PRO Regulatory Issues in Europe

EMEA reflection paper on HRQL
Analysis and interpretation

Olivier CHASSANY, MD, PhD
Clinical Research & Development Department *(institutional sponsor)*
Assistance Publique – Hôpitaux de Paris, Medical University Paris 7, France
olivier.chassany@sls.aphp.fr

FDA PRO guidance for industry versus EMEA HRQL reflection paper

FDA

Guidance for Industry

Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

DRAFT GUIDANCE

February 2006
www.fda.gov/cder/guidance/5460dft.pdf

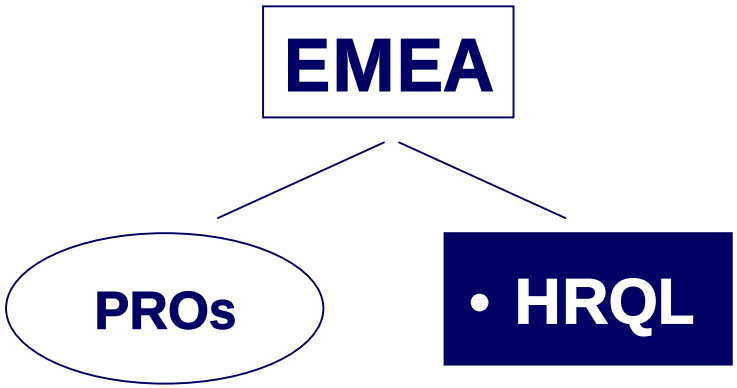
EMEA

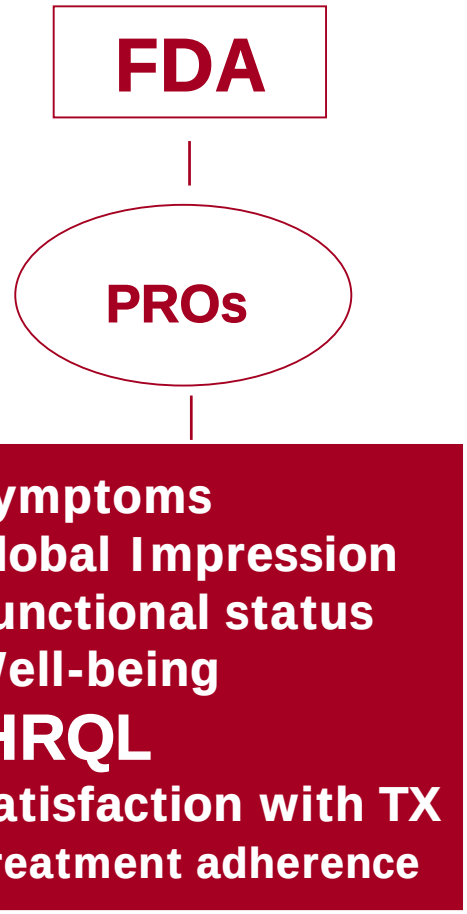
Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products

Scope : to discuss the place that HRQL, a specific type of PRO, may have in drug evaluation process and to give some broad recommendations on its use in the **context of already existing guidance documents**.

EMEA/CHMP/EWP/139391/2004
Adoption by CHMP : July 2005
Came into effect : January 2006
www.emea.eu.int

Discrepancy between FDA and EMEA guidance ?





EMEA reflection paper
(Introduction)

HRQL should be clearly differentiated from the core symptoms of a disease (e.g. pain, migraine, pyrosis...) assessed by the patient himself which are well-accepted primary and secondary efficacy endpoints in registration trials.

First step of recognition of HRQL by EMEA and national Agencies through the EWP (Efficacy Working Party), and the need to go beyond symptoms and functional status

A few issues raised by the Efficacy Working Party (EMEA) - Dr Mira Pavlovic (AFSSAPS representative)

- **How can we better understand the relationship between HRQL and Patient-Reported Outcomes in general ?**
- Should a validated HRQL scale go through all the steps of validation when translated in several different languages in order to be included in a drug development program ?
- Why developing a multi-domain HRQL scale, if at the end the sponsor predefines in a clinical trial one or two domains of this scale in order to gain a specific claim on these domains ?
- Should the sponsor develop generic or disease specific questionnaires ?
- How is it possible for a reviewer to check that the claim asked by the sponsor corresponds to what has been really measured in clinical trials ?
- What are the important issues to check when reviewing a dossier with HRQL data ?

Workshop on validation of scales (October 2005, EWP, EMEA, London)
Development and validation of “multidomain scales” (example 3, in quality of life)

Why Use Patient-Reported Outcome Instruments in Medical Product Development?

FDA Guidance

PRO instruments are included in clinical trials for new medical products because :

1 Some Treatment Effects Are Known Only to the Patient

For example, **pain** intensity and pain relief are the fundamental measures used in the development of analgesic products. There are no observable or physical measures for these concepts.

2 Patients Provide a Unique Perspective on Treatment Effectiveness

... improvements in clinical measures of a condition may not necessarily correspond to improvements in how the patient functions or feels.

3 Formal Assessment May Be More Reliable Than Informal Interview

Self-completed questionnaires capture directly the patient's perceived response to treatment, without a third party's interpretation...

