



## **Pan-Pacific Conference**

# **Guidance for the improved integration of Health-Related Quality of Life assessment in the drug regulatory process**

## **The ERIQA\* Working Group experience**

**\* European Regulatory Issues on Quality of life Assessment**

**Tokyo, Japan - April 13-15, 2001**

***Olivier Chassany, MD, Clinical Research Delegation,  
Assistance Publique - Hôpitaux de Paris (AP-HP),  
& Medical University, Paris, France***

# Drug Approval Process

## *Major biases encountered in reviewing dossiers*

- **Many** clinical trials with HRQL data
- **Few** specific labeling or promotional claims (US, Europe)
- **Problems of :**
  - Recognition by regulators
  - Trials quality
  - Interpretation
  - Lack of guidelines
- No justification of HRQL choice
- No evidence of questionnaire validation
- No objective of HRQL changes
- No justification of sample size
- No description of the follow up of patients
- No clear handling of missing data
- Not all patients are analysed
- No correct presentation of results
- No adjustment for multiple comparisons
- No interpretation of results

Chassany O et al. Reporting on quality of life in randomised controlled trials. BMJ 1999.

# HRQL is a PRO... (Patient-Reported Outcome)

*Harmonization meeting - FDA 16 Feb 2001*

*ERIQA, ISOQOL, ISPOR, PhRMA HOC*

## Positive

- Patient has a unique voice and valuable perspective that should play a role in medical decision making
- PROs can be measured in reliable and valid ways
- PROs are increasingly used as efficacy endpoints in randomized controlled trials

## Still in debate

- Clearly define the use of PRO (i.e. what is measured ?)
- What are the methods to handle missing data
- What is meant by clinical significance ?
- How should results be presented is the label ?
- What level of evidence is needed to support a claim ?

# Review of existing EMEA guidelines

1999-2000

**Objective :** to identify diseases or drugs in which a formal HRQOL assessment is recommended

EMEA website : [www.eudra.org/emea.html](http://www.eudra.org/emea.html)

*Apolone et al.  
Pharmacoeconomics 2001*



Type of Documents	Number
<b>CPMP / EWP (Efficacy Working Group)</b>	<b>19</b>
- Notes for guidance (NG)	9
- Concept Paper (CP)	3
- Points to Consider (PC)	5
- Position Statements (PS)	2
<b>CPMP / ICH</b>	
- Notes for guidance (NG)	1
<b>CPMP / EPAR</b> (representing 26 Products)	<b>104</b>
<b>Miscellaneous</b>	<b>9</b>
- CPMP	
- minutes reports, workshop, letters	5
- Assessment/opinion	2
- CVMP	2
<b>TOTAL</b>	<b>133</b>

- Cardiac Failure
- Angina Pectoris
- Peripheral Arterial Occlusive
- Asthma, COPD
- Osteoarthritis
- Rheumatoid Arthritis
- Crohn's Disease
- Weight Control
- Anti-Cancer
- Parkinson's Disease
- Alzheimer's Disease
- ALS
- Multiple Sclerosis

# **Negative recommendations**

---

- **CPMP/EWP/563/95 NG : Parkinson's Disease**

*The use of indirect efficacy variables as primary efficacy variable in pivotal studies, **such as...[..] quality of life is not recommended** unless the association between these variables and improvement in core symptoms or motor fluctuations or handicap has been proven*

- **CPMP/EWP/2284/99 PC : Crohn's Disease**

*Other end-points such as...[..]...improvement in HRQL can be subsumed as response variables or outcomes measures of either the treatment of active disease or maintenance of remission. **Unless otherwise justified, they should not be mentioned in the indication.***

# Results of EMEA recommendations

---

## “Negative”

- Vague in most cases, too generic
- Inconsistent between each other
- Reveal in some cases a lack of knowledge of the field
- Choice of questionnaire arguable
- Not updated
- Do not exist in relevant diseases: HIV/AIDS, HBP, IBS, liver dis...
- Interest in HRQL but “secondary supportive” claim possible
- **No definite guideline**

## “Positive”

- Do exist...: real interest
- HRQL recognized as a valuable endpoint (mainly secondary) : *valuable information, important dimension, recommended (CHF)*
- Reveal in some cases a certain level of knowledge of the field
- Word “claims” explicitly mentioned...

# Meeting with European Regulators (1999)

## *Role of HRQL Outcome in the European Drug Regulatory process*

EMEA CPMP Chairman (Pr Alexandre)

French Drug Licensing Committee President - AFSSPS (Pr Caulin)

