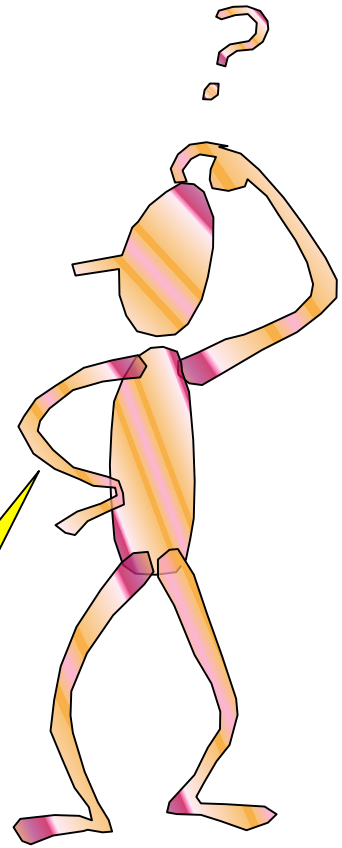
- Ability of an instrument to detect small but important changes (belongs to validity)
- Score changes related to a defined and measurable criteria of change
  - clinical status most of the time
  - physician or patient rating
- Tests (reported most of the time)
  - Paired statistics
  - Standardized ratio: Effect size - SRM - Guyatt Statistics