FDA Guidance, P3-4, L103-120

Weak correlation between Patient-Reported Outcomes and physiological endpoints

(n = 96)	*r*	*BPQ*	*CRQ*
6-min walk test		0.17	0.07
Pre SaO2		0.14	0.17

Symptoms BPQ : Breathing Problems Questionnaire
HRQL CRQ : Chronic Respiratory Disease Questionnaire

→ Variability in exercise capacity contributed to only 3% of the variability in BPQ score

Quality of life in elderly patients with COPD: measurement and predictive factors. Yohannes AM et al. Resp Med 1998.

Increasing cardiorespiratory fitness (CRF) does not systematically improve fatigue and quality of life

- 111 patients allocated to an intervention group (n = 59), which underwent a 14-week training program, or a control group (n = 52)
- Chemotherapy for lymphomas, breast, gynecologic, testicular cancer
- Primary outcome : change in CRF (Bicycle ergometer test between)
- Secondary outcomes :
  - Hospital Anxiety and Depression Scale
  - EORTC QLQ-C30

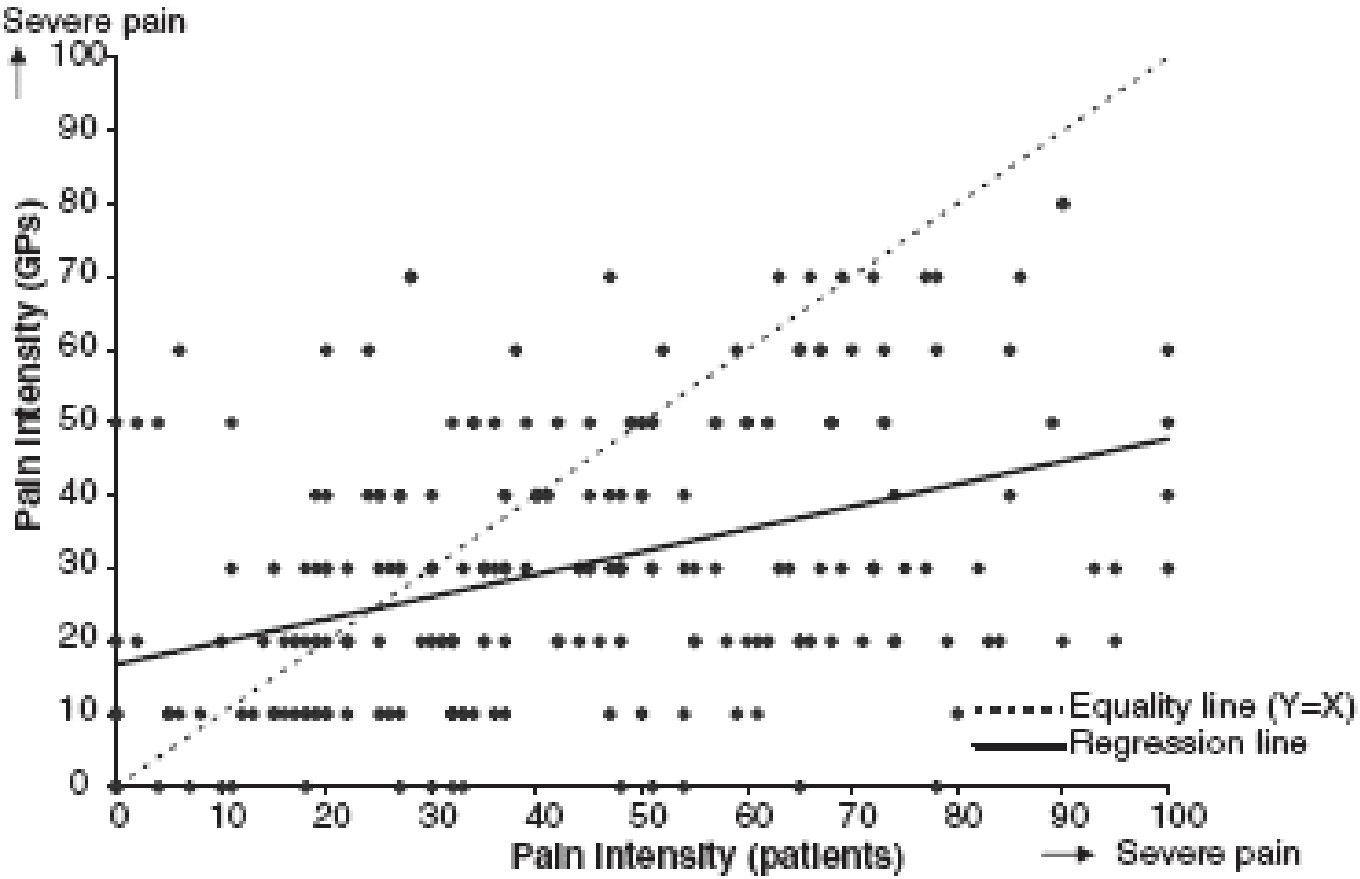
RESULTS

- CRF ($p<0.01$) : 6.4 mL/kg/min (**intervention**) vs 3.1 (control)
- Fatigue score ($p<0.01$) : -17.0 points (**control**) vs - 5.8 (intervention)
- No differences in mental distress or HRQOL.