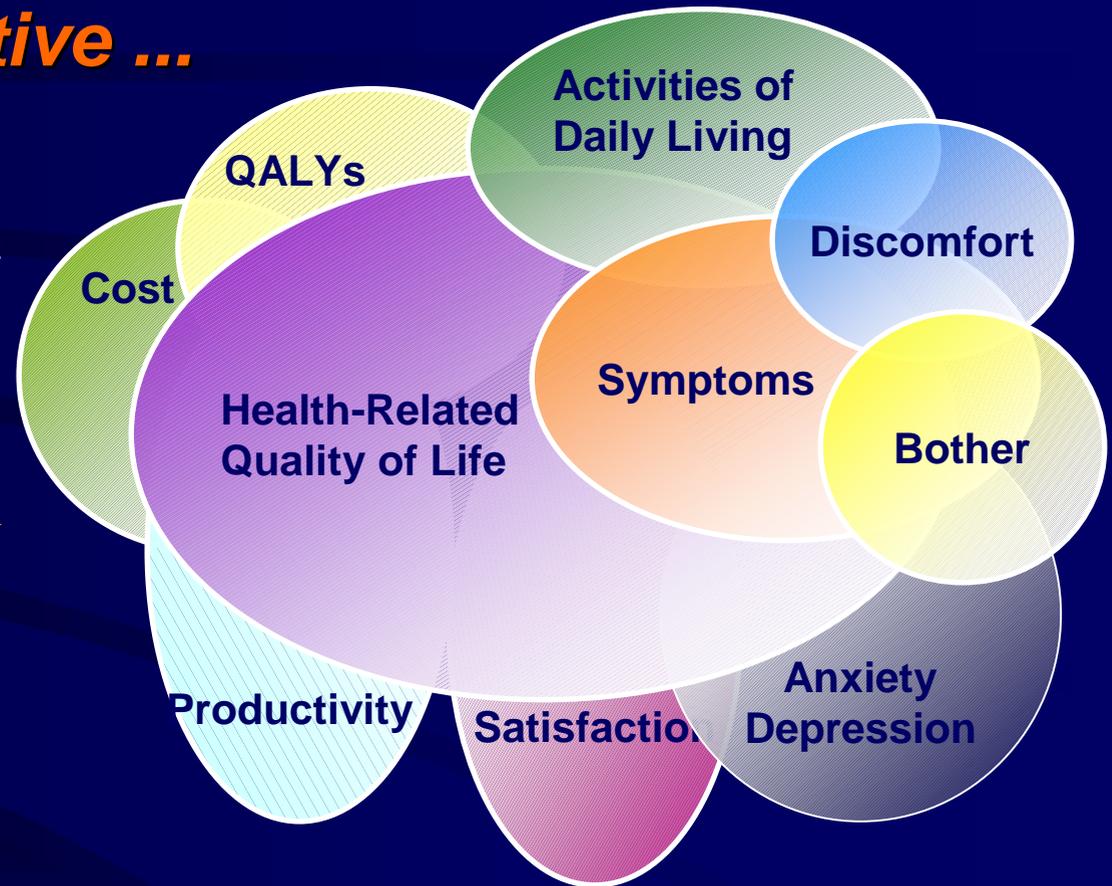
*→ HRQL to be considered as a credible criterion if there is enough evidence about the :*

- Added-value of HRQL with respect to other criteria
- Psychometric properties of the HRQL instruments
- International validation of the HRQL instruments
- Adequacy of the statistical analysis plan
- Clinical significance of observed changes

# Added-value of HRQL and PRO

## 1- Shift from acute disease to chronic illnesses The patient's perspective ...

- Increasing prevalence and greater longevity with chronic illnesses
- Treatment often not curative
- Aims to improve **function and well-being** by reducing the severity of an illness or by limiting disease progression (e.g. cancer, AIDS, Crohn's disease, rheumatoid arthritis, epilepsy, arthritis...)



Adapted from Laurie Burke (DDMAC, FDA) presentation, Acceptable Evidence for Pharmaceutical Advertising and Labeling, DIA, New Orleans, October 2000