This is the target

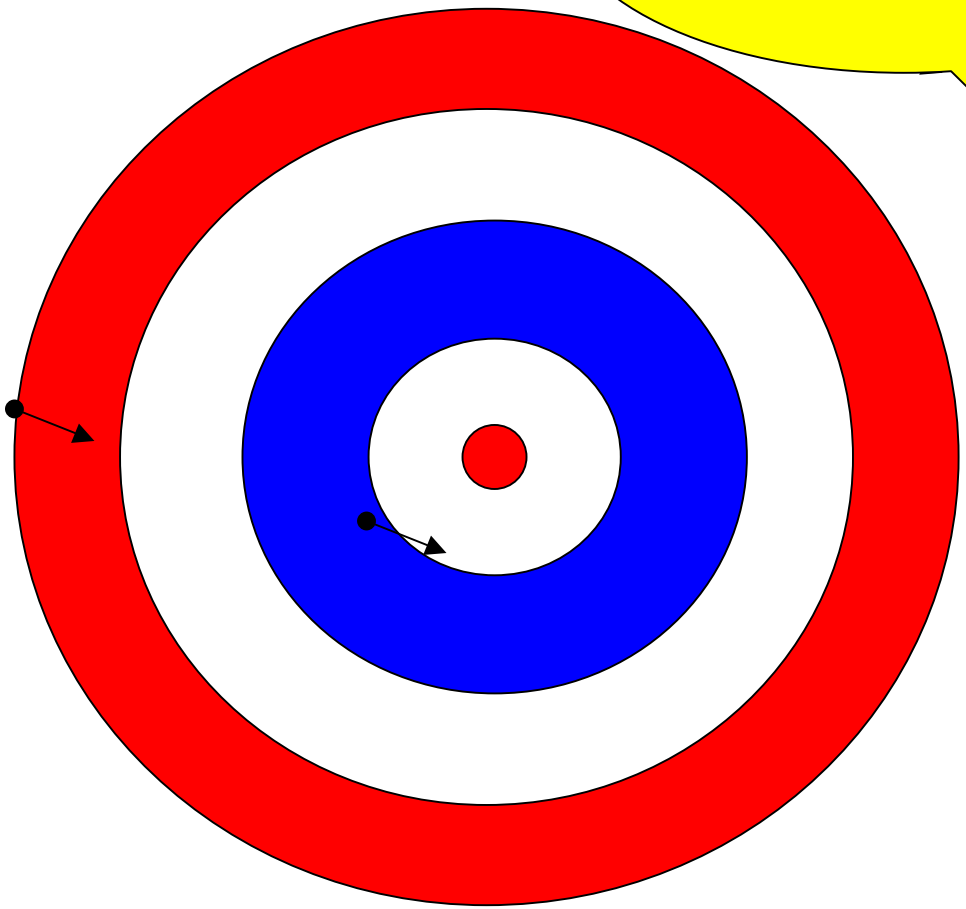
William Tell is standing there

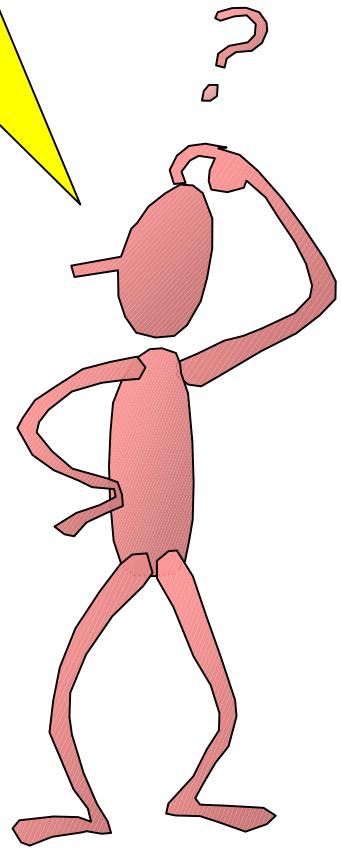






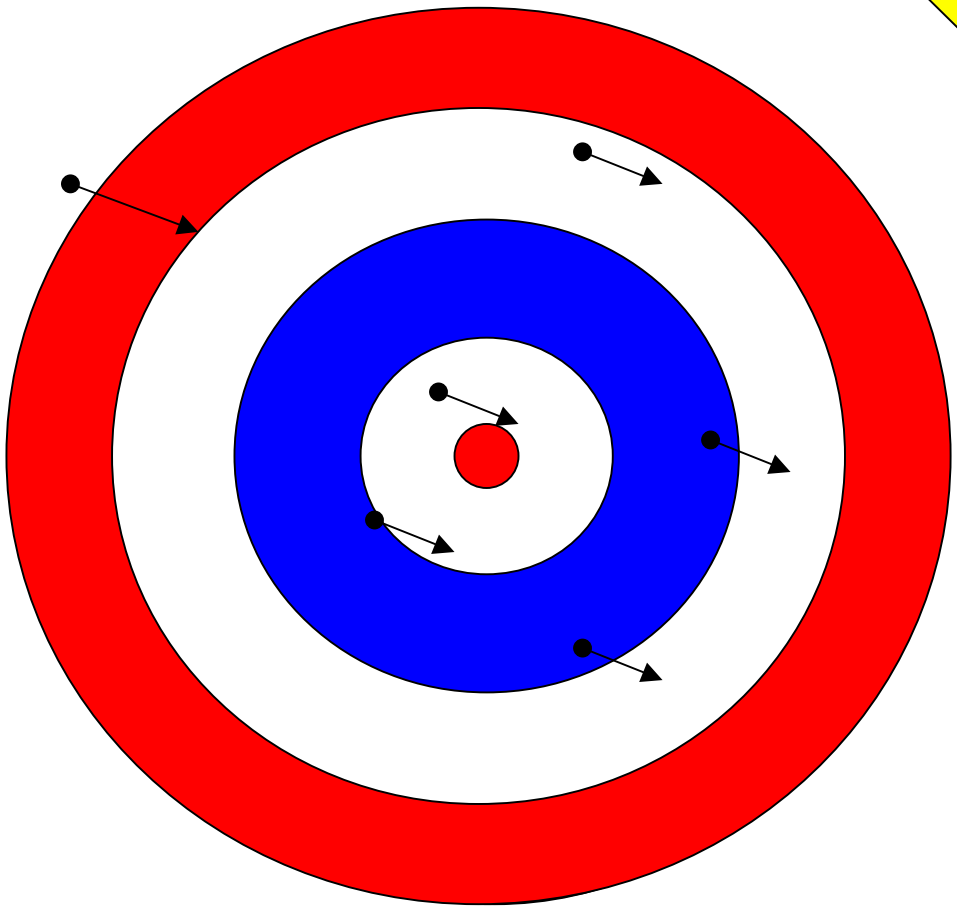
There you are

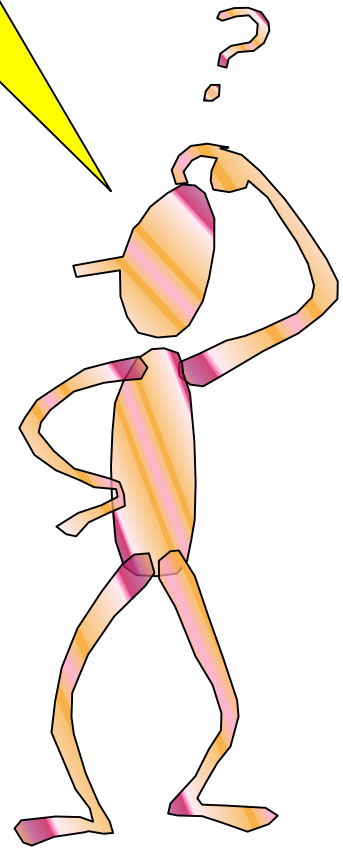