Thorsen L. Effectiveness of physical activity on cardiorespiratory fitness (CRF) and health-related quality of life in young and middle-aged cancer patients shortly after chemotherapy. J Clin Oncol 2005

Clinicians underestimate pain severity
(Irritable Bowel Syndrome)

Patients	**39.0 ± 24.9** (n=232)
GPs	**30.4 ± 21.0** (n=307)
Difference	**8.6 (28%)**
Correlation	**Kw = 0.31** (n=232)



Pain values ranged between 0 (no pain) and 100 (severe pain)
All values under the equality line indicate GP underestimation of pain by GPs

Chassany O, Duracinsky M, et al. Discrepancies between PROs and Clinician-Reported Outcomes in chronic venous disease, irritable bowel syndrome, and peripheral arterial occlusive disease. Value in Health 2006.

Clinicians estimate mistakenly the impact on QoL

Chronic Venous Disease (CVD)	Patients (CIVIQ) (n = 240)	GPs (4 items) (n = 291)	Relationship (n = 240)
Global score	61 ± 20	72 ± 19	Kw = 0.17

Peripheral Arteriopathy Obstructive Disease	Patients (CLAU-S) (n = 68)	GPs (5 items) (n = 90)	Relationship (n = 58)
Global score	66 ± 23	54 ± 21	Kw = 0.26

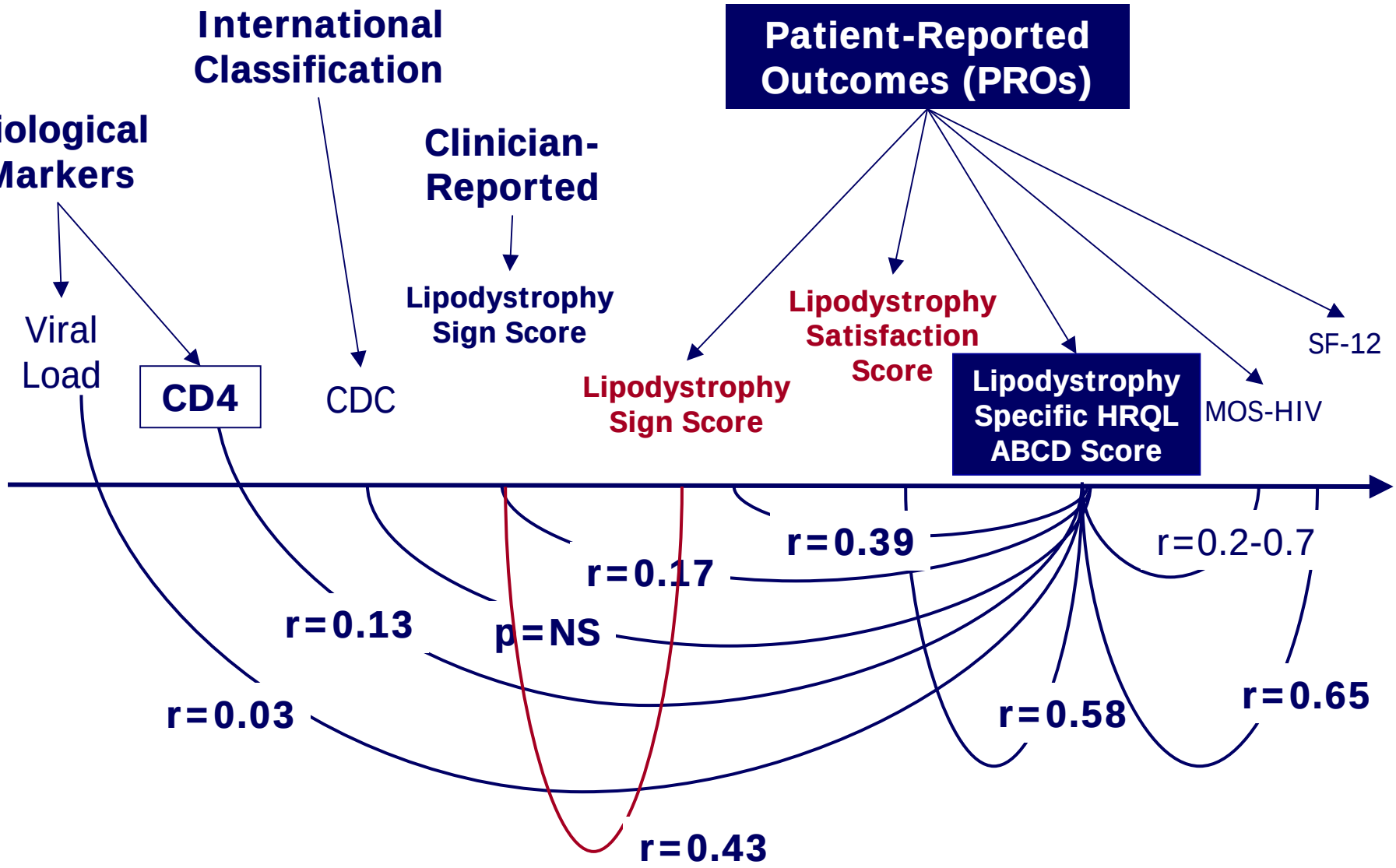
QoL scores ranged from 0 (bad QoL) to 100 (good QoL)

Kw : Weighted Kappa coefficient

GPs tend to underestimate the impact of “non severe” disease (CVD) on QoL and overestimate the impact of « more severe » disease (arteriopathy)

Chassany O, Duracinsky M, et al. Discrepancies between PROs and Clinician-Reported Outcomes in chronic venous disease, irritable bowel syndrome, and peripheral arterial occlusive disease. Value in Health 2006.