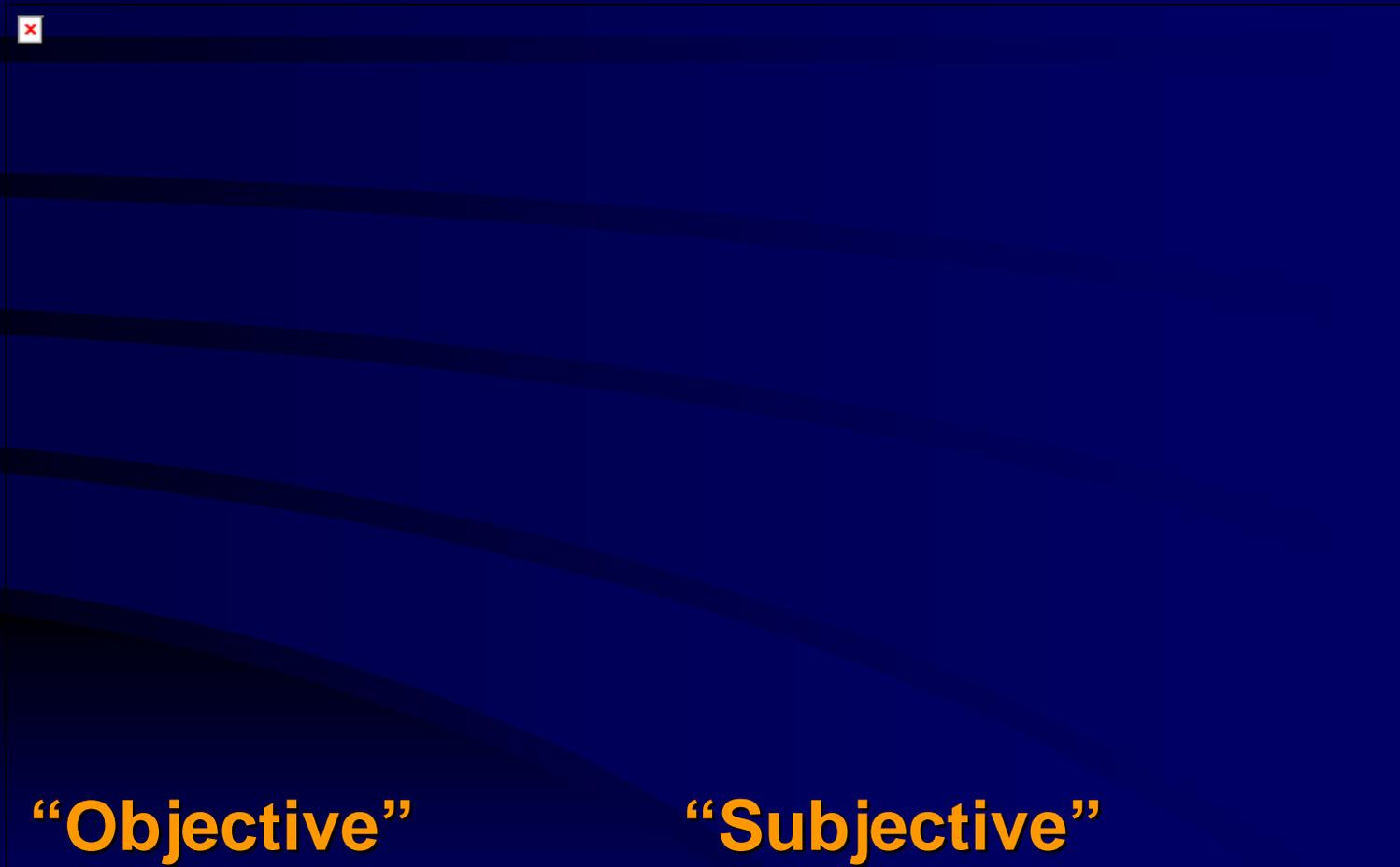
# Added-value of HRQL and PRO

## 2- Modest correlation

- Traditional measures of disease activity are poorly correlated with PRO.
  - They do not reflect the patient's perception of their function and well-being.
  - Ratings from patients frequently differ with those of physicians.
  - The direction of this discrepancy may not be predicted a priori.
- Back pain in lombalgia (pain score and HRQL)
  - Gastric acid secretion in GERD (Ph-metry and pyrosis)
  - Exercise tolerance in angina pectoris (performance and ability to go for shopping)
  - Spirometry in asthma or COPD (Forced Expiratory Volume and daily functioning,  $r = 0.10$  to  $0.30$ )