Neither valid nor reliable

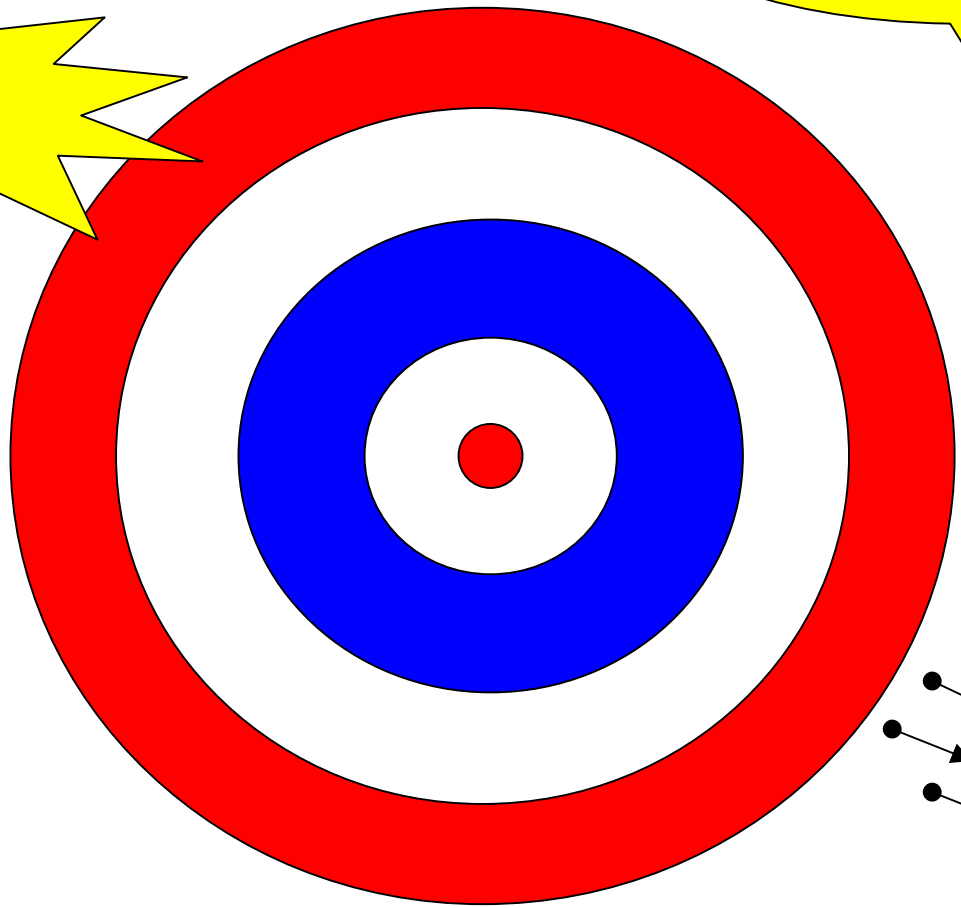
Valid but not very reliable

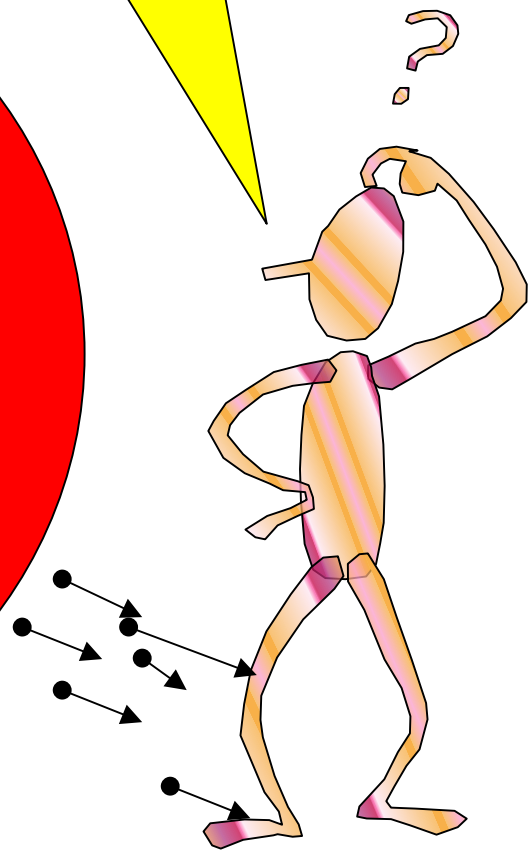




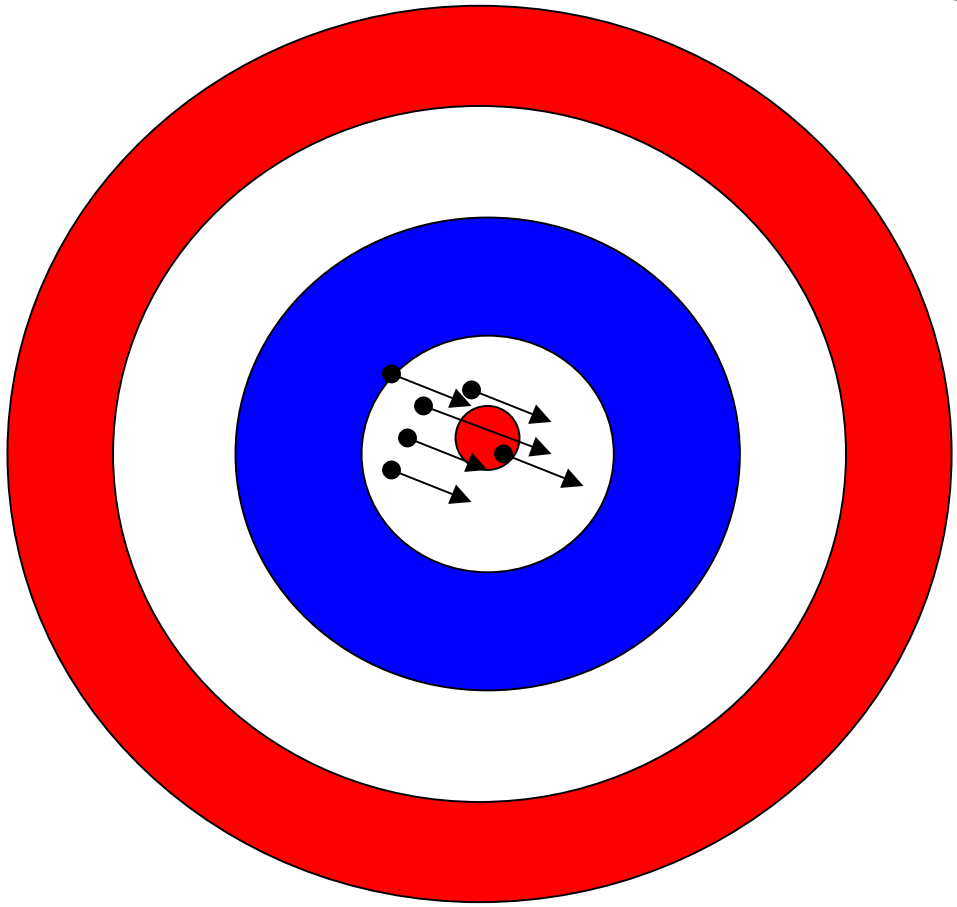
Reliable but not valid

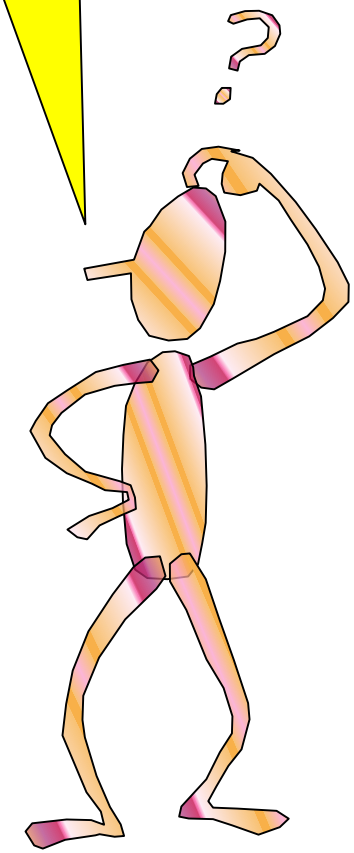
Sorry !