The specific impact of Lipodystrophy (HIV) on HRQL is not adequately captured by other criteria



Duracinsky M, Chassany O. Agreement between patients' and clinicians'-reported outcomes in lipodystrophy (HIV/AIDS). Value in Health 2004.

Psoriasis : Note for Guidance

CPMP/EWP/2454/02 (Nov. 2003)

HRQL

4.1.2. Patient's assessed outcome measures (cont'd)

Efficacy of a new drug **evaluated by patient** is important when ... **even relatively limited extent of skin psoriasis may severely socially and psychologically disable the patient.**

The assessment of **HRQL scales specific for psoriasis may represent an added value** for a new drug in comparative clinical trials, in addition to classical efficacy/safety measures. Patient-assessed drug efficacy may be a secondary or tertiary endpoint in pivotal clinical trial.

... Ideally, trials assessing psoriasis-specific HRQL should be designed to assess patient's perspective in the evaluation of drug-effect **in order to understand better the clinical significance** of the benefit observed and to be sure that the administered treatment does not impact adversely on patient's HRQL.

II. Recommended primary/secondary efficacy endpoints

a) Symptom modifying drugs

PRO

- **Pain** attributable to the target joint is recommended as primary endpoint. Functional disability is an important additional primary endpoint.

  Pain should be measured by self-assessment with validated methods, such as VAS or Likert scale.

PRO

- Functional disability

  A disease-specific and joint specific instrument such as the **WOMAC**...[...]...is recommended.

Secondary endpoints include:

HRQL

Global rating, Flares, Physical signs including range of motion, **Quality of Life**, Consumption of medications for pain relief

5. Recommended primary/secondary efficacy endpoints

PRO

Primary: The patient's global assessment of symptoms and abdominal discomfort/pain should be used as the two primary endpoints. Statistically significant changes must be found in both parameters.

HRQL

Secondary (supportive): choice of secondary efficacy variables should be justified by the applicant and should include variables such as bloating/distension, stool frequency and urgency, and quality of life parameters. **Health-related quality of life must, however, be considered most important secondary endpoints.**

Rheumatoid arthritis

CPMP/EWP/556/95 rev 1 (Dec 2003)

3. Tools to measure efficacy (primary or secondary endpoints)

d) Patient's global assessment of disease activity (VAS)

e) Pain score (patient's assessment : VAS, Likert Scale)

PRO

g) Physical function (assessed by patient, e.g. **HAQ**, **AIMS**)

4. Supportive evidence for efficacy

d) Emotional and social function (e.g. AIMS-1)

HRQL

e) Quality of life (RA-specific, e.g. AIMS, SF-36 or generic...)

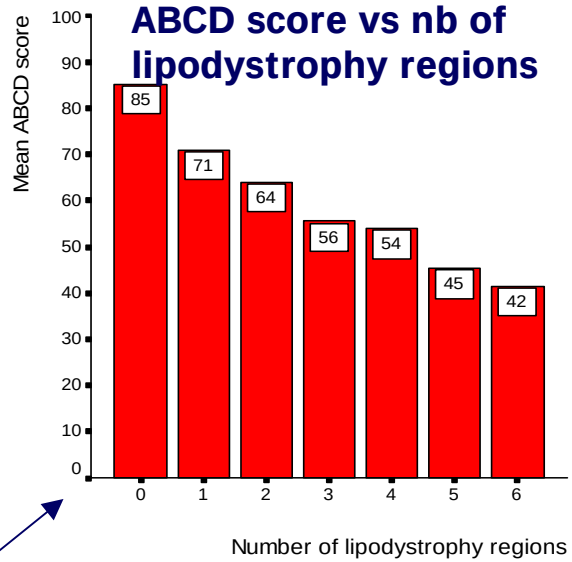
A few issues raised by the Efficacy Working Party (EMEA) - Dr Mira Pavlovic (AFSSAPS representative)

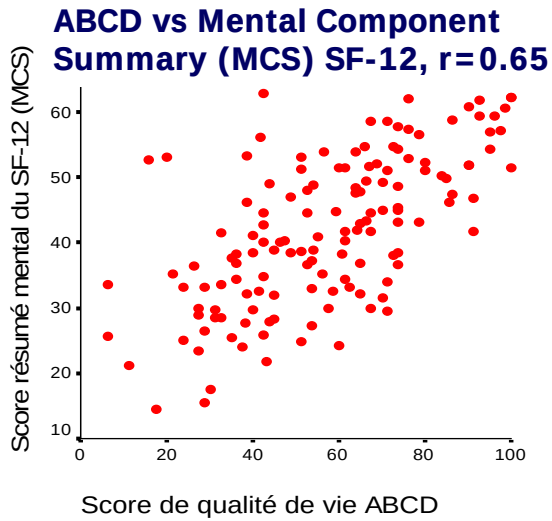
- How can we better understand the relationship between HRQL and Patient-Reported Outcomes in general ?
- **Should a validated HRQL scale go through all the steps of validation when translated in several different languages in order to be included in a drug development program ?**
- Why developing a multi-domain HRQL scale, if at the end the sponsor predefines in a clinical trial one or two domains of this scale in order to gain a specific claim on these domains ?
- Should the sponsor develop generic or disease specific questionnaires ?
- How is it possible for a reviewer to check that the claim asked by the sponsor corresponds to what has been really measured in clinical trials ?
- What are the important issues to check when reviewing a dossier with HRQL data ?