# Added-value of HRQL and PRO

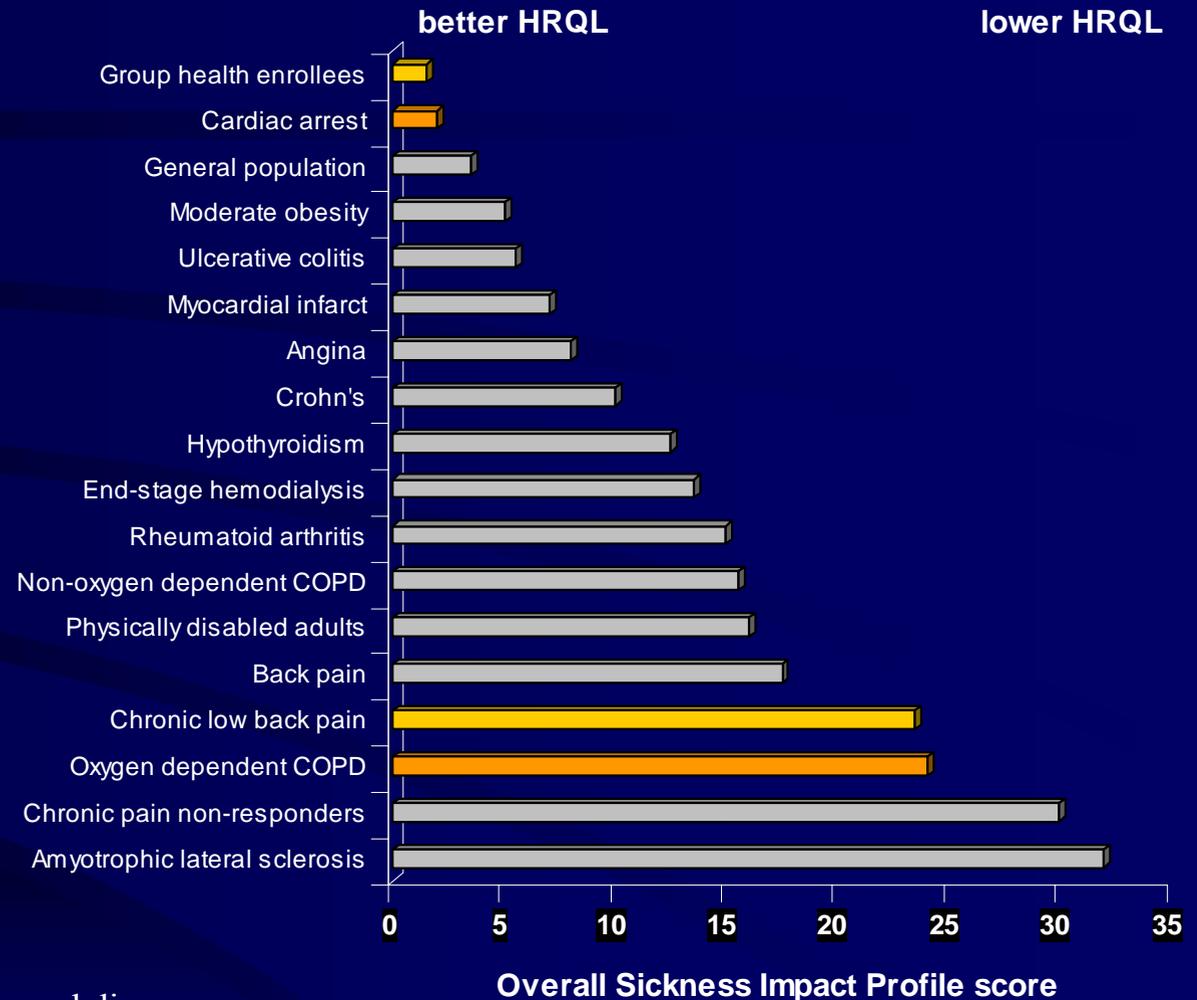
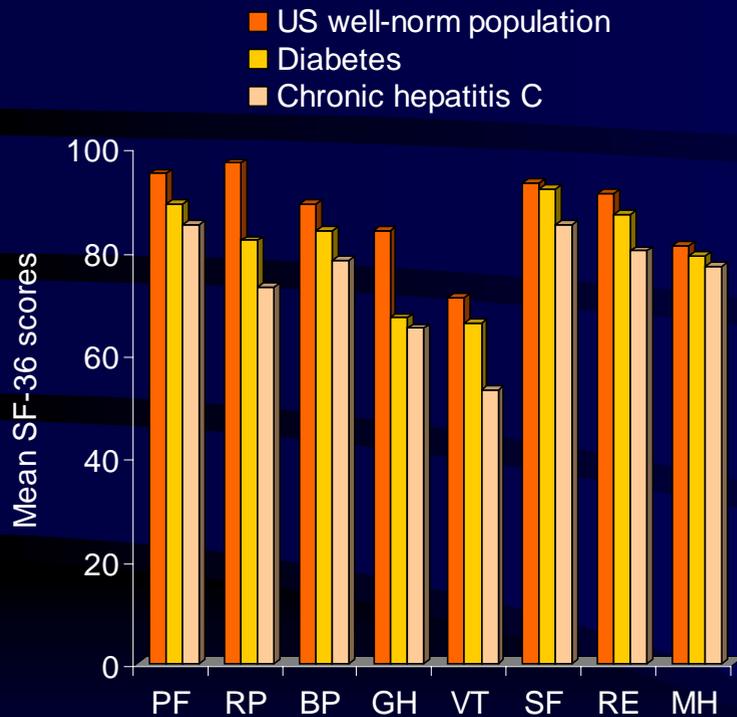
## 2- Modest correlation



Exercise test versus physical functioning,  $r = 0.40$

# Added-value of HRQL and PRO

## 3- Negative impact



Bayliss MS, et al. Questionnaire to assess the generic and disease-specific health outcomes of patients with chronic hepatitis C. *Qual Life Res* 1998; 7: 39-55.

Bergner M, et al. The sickness impact profile: validation of a health status measure. *Med Care* 1976.

# Added-value of HRQL and PRO

## 4- When HRQL is relevant in clinical trials ?

- Patient's self-report is the primary or sole indicator of disease activity
  - No objective marker or several possible markers of disease activity
  - Disease expressed by many symptoms
  - To ensure that treatments having a small impact of survival, do not adversely affect patients' lives due to morbidity, functional or psychological impairments or side effects
  - Equivalence trial where the drug under study may have PRO benefits
- IBS, dyspepsia
  - Migraine
  - Asthma, COPD
  - Arthritis
  - Rhinitis
  - Erectile dysfunction
  - Menopause
  - Psoriasis...
  - Rheumatoid arthritis
  - Crohn's disease
  - Chronic heart failure
  - Cancer
  - AIDS...

# Psychometric validation

*HRQL instruments are better validated...*

... than traditional measures

- Item generation, scaling
- Reliability
- Content validity
- Construct validity
  - Structural validity
  - Clinical validity
  - Concurrent validity
  - Predictive validity
- Responsiveness
- Clinical trial can serve as a validation study (item selection and scoring algorithm should not be biased toward the treatment effect)

## French Drug approval (1999)

Proton pump inhibitor / dyspepsia

Phase III, RCT, DB,

vs. comparator & placebo

n = 810, 2 weeks duration

## HRQL assessment (secondary)

- Unknown HRQL index
- No description of validation process, no reference
- Evidence of validation : 10 items concerning gastro-intestinal symptoms were added and approved by Pr X !

# Cultural adaptation

## *Linguistic validation*

Have you had any difficulties carrying out your leisure activities (DIY, gardening, **walks in the country ...** ) ?