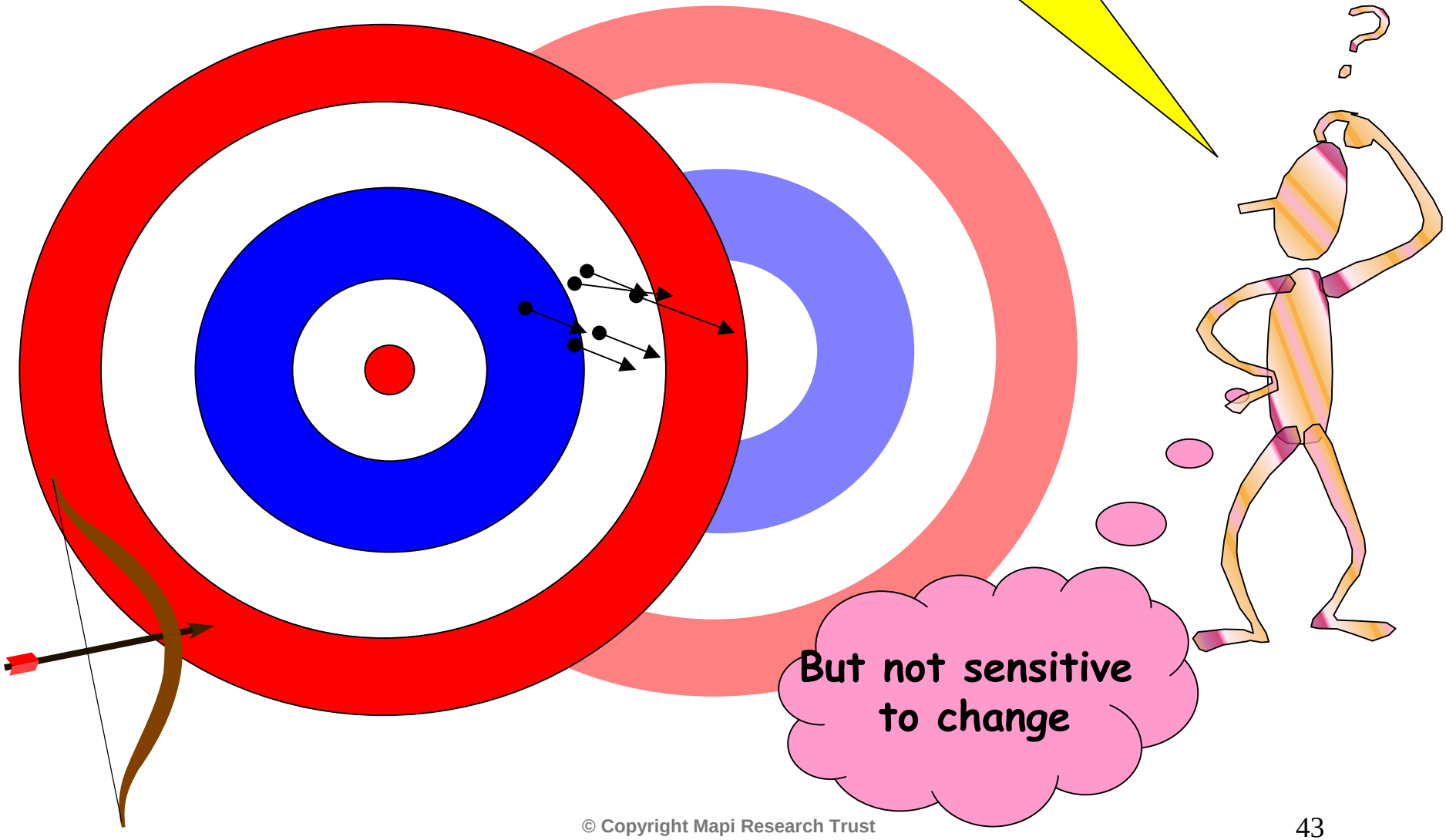


Valid and reliable





Valid and reliable...



THE COCHRANE COLLABORATION®
HRQL Methods Group

4- HOW IS A NEW PRO QUESTIONNAIRE DEVELOPED?
Psychometric validation and cultural adaptation

The ERIQA Group

BACKGROUND

The development of a measurement instrument is a long process, involving: item generation; item scaling; item aggregation into domains (e.g. HRQL); item reduction; and demonstration of acceptability, reliability, validity, and responsiveness.

This PROmat explores the second steps of the Patient-Reported Outcomes (PRO) questionnaire development, i.e.: psychometric validation and cultural adaptation.

ASSIGNMENTS

Assignment 1:
This assignment presents several different validation steps. Please comment each of them and explain your findings using table 3.1 provided in the workbook .

Assignment 2:
The cultural adaptation of a PRO measure is a rigorous and complex process (see the Workbook, Fig.4).

The main objective is to obtain a conceptual equivalence between the source and target versions allowing, among other things, "pooling" and comparison of data from international studies.

In this assignment, you will define the concepts behind several items, find appropriate equivalents for a same item in different languages and finally explore the relevance of patient testing.

ASSIGNMENT 1 — Psychometric validation

A) RELIABILITY : You will find below data from published papers. What information do you need to assess the reliability of these questionnaires?

Example 1: Irritable Bowel Syndrome-Quality of Life Instrument

IBS-QOL : 34 items	Internal consistency: Cronbach's α coefficient	Test-retest: Intraclass Correlation Coefficient (ICC)
Dysphoria (8)	0.92	0.89
Interference with activity (7)	0.84	0.88
Body image (4)	0.75	0.85
Health worry (3)	0.70	0.86
Food avoidance (3)	0.76	0.76
Social reaction (4)	0.74	0.84
Sexual (2)	0.83	0.77
Relationships (3)	0.65	0.69
Overall scale	0.95	0.86

Patrick DL, Drossman DA, Frederick IO, DiCesare J, Puder KL. Quality of life in persons with irritable bowel syndrome: development and validation of a new measure. Digestive Diseases and Sciences 1998; 43: 400-411.

Example 2: European Evaluation of Vertigo (EEV) – Symptoms scale

EEV: 5 items	Test-retest: Intraclass Correlation Coefficient (ICC)
Illusion of movement	0.93
Duration of illusion	0.58
Motion intolerance	0.90
Neurovegetative signs	0.97
Instability	0.87
Global score	0.93

Megnigbeto CA, Sauvage JP, Launois R. The European Evaluation of Vertigo (EEV) scale: a clinical validation study. Rev Laryngol Otol Rhinol (Bord). 2001; 122: 95-102

B) MEASUREMENT VALIDITY

1- Factorl analysis: How many subscales do you identify? Assemble the items from each subscale and give a title to each subscale

Items (QOLRAD)	I	II	III	IV	V
Discouraged or distressed	0.74	0.21	0.27	0.29	0.20
Generally unwell	0.52	0.32	0.31	0.26	0.49
Unable eat food one likes	0.20	0.26	0.78	0.27	0.13
Difficulty socializing	0.48	0.27	0.33	0.61	0.05
Food seems unappealing	0.35	0.19	0.63	0.24	0.27
Discomfort due to eating/drinking	0.26	0.20	0.71	-0.05	0.28
Feeling tired/worn out	0.37	0.31	0.23	0.25	0.65
Tired-lack of sleep	0.33	0.75	0.24	0.22	0.28
Wake up at night	0.21	0.84	0.18	0.21	0.08
Frustrated	0.69	0.26	0.38	0.24	0.20
Trouble getting to sleep	0.33	0.73	0.29	0.26	0.15
Lack of energy	0.43	0.34	0.26	0.36	0.54
Intolerance to food	0.31	0.24	0.73	0.17	0.05
Worry serious disease	0.75	0.22	0.21	0.23	0.09
Eat smaller meals	0.15	0.15	0.65	0.32	0.37
Irritable	0.64	0.29	0.33	0.25	0.25
Anxious or upset	0.74	0.25	0.28	0.25	0.22
Unable to carry out physical activities	0.27	0.24	0.22	0.78	0.19
Avoid bending over	0.16	0.31	0.22	0.42	0.50
No good night sleep	0.21	0.83	0.27	0.13	0.20
Kept from doing things with family/friends	0.29	0.24	0.32	0.68	0.33
Worries/Fears	0.77	0.26	0.19	0.16	0.14
Unable carry out daily activities	0.39	0.26	0.22	0.72	0.21
Not waking fresh/rested	0.40	0.61	0.24	0.30	0.25
Avoid certain food/drink	0.24	0.21	0.74	0.30	-0.04

Kulich K, Wiklund I, Junghard O. Factor structure of the Quality of Life in Reflux and Dyspepsia (QOLRAD) questionnaire evaluated in patients with heartburn predominant reflux disease. Quality of life research 2003; Vol 12: 699-708.