Workshop on validation of scales (October 2005, EWP, EMEA, London)
Development and validation of “multidomain scales” (example 3, in quality of life)

Rigorous process of development / validation of HRQL / PRO scales

- Conceptual framework
- Item generation
- Scaling, scoring
- Item reduction
- Reproductibility
- Content validity
- Construct validity
- Discriminant validity
- Convergent validity
- Responsiveness
- Cultural adaptation





Factorial analysis ABCD Score

ABCD 20 items	Factor 1	2	3	4
a	**,723**	,084	,284	,177
b	**,529**	,067	,427	,293
c	**,696**	,359	,152	,290
d	**,580**	**,488**	,149	,318
e	**,625**	,143	,471	,096
f	**,684**	,118	,347	-,105
g	**,609**	,195	,381	,125
h	**,767**	,417	-,050	,089
i	,181	,323	**,728**	,132
j	,387	**,697**	,369	,104
k	,110	,293	**,740**	,119
l	,174	**,732**	,317	,000
m	,181	**,775**	,298	,121
n	**,542**	**,611**	-,078	,358
o	,195	**,731**	,265	,249
p	,378	**,490**	,123	,478
q	**,778**	,412	-,101	,290
r	,149	,136	**,505**	,221
s	,241	,247	,339	**,662**
t	,100	,089	,166	**,821**

Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002

Should a validated HRQL scale go through all the steps of validation when translated in several different languages in order to be included in a drug development program ?

Regulators would tend to say yes ? No clear answer

3- Study design for HRQL assessment

- As a general rule, the validation of HRQL instrument should **preferably** have been completed before its use in therapeutic confirmatory trials. In principle, the same study should not be used to validate the HRQL instrument and to test for the HRQL change.

- If the HRQL instrument planned to be used is not validated (or is insufficiently validated), it is **recommended** to test it already in the therapeutic exploratory trials to be able to retest it again in therapeutic confirmatory trials. Indeed, if HRQL is planned to be assessed, it should be implemented in drug development plan as early as possible.

Should a validated HRQL scale go through all the steps of validation when translated in several different languages in order to be included in a drug development program ?

- Depends :
  - on timing in the drug development program,
  - on number of translated languages
- Translation process is unique and translator-dependant,
  → but if well performed, **should be enough** ?
- Verification of **psychometrics on baseline** (blinded) data for each language/country/culture ?
  - Although regulators seem reluctant to such analysis
    → Should be prespecified in the protocol
- Issue not the same between different European cultures, and between Western World and Asia ?

A few issues raised by the Efficacy Working Party (EMEA) - Dr Mira Pavlovic (AFSSAPS representative)

- How can we better understand the relationship between HRQL and Patient-Reported Outcomes in general ?
- Should a validated HRQL scale go through all the steps of validation when translated in several different languages in order to be included in a drug development program ?
- **Why developing a multi-domain HRQL scale, if at the end the sponsor predefines in a clinical trial one or two domains of this scale in order to gain a specific claim on these domains ?**
- Should the sponsor develop generic or disease specific questionnaires ?
- How is it possible for a reviewer to check that the claim asked by the sponsor corresponds to what has been really measured in clinical trials ?
- What are the important issues to check when reviewing a dossier with HRQL data ?

Workshop on validation of scales (October 2005, EWP, EMEA, London)
Development and validation of “multidomain scales” (example 3, in quality of life)

Why developing a multi-domain HRQL scale, if at the end the sponsor predefines in a clinical trial one or two domains of this scale in order to gain a specific claim on these domains ?

- There are situations where treatment improves specific domains of HRQL (such as physical or social functioning), which are considered important to patients.
- A company may seek specific claim based on the subset (one or two) of domains of HRQL, if the analysis plan **pre-specifies** which domains will be targeted as endpoints in the study.
- In addition, the use of specific HRQL domains as study endpoints pre-supposes that the HRQL instrument was adequately developed and fully validated prior to measuring the subset of domains chosen.

Why developing a multi-domain HRQL scale, if at the end the sponsor predefines in a clinical trial one or two domains of this scale in order to gain a specific claim on these domains ?

- A Company needs to document the change on the predefined HRQL domains of interest, and to provide information about the amount of change that is required to be considered as clinically meaningful.

- **In case of positive / relevant results, a specific claim reflecting domain(s) with improvement might be mentioned in the SmPC.**

- It is recommended that the claim always specify the changes observed in **all HRQL domains** for a given condition, including the domains with the improvement, the domains with no change and the **domains with the worsening**, if any. A **full disclosure** of complete results should be provided.