**French:** Ballades en forêt  
(**Walks in the forest**)

**German:** Waldspaziergänge  
(**Walks in the forest**)

**Slovak:** Zbierat' Huby  
(**To search for Mushrooms**)

## *Psychometric validation*

- The translated versions must display the same scale structure (item aggregation into dimensions)
- And should have been tested in terms of reliability and validity
- **In order to aggregate HRQL data from the different countries of the trial**

# Selection of HRQL questionnaire

*Hypotheses of HRQL change*

*Justification of the choice of HRQL questionnaire*

*What is measured ?*

GSRS : Gastrointestinal Symptom Rating Scale (5 scales)

## **1- reflux syndrome (2 items)**

Have you been bothered by acid reflux during the past week ?

Have you been bothered by heartburn during the past week ?

2- abdominal pain

3- diarrhoea syndrome

4- indigestion syndrome

5- constipation syndrome

Reliability and validity of the gastrointestinal symptom rating scale in patients with GERD. Qual Life Res 1998.

## **French Drug Approval (1999)**

- Proton pump inhibitor / oesophagitis
- Phase III : 2 studies in USA, 1 study in Europe
- > 700 patients included

## **HRQL claim**

No clear definition of HRQL in the study report, neither in the protocol

→ Overall physical well being (0 to 4) ?

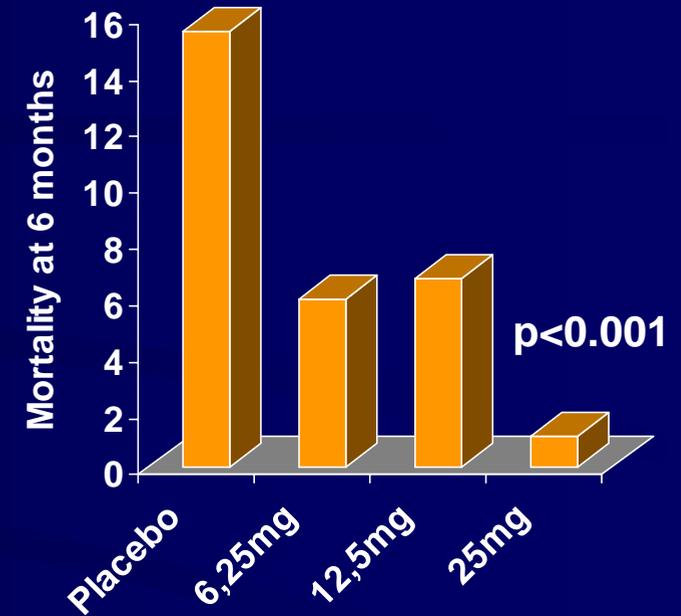
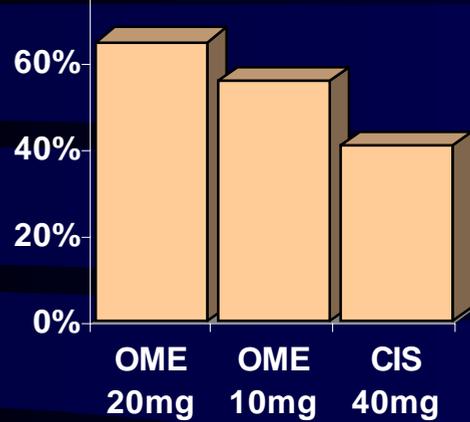
→ **Time lost from usual activities of daily living ?**

(less time lost in placebo group !)

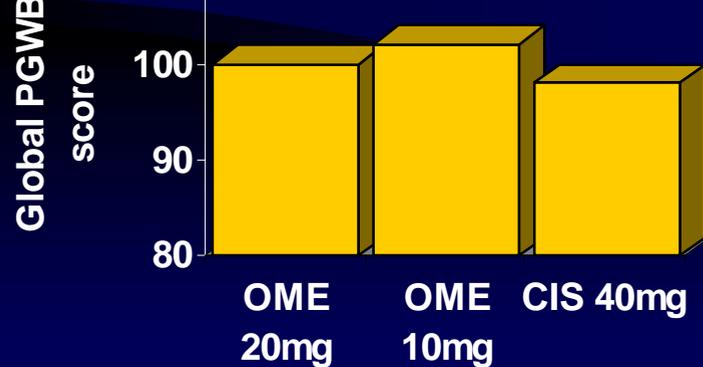
# Selection of a HRQL instrument

*Is the questionnaire responsive ?*

Resolution of symptoms at 4 weeks



Global PGWB score After 4 weeks of treatment



MLwHF (0-105)	Placebo	6,25mg	12mg	25mg
n	72	77	80	78
Baseline	47.7	45.8	43.9	43.6
Endpoint	40.4	38	36.5	38.2

Bristow MR et al. Carvedilol. Circulation 1996.

# Study design

*Comparative randomized trials are a pre-requisite and double-blinded...*

- HRQL claims cannot be based on **non-comparative** and **non-blind** clinical trials.
- They always lead to a **higher rate** of positive results.
- No causal link can be established between the therapeutic intervention and the HRQL change.

## French Drug Approval (2000)

HRQL claim in Benign Hypertrophy Prostate

1- Cohort study (n = 7093) - specific scale.

- HRQL score improving from  $91 \pm 32$  (J0) to  $109 \pm 31$  (J3) [75% patients] : + 29%

- Improvement by 50% of symptoms (similar to the one observed in study versus placebo, they forget to say that placebo leads to 40% improvement)

2- Cohort study (n = 5849)

3- Cohort study (n = 4951) abstract

→ Why not only 1 trial vs placebo ?