2- Convergent validity: Below are correlations between a self-report quality of life measure specific to urinary incontinence (I- QOL) and domains of the psychologic well-being (PGWB), and correlations between the I-QOL and functional status domains on the SF-36. Give your comments.

PGWB	Global score Correlation with I-QOL	MOS SF-36	Global score Correlation with I-QOL
		Role, physical	0.67
Health worry concern	0.59	Social function	0.62
Energy and vitality	0.55	Physical functioning	0.53
Depressed mood	0.52	Role, emotion	0.53
Tension and anxiety	0.51	Mental health	0.53
Positive well-being	0.48	General health	0.52
Behavioral and emotional control	0.45	Vitality	0.43
Total PGWB	0.62	Bodily pain	0.35

Wagner TH, Patrick DL, Bavendam TG, Martin ML, Buesching DP. Quality of life of persons with urinary incontinence: development of a new measure. Urology 1996;47:67-72

C) RESPONSIVENESS

- Within subject difference (6 wk minus baseline) in Salmeterol group divided by the pooled within-subject SD of change
- Randomized – Double Blind – Parallel Group – Multicenter - 6-wk
- Salmeterol 50 µg or Salbutamol 400 µg bid
- 120 patients randomized
- FEV_1 % predicted : 59%

	Effect size
AQLQ (32 items)	0.820
Activities (11)	0.860
Symptoms (12)	0.723
Environment (4)	0.550
Emotions (5)	0.302
	Effect size
LWAQ (68 items)	0.694
Health knowledge	0.625
Health appraisal	0.333
	Effect size
SIP (136 items)	0.320

Rutten -van Mölken MP, Custers F, van Doorslaer EK, Jansen CC, Heurman L, Maesen FP, Smeets JJ, Bommer Am, Raajmakers JA. Comparison of performance of four instruments in evaluating the effects of salmeterol on asthma quality of life. European Respiratory Journal 1995; 8: 888-98.

This is the comparison of three instruments used to evaluate the effects of Salmeterol on a patient's life. Give your comments.

AQLQ: Asthma Quality of Life Questionnaire

LWAQ: Living With Asthma Questionnaire

SIP: Sickness Impact Profile

ASSIGNMENT 2 — Cultural adaptation

A) Generic Questionnaire: MOS SF-36 Health Survey – Original US

Item: How much of the time during the past 4 weeks, have you felt downhearted and blue?

➔*The questionnaire had to be translated into UK-English and New Zealand*

1) What is the concept explored?

2) Do you think that the US version is acceptable in England and New Zealand? Give your reasons.

B) Disease: Asthma - Original version developed in Canada

Item: Here is a list of activities in which some people with asthma are limited, among them: « shoveling snow »

➔*The questionnaire had to be translated into Japanese*

1) What is the concept explored?

2) Was the literal translation chosen?

Give your suggestions for the Japanese version.

ASSIGNMENT 1

Psychometric validation

A) RELIABILITY : You will find below data from published papers.

What information do you need to assess the reliability of these questionnaires?

Example 1: Irritable Bowel Syndrome-Quality of Life Instrument

IBS-QOL : 34 items	Internal consistency: Cronbach's α coefficient	Test-retest: Intraclass Correlation Coefficient (ICC)
Dysphoria (8)	0.92	0.89
Interference with activity (7)	0.84	0.88
Body image (4)	0.75	0.85
Health worry (3)	0.70	0.86
Food avoidance (3)	0.76	0.76
Social reaction (4)	0.74	0.84
Sexual (2)	0.83	0.77
Relationships (3)	0.65	0.69
Overall scale	0.95	0.86

Patrick DL, Drossman DA, Frederick IO, DiCesare J, Puder KL. Quality of life in persons with irritable bowel syndrome: development and validation of a new measure. Digestive Diseases and Sciences 1998; 43: 400-411.

ASSIGNMENT 1

Psychometric validation

A) RELIABILITY

Example 2: European Evaluation of Vertigo (EEV) – Symptoms scale

EEV: 5 items	Test-retest: Intraclass Correlation Coefficient (ICC)
Illusion of movement	0.93
Duration of illusion	0.58
Motion intolerance	0.90
Neurovegetative signs	0.97
Instability	0.87
Global score	0.93

Megnigbeto CA, Sauvage JP, Launois R. The European Evaluation of Vertigo (EEV) scale: a clinical validation study. Rev Laryngol Otol Rhinol (Bord). 2001; 122: 95-102

ASSIGNMENT 1

Psychometric validation

B) MEASUREMENT VALIDITY

1- Factor analysis:

How many subscales do you identify? Assemble the items from each subscale and give a title to each subscale.

Items (QOLRAD)	I	II	III	IV	V
Discouraged or distressed	**0.74**	0.21	0.27	0.29	0.20
Frustrated	**0.69**	0.26	0.38	0.24	0.20
Anxious or upset	**0.74**	0.25	0.28	0.25	0.22
Worries/fears	**0.77**	0.26	0.19	0.16	0.14
Irritable	**0.64**	0.29	0.33	0.25	0.25
Worry serious disease	**0.75**	0.22	0.21	0.23	0.09
No good night sleep	0.21	**0.83**	0.27	0.13	0.20
Tired-lack of sleep	0.33	**0.75**	0.24	0.22	0.28
Wake up at night	0.21	**0.84**	0.18	0.21	0.08
Not waking fresh/rested	0.40	**0.61**	0.24	0.30	0.25
Trouble getting to sleep	0.33	**0.73**	0.29	0.26	0.15
Discomfort due to eating/drinking	0.26	0.20	**0.71**	-0.05	0.28
Eat smaller meals	0.15	0.15	**0.65**	0.32	0.37
Unable eat food one likes	0.20	0.26	**0.78**	0.27	0.13
Food seems unappealing	0.35	0.19	**0.63**	0.24	0.27
Intolerance to food	0.31	0.24	**0.73**	0.17	0.05
Avoid certain food/drink	0.24	0.21	**0.74**	0.30	-0.04
Avoid bending over	0.16	0.31	0.22	**0.42**	0.50
Kept from doing things with family/friends	0.29	0.24	0.32	**0.68**	0.33
Difficulty socializing	0.48	0.27	0.33	**0.61**	0.05
Unable carry out daily activities	0.39	0.26	0.22	**0.72**	0.21
Unable to carry out physical activities	0.27	0.24	0.22	**0.78**	0.19
Feeling tired/worn out	0.37	0.31	0.23	0.25	**0.65**
Generally unwell	0.52	0.32	0.31	0.26	**0.49**
Lack of energy	0.43	0.34	0.26	0.36	**0.54**