A few issues raised by the Efficacy Working Party (EMEA) - Dr Mira Pavlovic (AFSSAPS representative)

- How can we better understand the relationship between HRQL and Patient-Reported Outcomes in general ?
- Should a validated HRQL scale go through all the steps of validation when translated in several different languages in order to be included in a drug development program ?
- Why developing a multi-domain HRQL scale, if at the end the sponsor predefines in a clinical trial one or two domains of this scale in order to gain a specific claim on these domains ?
- **Should the sponsor develop generic or disease specific questionnaires ?**
- How is it possible for a reviewer to check that the claim asked by the sponsor corresponds to what has been really measured in clinical trials ?
- What are the important issues to check when reviewing a dossier with HRQL data ?

Workshop on validation of scales (October 2005, EWP, EMEA, London)
Development and validation of “multidomain scales” (example 3, in quality of life)

2. HRQL in drug approval process

- A claim about improvement in HRQL needs to be supported by data collected by instruments **validated for use in the corresponding condition**.
- In theory, both generic and disease specific questionnaires may be used for a given condition.
- In practice, it is very important to choose the questionnaire which contains/is adapted to explore the **domains relevant** for the disease and its treatment(s).
- Indeed, “HRQL improvement” as a claim implies **that the most important and clinically relevant health-related domains of functioning that impact patient’s quality of life are known and measured**.

What does mean a validated questionnaire ? What is the level of evidence ?

- Past : abuse of the term “quality of life”
- Present : abuse of “we have used a validated questionnaire”
- **Validated and relevant are not synonymous**

→ A **validated** scale doesn’t imply systematically that it is **relevant** for the population studied

→ A **validated** scale in a condition may not be valid anymore in **another** condition

Justification of the choice of a PRO instrument : A “Validated” instrument is not enough

Rationale (relevance) for the choice of a well-established and validated HRQL questionnaire, applied in a given condition

- **SF-36** for everything : “it has been used and published before”, “we will be able to compare”
- **MOS-HIV** validated before HAART
- **QLQ-C30** for peripheral arteriopathy obstructive disease (academic sponsored trial) : “it has been published before”
- **QLQ-C30** in some anemia (non cancer) disease (Dossier for approval) : “the principal investigator has used this questionnaire before in cancer trials...”

A few issues raised by the Efficacy Working Party (EMEA) - Dr Mira Pavlovic (AFSSAPS representative)

- How can we better understand the relationship between HRQL and Patient-Reported Outcomes in general ?
- Should a validated HRQL scale go through all the steps of validation when translated in several different languages in order to be included in a drug development program ?
- Why developing a multi-domain HRQL scale, if at the end the sponsor predefines in a clinical trial one or two domains of this scale in order to gain a specific claim on these domains ?
- Should the sponsor develop generic or disease specific questionnaires ?
- **How is it possible for a reviewer to check that the claim asked by the sponsor corresponds to what has been really measured in clinical trials ?**
- What are the important issues to check when reviewing a dossier with HRQL data ?

Workshop on validation of scales (October 2005, EWP, EMEA, London)
Development and validation of “multidomain scales” (example 3, in quality of life)

1. Introduction

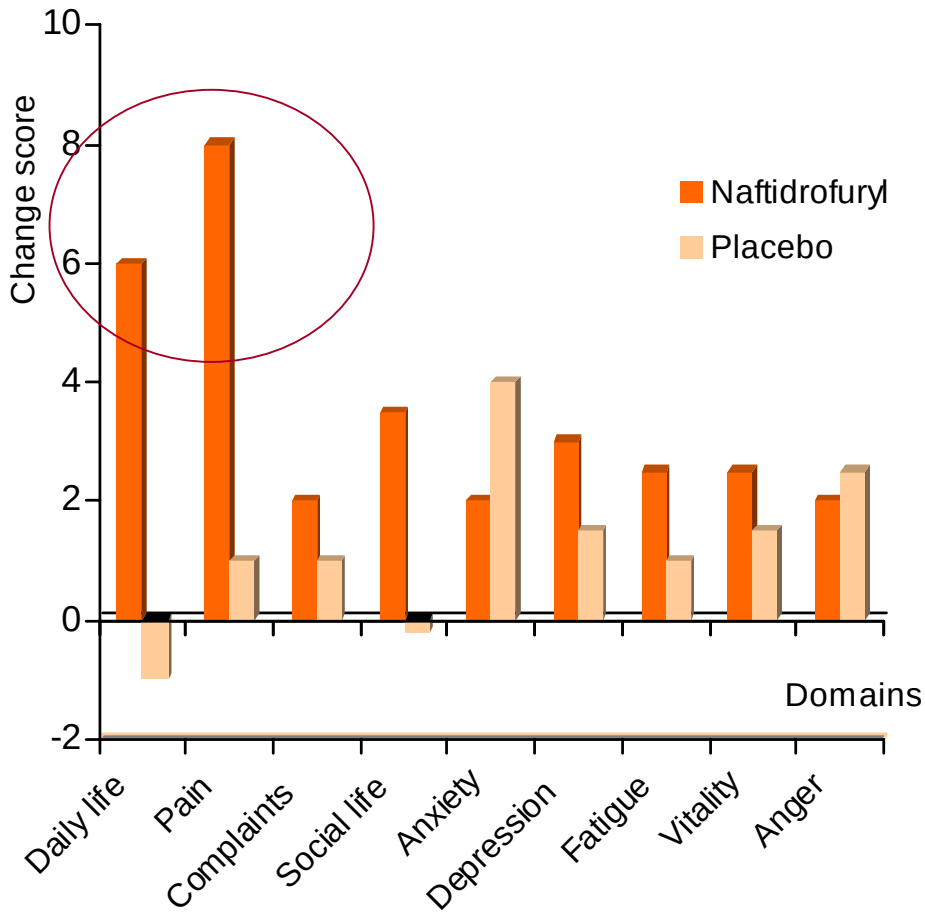
- As stated above, the notion of multidimensionality is a key component of definition of HRQL. A single domain, e.g., physical functioning or fatigue, is not considered as a HRQL (i.e. it cannot be the basis for a claim for a **global** HRQL improvement), even though it is a patient-reported.