# Study design

## *Practical issues related to HRQL component*

- **Eligibility criteria** : if HRQL primary endpoint, set a minimal impairment of HRQL (as for other criteria, e.g. pain, asthma onset... )
- **Timing and frequency** of HRQL assessment :
  - At baseline, at the end of the study or at withdrawal
- **Mode and site** of HRQL administration :
  - Self-administered whenever possible 
  - Assure the confidentiality
  - Before the medical consultation
- **Data monitoring** and quality assurance
- **Procedures** for prevention and handling of missing data

In one study evaluating sexual impairment induced by antihypertensive treatment in male patients, the answers given by nurses, patients and their spouses were quite different, respectively **low, moderate and important...**

# Adequate statistical analysis plan

## Sample size

- Estimation should be performed whether HRQL is a primary or **secondary endpoint**
- Without pre-specified hypotheses, there is a high risk of under-powering the study, leading to a non-significant difference between groups
- Conversely, including hundreds of patients is likely to reveal a **non pertinent** but statistically significant difference in HRQL scores

Ranitidine vs. Placebo among > 500 patients with gastro-esophageal reflux disease

SF-36	Ran	Pla		p
<b>PF</b>	<b>82.6</b>	<b>80,0</b>	<b>2.6</b>	<b>0.019</b>
RP	77,0	74.6		NS
<b>BP</b>	<b>73.8</b>	<b>69.1</b>	<b>4.7</b>	<b>0.003</b>
GH	69.7	68.7		NS
<b>VT</b>	<b>58,0</b>	<b>54.4</b>	<b>3.6</b>	<b>0.005</b>
SF	85.5	83.7		NS
RE	81.9	78.2		NS
MH	72.5	71.7		NS

Rush DR et al. Clinical effectiveness and quality of life with ranitidine vs placebo in gastroesophageal reflux disease patients. J Fam Pract 1995.

# Adequate statistical analysis plan

## Sample size

### European Drug Approval (2000)

Treatment in CHF

Metoprolol vs placebo

Primary endpoint : mortality

**Tertiary endpoint : HRQL**

- **MLwHF (21 items)**
- **Change in the total score from baseline to 21 months is the main efficacy variable.**
- **Range score : 0 (good) - 105**

### Sample size determination :

- 419 patients per group
- To detect a difference of **3 units** between the groups
- With a power of **80%**.
- From previous studies : estimated within-subject standard deviation : **16**.
- Significance level set at **5%** (two-sided).

# Adequate statistical analysis plan

## *Intent to treat analysis*

- May disadvantage the treatment under study, but excluding patients from the analysis, whatever the reason, may result in bias (*i.e. destruction of the comparability of treatment groups*)
- Moreover, patients who do not complete the trial *may have the poorest HRQL*

## European mutual recognition

- Treatment in claudication
- Phase III, RCT, DB, n = 422, 6 months
- New treatment vs placebo
- Results : initial change distance :  $\Delta$  32% vs placebo

## HRQL assessment (secondary)

- PQVS French generic questionnaire
- Global satisfaction (p = 0.049, t-test)
- Analysis performed on **324 patients**
- How many and how were handled missing data ?

# Adequate statistical analysis plan

## Missing data

- HRQL data may be missing **more frequently** than for other endpoints, as many items are to be completed over time
- There is strong evidence that missing data are not “**missing at random**” (i.e. they are related to either the treatment or the underlying disease) and cannot be ignored without introducing bias

## French drug approval (1999)

- Treatment in rheumatoid arthritis
- Phase III, RCT, DB, n = 485, 52 weeks
- New treatment vs comparator & placebo,

Results : some improvement in clinical end-points

HRQL assessment (secondary) :

### HAQ and SF36

- HRQL better with new drug, but **only 280** patients analysed, how missing data
- Type I error ? (multiple comparisons)
- Many adverse events (withdrawal of patients 22% vs 8% placebo)

# Adequate statistical analysis plan

## *Missing data*

- **Prevention** of missing data during the trial
- **Description** of missing data in the report (items, questionnaires)
- **Imputation** of missing data in the analysis (Last observation carried forward, regression...)

## French drug approval (2001)

- Treatment in claudication
- Phase III, RCT, DB, n = 269, 6 months
- Treatment vs placebo

Results : Improvement of distance walk over placebo

HRQL assessment (primary) :

**CLAU-S : specific questionnaire**

→ Analysis on 255 patients

→ 8 patients in the active group not analyzed due to adverse effect (1 patient in placebo group)

- Gangrene (n = 1)
- Claudication worsening (n = 1)

# Adequate statistical analysis plan

## *Multiple statistical tests*

- Statistical analysis of HRQL data may be associated with a high incidence of type I error.
- Significant effects are found simply because of the multiple comparisons between treatment groups of many scales repeated over time.

→ Reducing the number of statistical comparisons (selection of time point or domains)

→ Statistical adaptation to the number of comparisons (e.g. Bonferroni)

Number of tests	Global type I error
1	0.05
2	0.08
3	0.11
4	0.13
5	0.14
10	0.19
20	0.25
50	0.32

according to the number of statistical tests performed within a study at the 0.05 level of significance

# Adequate statistical analysis plan

## *Multiple statistical tests*

- For example, performance of 20 statistical tests without adjustment on a questionnaire with 20 items or scales is associated with a **0.25 type I error**.
- This means that a false statistically significant difference at the 0.05 level is likely to appear by random in approximately **four items** or scales.