ASSIGNMENT 1

Psychometric validation

B) MEASUREMENT VALIDITY

2- Convergent validity: Below are correlations between a self-report quality of life measure specific to urinary incontinence (I- QOL) and domains of the psychologic well-being (PGWB), and correlations between the I-QOL and functional status domains on the SF-36. Give your comments.

PGWB	Global score Correlation with I-QOL	MOS SF-36	Global score Correlation with I-QOL
		Role, physical	0.67
Health worry concern	0.59	Social function	0.62
Energy and vitality	0.55	Physical functioning	0.53
Depressed mood	0.52	Role, emotion	0.53
Tension and anxiety	0.51	Mental health	0.53
Positive well-being	0.48	General health	0.52
Behavioral and emotional control	0.45	Vitality	0.43
Total PGWB	0.62	Bodily pain	0.35

Wagner TH, Patrick DL, Bavendam TG, Martin ML, Buesching DP. Quality of life of persons

ASSIGNMENT 1

Psychometric validation

C) RESPONSIVENESS

This is the comparison of three instruments used to evaluate the effects of Salmeterol on a patient's life. Give your comments.

- Within subject difference (6 wk minus baseline) in Salmeterol group divided by the pooled within-subject SD of change

- Randomized – Double Blind – Parallel Group – Multicenter - 6-wk
- Salmeterol 50 µg or Salbutamol 400 µg bid
- 120 patients randomized
- FEV_1 % predicted : 59%

	Effect size
AQLQ (32 items)	0.820
Activities (11)	0.860
Symptoms (12)	0.723
Environment (4)	0.550
Emotions (5)	0.302
	Effect size
LWAQ (68 items)	0.694
Health knowledge	0.625
Health appraisal	0.333
	Effect size
SIP (136 items)	0.320

AQLQ: Asthma Quality of Life Questionnaire

LWAQ: Living With Asthma Questionnaire

SIP: Sickness Impact Profile

Rutten -van Mölken MP, Custers F, van Doorslaer EK, Jansen CC, Heurman L, Maesen FP, Smeets JJ, Bommer Am, Raajmakers JA. Comparison of performance of four instruments in evaluaing the efects of salmeterol on asthma quality of life. European Respiratory Journal 1995; 8: 888-98.

ASSIGNMENT 2

Cultural adaptation

A) Generic Questionnaire: MOS SF-36 Health Survey – Original US

Item: How much of the time during the past 4 weeks, have you felt downhearted and blue?

➔ *The questionnaire had to be translated into UK-English and New Zealand*

1) What is the concept explored?

2) Do you think that the US version is acceptable in England and New Zealand? Give your reasons.

B) Disease: Asthma - Original version developed in Canada

Item: Here is a list of activities in which some people with asthma are limited, among them: « shoveling snow »

➔ *The questionnaire had to be translated into Japanese*

1) What is the concept explored?

2) Was the literal translation chosen?

Give your suggestions for the Japanese version.

Learning points

- Instrument development is a rigorous scientific process
- Psychometric validation should include evidence of reliability, validity, and responsiveness
- There is no single right way to develop an instrument although best practices are available for the steps in the process

Choosing an appropriate existing measure

Objectives

- To explore the process for selecting an appropriate PRO instrument for use in a specific clinical trial scenario
- To examine the trade-offs in the selection process
- To review the criteria necessary for appropriately evaluating a PRO instrument
- To identify and evaluate established questionnaires for use in a specific patient group

Choice of a PRO Measure

Relevance to the Pathology

- Literature: medical and specific data bank
- Patient interview
- Expert interviews

Highlighting Differences Between the Treatments

- Product hypotheses
- Comparator

Specific Matter

- Design of the study
- Targeted patients

Instrument Selection: Criteria to examine

- Patient-based (but unbiased) data collection
- Psychometric standards and goals *(reliability, validity, responsiveness to change)*
- Practical considerations *(wording, formatting, administrative/respondent burden)*

The ERIQA Group

THE COCHRANE COLLABORATION®
HRQL Methods Group

5- CHOOSING AN APPROPRIATE EXISTING MEASURE

BACKGROUND

The development of a new instrument requires an investment of both time and resources. Most frequently, disease-specific instruments are developed to measure concepts not covered by existing instruments. Decisions about whether to invest time and resources into the development of a new instrument depend on the importance of the outcome in the drug development process, the availability of existing instruments to assess the relevant outcomes for the drug and disease indication, and the commitment of the developer. The development of a new Patient-Reported Outcomes (PRO) instrument requires about two years of effort and several studies to pilot test the instrument and to evaluate its psychometric characteristics.

The effect of treatment or intervention on a patient's perspective is a balance between the safety, efficacy, and convenience of the treatment in question vs. the impact of the disease itself. The first step in developing a measurement strategy is to develop a hypothesis for the domains which may benefit from treatment. In order to avoid biasing the instrument in favor of a preferred hypothesis, a specific clinical research scenario must be used.

These assignments will provide you with criteria to help you evaluate whether an appropriate patient's perspective measure already exists for a particular clinical trial.

ASSIGNMENTS

Assignment 1:

The relevance of PRO as an outcome of treatment intervention depends not only on the impact of the disease and population but also on the clinical and psychological impact of the treatment in question including any side effects.

In this assignment, you will consider a clinical research scenario and try to find a net balance of the effects on a patient's life between improved clinical parameters and adverse effects of treatment. Study the research scenario for osteoarthritis and answer the questions to help you define a suitable measurement approach.