2. HRQL in drug approval process

- In order to approve a global claim that a product “improves HRQL”, it would be necessary to demonstrate robust improvements in all or **most of these domains**.

How many and which PRO domains should improve for a claim ?

- 234 Patients with Peripheral Arteriopathy Occlusive Disease (PAOD)
- **HRQL primary endpoint** using the specific questionnaire : CLAU-S (9 domains, 80 items)
- **Results** : 2 domains significantly improved with drug (daily life, p=0.004; pain, p=0.001)
- **Should the planners have hypothesized that only these 2 domains would improve?**



The effects of naftidrofuryl on quality of life. Liard F et al. Dis Manage Health Outcomes 1997.

A few issues raised by the Efficacy Working Party (EMEA) - Dr Mira Pavlovic (AFSSAPS representative)

- How can we better understand the relationship between HRQL and Patient-Reported Outcomes in general ?
- Should a validated HRQL scale go through all the steps of validation when translated in several different languages in order to be included in a drug development program ?
- Should the sponsor develop generic or disease specific questionnaires ?
- Why developing a multi-domain HRQL scale, if at the end the sponsor predefines in a clinical trial one or two domains of this scale in order to gain a specific claim on these domains ?
- How is it possible for a reviewer to check that the claim asked by the sponsor corresponds to what has been really measured in clinical trials ?
- **What are the important issues to check when reviewing a dossier with HRQL data ?**

Workshop on validation of scales (October 2005, EWP, EMEA, London)
Development and validation of “multidomain scales” (example 3, in quality of life)

EMEA reflection paper
How to support a HRQL claim ?

- The claim in the SmPC with the respect to HRQL (i.e. in section 5.1) will always be considered depending on the **strength of the evidence and the relevance (pertinence and importance)** of the finding.

- The strength of the evidence should be based on
  - the rationale for HRQL assessment in the context of the disease/medicinal product
  - the justification of the choice of the HRQL questionnaire(s)
  - the objectives of HRQL assessment and the hypotheses of HRQL changes
  - the evidence of validation (and of cultural adaptation / translation if applicable) of the HRQL questionnaire(s)
  - the adequacy of the statistical analysis plan
  - and the relevance of observed changes

4. STATISTICAL ANALYSIS AND HYPOTHESIS

- The number of patients, necessary to support the change in the primary endpoint, is frequently sufficient to test for the HRQL change.
- In some situations, the number of patients is far too large and the trial is then overpowered, and allows to demonstrate significant but very small differences in HRQL scores, which are not relevant.
- **Therefore, every effort should be made to ensure that the sample size calculated for the primary endpoint is adequate for demonstrating hypotheses made a priori on the HRQL assessment.**
- The assessment of HRQL in a subset of the sample should be justified.

PROs (including HRQL) are unavoidably part of the Approval decision

Example of an drug in Irritable Bowel Syndrome
(European mutual recognition procedure)

PRO

- **Small** difference on pain versus placebo (primary endpoint)

HRQL

Tertiary endpoints (quality of life, satisfaction, utility and work productivity) bring **consistency** with the other endpoints, and they may thus reinforce the rather small clinical benefit observed on the co-primary endpoints, and **thus enhance the benefit/risk ratio**

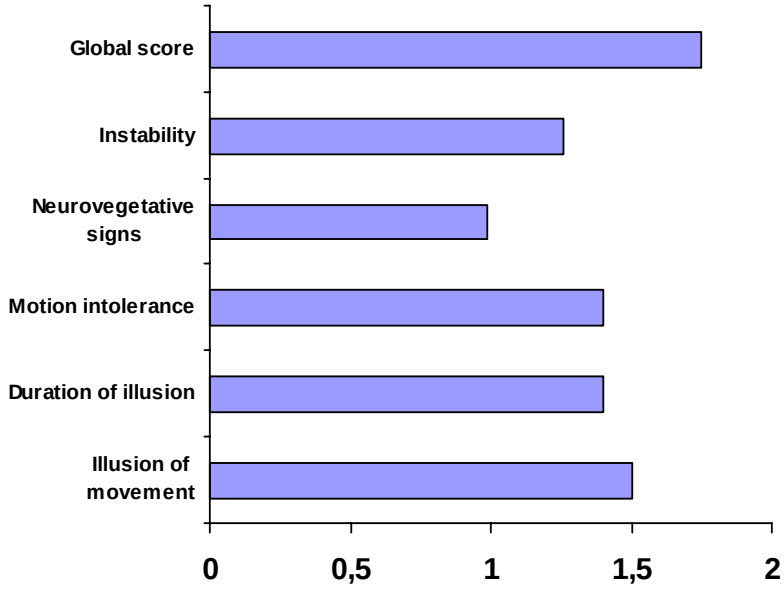
- **Not only patients tend to feel a little bit better for pain and symptoms, but they express a small improvement in some aspects of their daily life, and they are a little bit more productive for work**