## European mutual recognition

- Treatment in claudication
- Phase III, RCT, DB, n = 324, 3 months
- New treatment vs comparator
- Results : **NO** difference in walking in ITT

## HRQL assessment (secondary)

- PQVS French generic questionnaire
- univariate analysis : statistical difference for **4 items** among 19  
(Per protocol analysis, n = 268)

# Reporting of results

*Results presented in accordance with the protocol ?  
Full disclosure of all results ?*

## European Drug Approval (2000)

Treatment in CHF

In the Clinical expert and study reports : important improvement of HRQL on a global assessment scale (OTE, single item)

In the protocol :

- Principal statistical analysis : MLwHF specific questionnaire (Minnesota Living with Heart failure)
  - No between-group difference
- Secondary analysis : OTE

## French Drug Approval (2001)

- Treatment in Claudication
- Phase III, RCT, DB, n = 269, 6 months
- Treatment vs placebo
- **Results** : Improvement of distance walk over placebo

### HRQL assessment (primary) :

- **CLAU-S** : specific questionnaire
- Assessment at 3 and 6 months
- ~~R~~Results presented as **final evaluation** : “daily activities” and “pain” domains in favor of the active treatment.
- ~~I~~n fact : results at 3 months

# Interpretation of results

*Report of values, means, SD, 95%IC, range...*

## 1. Understanding the content of the scales

- Description of the **content** of domains

## 2. Understanding size of changes from baseline and differences between groups

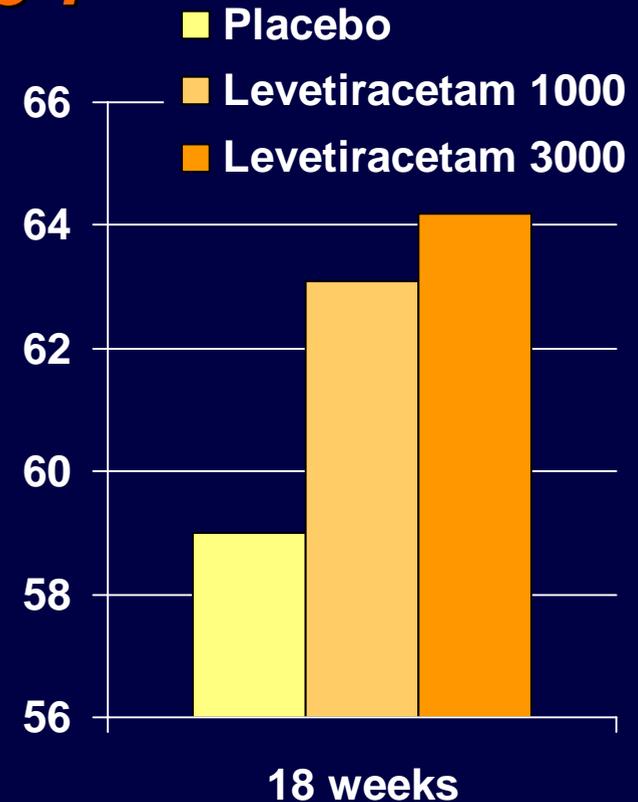
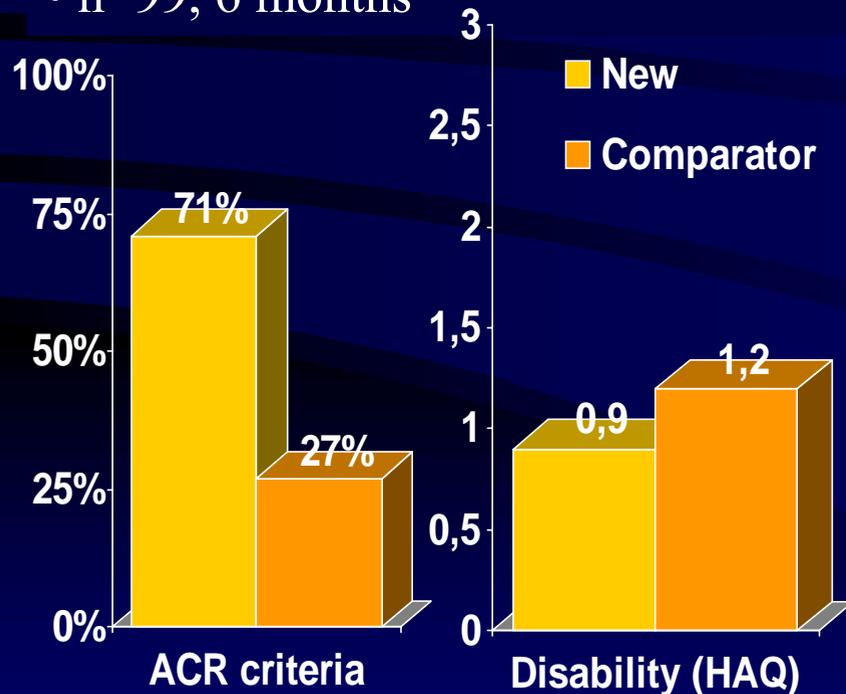
- **Distribution** of HRQL scores within- and between group
- Description of score range : min-max
- 95%IC of the difference and/or odds ratio of the difference
- **Effect size** :  $\Delta^{\text{pre-post}} / \sigma^{\text{baseline score}}$
- SRM Guyatt's :  $\Delta^{\text{pre-post}} / \sigma^{\text{change in stable patients}}$

# Interpretation of results

*What is the magnitude of change ?*

## *Dossier for drug approval*

- Treatment in rheumatoid arthritis
- Phase III, RCT, DB, vs comparator,
- n=99, 6 months



Mean QOLIE-31 total scores at 18 wk for Seizure patients (n = 246) .  
Cramer et al, Epilepsia 2000.