Assignment 2:

Data on the attributes of some existing scales for osteoarthritis are provided in the table.

a) Start by taking the list of PRO domains from Assignment 1 that you identified as being the important outcome measures for the clinical research scenario. Indicate the extent to which each of these instruments covers these domains by marking the relevant properties.

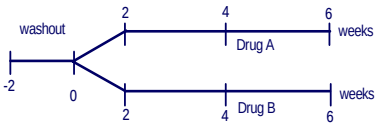
b) Review the "10 Questions to ask yourself" and tables 3.1 & 3.2 Provided in the Workbook . Now describe which of the instruments you would choose for this research scenario. Would you add any additional instruments or modules?

c) Give the reasons for your choice(s).

ASSIGNMENT 1

Developing hypotheses

Read the following research scenario and discuss the main clinical effects of treatment. Write in the table what you consider to be the most important PRO and any additional ones that might be affected by the treatment.

Clinical Research Scenario	a. List up to 5 of the most important PRO.	b. Describe how each of these might be affected by the treatment.
Comparative trial of anti-inflammatory treatment in patients with mild to moderate osteoarthritis. Background: Osteoarthritis, generally described as joint pain, is a common disease in the elderly population. The main symptom is pain, which may arise from the affected joint, surrounding bone, or inflammation. The pain is characteristically worse at the end of the day and accompanied with stiffness after inactivity and limitation of joint movement. In the absence of specific agents that might prevent or reverse osteoarthritis, treatment is symptomatic and directed primarily towards pain control and increased mobility. Purpose: to evaluate the effects of two different forms of non-steroidal anti-inflammatory drugs (NSAIDs) in patients suffering from mild to moderate osteoarthritis. In order to aid the decision and policy making on the appropriate use of resources in treating osteoarthritis, it is important that the results can be interpreted in the context of a wider population. Setting: This multinational clinical trial has been designed to compare the safety and efficacy of two NSAIDs in patients between 45 and 65 years of age. Although Drug A has been shown to be effective in treating osteoarthritis, it has a high incidence of gastrointestinal side effects. Drug B is a new enteric coated tablet, which should reduce the incidence and severity of these side effects, and it is expected to have at a minimum, equal efficacy. Three hundred patients will be recruited into the study from primary care centers in the USA, Canada, the UK, and Scandinavia. Design: After a two-week washout period, patients will be randomly assigned to receive either Drug A or B for a six-week treatment period. The primary endpoints are pain and mobility which will be assessed at weekly intervals. The measure of the impact of the treatment on a patient's life is considered to be a secondary endpoint.	c. Which PRO subscales do you expect to respond (i.e. what are your hypothesis)?	
	d. The scheme below shows the study design. Mark the points in the study in relation to the primary endpoints where PRO should be measured with an arrow.  	

10 Questions to ask yourself prior to searching for a PRO instrument

Questions:	Potential Responses:
1) What is the purpose of your study?	Clinical trial; epidemiological; needs assessment; caregiver burden
2) What is the population to be studied?	Disease, conditions, ethnic or cultural group, gender, age, socioeconomic status
3) What are the scales most relevant to that population?	Generic, disease-specific, condition-specific, adult vs. child
4) Who is the appropriate respondent?	Adult patient; adult proxy caregiver; child; parent of child; primary caregiver of child other than parent; inpatient/outpatient
5) What is the optimal mode of administration?	Interview (face to face/phone); self-report (home (mail-out/mail-back); clinic, trial laboratory); computer-assisted
6) What is the background of the majority of respondents?	Reading/comprehension level, language, gender, age, culture, education
7) What is the appropriate instrument length?	Must be sufficiently long to be scientifically robust, yet short enough to be practical in the "real world"; smaller samples need a greater number of items per scale; larger samples need a lesser amount of items per scale
8) How will you define a meaningful difference? What is a meaningful difference?	Statistically significant vs. clinically relevant vs. socially meaningful
9) Which subscales do you expect to respond?	All hypotheses should be pre-specified in the protocol. Answer will be dependent on the population being studied and will influence the type of measurement tool to use
10) What is the extent of time between administrations?	Answer will affect recall periods and expectations for meaningful differences

ASSIGNMENT 2

Designing an instrument using existing PRO measures; research scenario for osteoarthritis

a) For the questionnaires below, and referring to the table on the right and the document "Criteria for evaluating PRO instrument", check those properties that match the hypotheses and description of the trial in Assignment 1.