January 2005

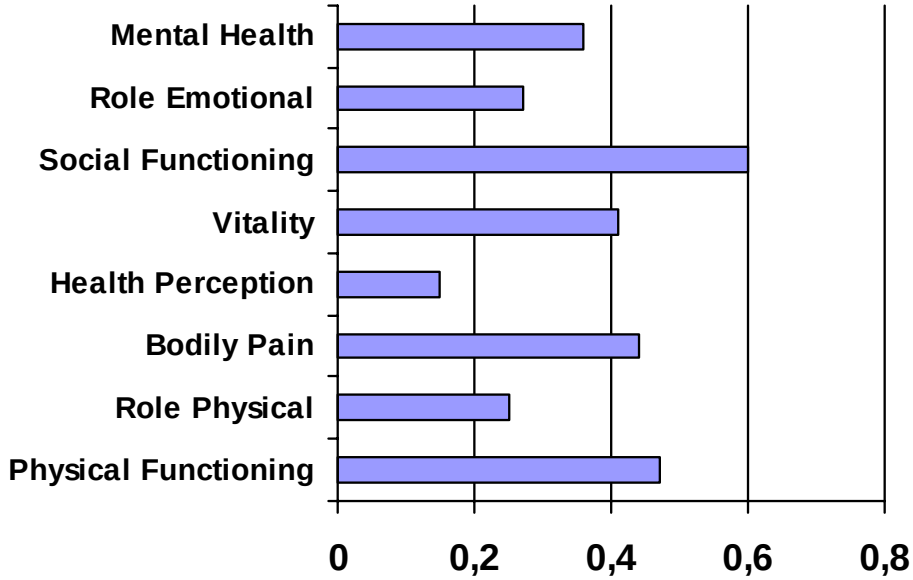
Interpretation of PROs results
Effect size

Longitudinal validation study : Effect Size (ES) of a symptomatic specific questionnaire (EEV) and the SF-36 calculated from the change as perceived by over 100 patients with vertigo after 4 weeks of treatment

European Evaluation of Vertigo (EEV)



Generic quality of life SF-36



Effect size (Distribution-based approach to express change in a standardized metric)
Dividing a difference between 2 groups by the SD

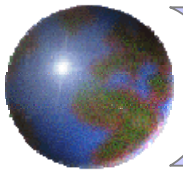
Effect Size	*Small*	*Moderate*	*Large*
Benchmark	> 0.20	> 0.50	> 0.80

Nobody knows if a 9 % difference of responders in IBS is worth giving a claim?

- Tegaserod / Irritable Bowel Syndrome
- Endpoints:
  - "did you have satisfactory relief of your overall IBS symptoms during last week?" (Y/N)
  - "did you have satisfactory relief of your abdominal discomfort or pain symptoms during last week?" (Y/N)
- **Responder** : satisfactory relief for at least 3 out of the 4 first 4 weeks
- Relief of overall IBS symptoms 33.7% vs 24.2% (placebo) → **9.3%**
- Relief of abdominal discomfort/pain : 31.3% vs 22.1% → **9.1%**

European mutual procedure (2005)

Tack J et al. A randomised controlled trial assessing the efficacy and safety of repeated tegaserod therapy in women with irritable bowel syndrome with constipation. Gut 2005; 54: 1707-1713.



Checklist for designing, conducting, reporting and reviewing HRQL - PRO in clinical trials

HRQL / PRO objectives

- Added value of HRQL / PRO — What is the claim expected ?
- Choice of the questionnaire(s) — Concepts important for the patients ?
- Hypotheses of HRQL / PRO changes — Even if secondary endpoints

Study design

- Basic principles of RCT fulfilled ? — Double-blind whenever possible
- Timing and frequency of assessment — Adapted to the concepts measured ?
- Mode and site of administration... — Procedures to ensure quality of data

HRQL / PRO measure

- Description of the measure (items, domains…) — Clarity, readability, recall period ?
- Evidence of validity — In the population under study ?
- Evidence of cultural adaptation — Through recommended procedure ?

Patient Reported Outcomes (PRO) and Regulatory Issues : A European Guidance Document for the improved integration of health-related quality of life assessment in the drug regulatory process. Chassany O et ERIQA Working Group. Drug Information Journal 2002.

Checklist for designing, conducting, reporting and reviewing HRQL - PRO in clinical trials

Statistical analysis plan

- Primary or secondary endpoint
- Sample size, subset of patients — Even for secondary endpoints
- ITT, type I error, missing data — Missing data & multiplicity issues

Reporting of results

- Participation rate, data completeness — Full disclosure of all scores
- Distribution of HRQL / PRO scores — Which and how many domains improve ?

Interpreting the results

- Effect size — Smaller ES with broader concepts
- Minimal Important Difference — MID is study specific
- Responders, number needed to treat… — May not be similar to MID

Giving the desired claim ? — What has been measured ?

Patient Reported Outcomes (PRO) and Regulatory Issues : A European Guidance Document for the improved integration of health-related quality of life assessment in the drug regulatory process. Chassany O et ERIQA Working Group. Drug Information Journal 2002.