Instruments	Arthritis Impact Measurement Scales (AIMS 2)	Check	WOMAC Osteoarthritis Index (WOMAC)	Check	Health Assessment Questionnaire (HAQ)	Check	SF-36	Check	Psychological General Well Being Index (PGWBI)	Check
Number of items	78		24		20		36		22	
Purpose	To assess health outcomes among persons 18 years and older with arthritis		To assess health outcomes in clinical trials of patients 55 years and older with osteoarthritis of the hip or knee		To assess disability and pain associated with arthritis		To express health outcomes from a wide range of interventions on a common scale in persons aged 18 years and older		To assess self-reported well being and distress in the community from ages 25–74 years	
Concepts measured	Mobility level (5) Walking and bending (5) Hand & finger function (5) Arm function (5) Self-care tasks (4) Household tasks (4) Social activity (5) Support from family& friends (4) Arthritis pain (5) Work (5) Level of tension (5) Mood (5) Overall concepts (21)		Pain (5) Physical function (17) Stiffness (2)		Dressing (2) Eating (2) Walking discomfort (2) Hygiene (3) Reach (2) Arising (2) Grip (3) Activities (4)		Mental Health (5) General health perceptions (6) Pain (2) Physical functioning (10) Role limitations (7) Social functioning (2) Vitality (4)		Positive well-being (4) Depression (3) Anxiety(5) Vitality (4) Health perceptions (3) Self-control (3)	
Development language	English (US)		English (US)		English (US)		English (US)		English (US)	
Applications	Clinical use (juvenile version available)		Clinical and research use		Clinical and research use		Clinical and research use		Clinical and research use	
Language adaptations	Chinese, Dutch, French, Greek, Hebrew, Italian, Japanese, Norwegian, Portuguese for Brazil, Russian, Spanish, Swedish		Afrikaans, Arabic for Egypt, Bulgarian, Croatian, Czech, Danish, Dutch, English (for Australia Canada, UK, US), Estonian, Finnish, French, Greek, Hebrew, Italian, Japanese, Mandarin for China, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish, Turkish + 39 other languages		Afrikaans for South Africa, Arabic, Chinese, Croatian, Czech, Danish, Dutch, English (UK), Finnish, French, French Canadian, German, Greek, Hebrew, Hungarian, Italian, Norwegian, Portuguese, Russian, Spanish, Swedish, Turkish + 19 other languages.		Afrikaans, Chinese for Taiwan, Danish, Dutch, English(for Australia / New Zealand, UK), Finnish, French(for france, Canada), German, Hebrew, Hungarian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Slovenian, Spanish , Swedish, Turkish + 38 other languages		Afrikaans, Catalan, Czech, Danish, Dutch, English (for Australia, Canada, Ireland, UK), Finnish, French, German, Greek, Hebrew, Hungarian, Italian, Japanese, Korean, Latvian, Norwegian, Polish, Portuguese, Slovak, Slovenian, Spanish, Swedish + 23 other languages.	
Reliability	Internal consistency high for all measures		Good test–retest correlation (apart from stiffness scale)		High test–retest correlation and internal consistency		Assessed in more than 10,000 patients and 2,000 non-patients		Good internal consistency and test–retest reliability in a variety of general and mental health populations	
Validity	Good correlation with traditional rheumatology measures		Some correlation with other scales		Good correlation between scores and observer ratings		Construct validity established by comparison of patients and healthy populations		Validity assessed using standard psychological tests (e.g., MMPI for depression)	
Responsiveness	Able to detect changes		Unknown		Able to detect changes but not superior to other instruments (e.g., AIMS).		Able to detect changes		Unknown	
Interpretability	Arthritis data available, used as "gold standard" in arthritis		Good correlation with clinical indices of arthritis severity but no large surveys documenting norms or benchmarks		Norms and arthritis benchmarks available for comparison		Norms and benchmark data available for comparison		US norms are available	
Scoring	Scores for individual concepts can be summed up to provide an overall score		Concept scores obtained by summing up all scores within a scale; no overall score		4-point scale to measure ability to perform task; overall scores for physical disability and physical discomfort; no system to value tradeoffs		Single score per concept after appropriate transformation of response (0-100); profile option; physical and mental summary score option		Total score obtained from summing up individual item responses	
Burden	Self-administered ; 20 minutes		Self- or interviewer-administered		Interviewer- or self-or telephone administered; 5–8 minutes		Self-or interviewer- or telephone-administered; computerized administration is also available; 5–10 minutes		Self-administered, self response only; 8–15 minutes	
Recall periods	Past month		48 hours		Past week		Time reference 4 weeks (acute version available)		Time reference is last 4 weeks	

b) What would be your preferred instrument for measuring PRO?

_____

_____

_____

_____

_____

c) Give the reasons for your choice

_____

_____

_____

_____

_____

ASSIGNMENT 1

Developing hypotheses

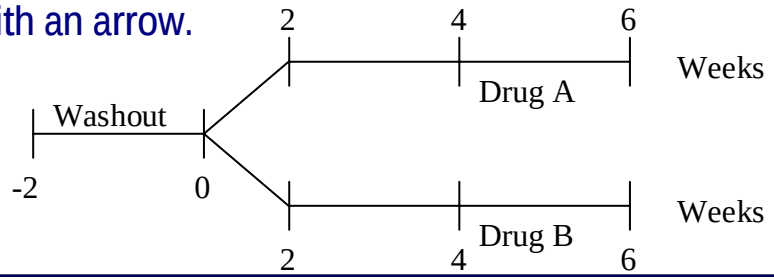
Read the following research scenario and discuss the main clinical effects of treatment. Write in the table what you consider to be the most important PRO and any additional ones that might be affected by the treatment.

<table>
<tr><th>Clinical Research Scenario</th><th>a. List up to 5 of the most important PRO.</th><th>b. Describe how each of these might be affected by the treatment.</th></tr>
<tr><td rowspan="3">Comparative trial of anti-inflammatory treatment in patients with mild to moderate osteoarthritis.

Background: Osteoarthritis, generally described as joint pain, is a common disease in the elderly population. The main symptom is pain, which may arise from the affected joint, surrounding bone, or inflammation. The pain is characteristically worse at the end of the day and accompanied with stiffness after inactivity and limitation of joint movement. In the absence of specific agents that might prevent or reverse osteoarthritis, treatment is symptomatic and directed primarily towards pain control and increased mobility.

Purpose: to evaluate the effects of two different forms of non-steroidal anti-inflammatory drugs (NSAIDs) in patients suffering from mild to moderate osteoarthritis. In order to aid the decision and policy making on the appropriate use of resources in treating osteoarthritis, it is important that the results can be interpreted in the context of a wider population.

Setting: This multinational clinical trial has been designed to compare the safety and efficacy of two NSAIDs in patients between 45 and 65 years of age. Although Drug A has been shown to be effective in treating osteoarthritis, it has a high incidence of gastrointestinal side effects. Drug B is a new enteric coated tablet, which should reduce the incidence and severity of these side effects, and it is expected to have at a minimum, equal efficacy. Three hundred patients will be recruited into the study from primary care centers in the USA, Canada, the UK, and Scandinavia.

Design: After a two-week washout period, patients will be randomly assigned to receive either Drug A or B for a six-week treatment period. The primary endpoints are pain and mobility which will be assessed at weekly intervals. The measure of the impact of the treatment on a patient's life is considered to be a secondary endpoint.</td><td colspan="2"></td></tr>
<tr><td colspan="2">c. Which PRO subscales do you expect to respond (i.e. what are your hypothesis)?</td></tr>
<tr><td colspan="2">d. The scheme below shows the study design. Mark the points in the study in relation to the primary endpoints where PRO should be measured with an arrow.
 
</td></tr>
</table>

ASSIGNMENT 2

Designing a strategy using existing PRO measures; research scenario for osteoarthritis

a) For the questionnaires below, and referring to the table on the right and the document “Criteria for evaluating PRO instrument”, check those properties that match the hypotheses and description of the trial in Assignment 1.

b) What would be your preferred instrument for measuring PRO?

_____

_____

_____

c) Give the reasons for your choice

_____

_____

_____

Learning points

- The first step is to ask yourself Key Questions

- The selection of a PRO instrument or the choice of domains (in the case of HRQL) is influenced by many factors, among them the severity and nature of the disease, the expected benefits and side effects of treatment, the evidence of its psychometric properties, its availibility in the targeted languages

Lecture 2