# Interpretation of results

## Effect Size

GERD	Treatment group		Difference	ES
PGWB global score* (Revicki, Dig Dis 1998)	OME 82.5	RAN 78.8	3.7	<b>0.22</b>
PGWB global score* (Havelund, Am J Gastro 1999)	OME 103.9	PLA 100.6	4.5	<b>0.26</b>
GSRs global score** (Festen, Am J Gastro 1999)	OME 12.3	RAN 10	3.3	<b>0.30</b>

range score : \* (22-132), \*\* (5-35)

Effect Size	No change	Small change (non pertinent)	Moderate change	Large change
	< 0.20	0.20-0.50	<b>0.50-0.80</b>	> 0.80

# Interpretation of results

*What is the minimal meaningful change ?*

3. Comparing changes to :

- norms or known-group references
- patient global rating of change : **minimal important difference (MID)**

4. Number needed to treat

- Number of patients that have to be treated for one patient to improve his HRQL (responder).
- The major limitation is to define the status of responder/non responder (i.e. how is set the clinical important difference ?)

# Interpretation of results

## *Predictive value of...*

- **Clinical end-points :**  
morbidity or mortality
- **Patient behaviour :**  
health resource utilisation
- **Economic endpoint :**  
productivity, days out of work

	Male (n = 176)	Female (n = 349)
<b>SF-36</b>		
PF	0,02	0,156
RP	0,038	0,148
<b>BP</b>	<b>0,003</b>	<b>0,0002</b>
GH	0,045	0,045
VT	0,002	0,2
<b>SF</b>	<b>0,0003</b>	0,019
RE	0,103	0,492
MH	0,027	0,179

Prediction of SF-36 scales of the number of medical visits for IBS during the past 6 months. Fullerton S et al. Gastroenterology 1998; 114: A14.

# HRQL claim

*How many domains should improve ?*

*What is measured ?*

- 90 (6 x 15) statistical tests
- Difference of **0.2** (range 1-7) at 3 months
- No difference at 12 months

**Abstract** “**Aerobic group-training of elderly patients recovering from an acute coronary event beneficially influences physical fitness and several parameters expressing quality of life**”

	J3	J12
Symptoms		
- Chest pain	NS	NS
- <b>Shortness of breath</b>	<b>&lt;0.05</b>	NS
- Dizziness	NS	NS
- <b>Palpitation</b>	<b>&lt;0.05</b>	NS
- Cognitive ability	NS	NS
Alertness	NS	NS
Quality of sleep	NS	NS
Physical ability	NS	NS
Daily ability	NS	NS
Depression	NS	NS
Self perceived health	NS	NS
Ladder of life: future	NS	NS
<b>Fitness</b>	<b>&lt;0.05</b>	NS
<b>Physical activity</b>	<b>&lt;0.01</b>	NS

# HRQL claim

*How many domains should improve ?*

*HRQL should be consistent with clinical results*

<b>Claudication</b>	<b>Naftidrofuryl</b>	<b>New treatment</b>
Study design	III, RCT, DB, placebo	II, RCT, DB, placebo
Length	6 months	6 months
Primary endpoint	CLAUS	<b>Absolute change distance : no difference</b>
Patients (ITT)	250 (234)	340
Domains improved	<b>Daily Life Pain</b>	<b>Social Function (SF-36) Mental Health</b>
Domains not improved	<b>7 other domains</b>	<b>Physical functioning Pain</b>
	Liard et al. Dis Manag Health Outcomes 1997	European Mutual Recognition

# Level of evidence to support a claim

***Prerequisite : protocol, study report & clinical expert report must document sufficient HRQL assessment to permit a critical review***

- Evidence for claims of HRQL benefits need to be consistent with the standards for claims of clinical efficacy.
- Consider HRQL or PRO like any other evaluation criteria
- Think about it early in the clinical development
- Perform trials with claim in mind
- Pre-specify a priori the major domains that are expected to improve, consistent with the disease
- Domain specific claims require statistical significance within that domain. But it does not constitute a general HRQL claim

**Currently : more is asked for HRQL data than for other criteria**

# Checklist for designing, conducting and reporting HRQL studies in clinical trials

## 1- Objective of the HRQL described

- Relevance for assessing HRQL
- Choice of the questionnaires
- Hypotheses of HRQL changes

## 2- Study design described

- Basic principles of RCT fulfilled ?
- Timing and frequency of assessment
- Mode and site of administration...

## 3- Description of the HRQL measure

- Description of the content of domains
- Reliability, Validity, Responsiveness
- Cultural adaptation

## 4- Description of the statistical analysis

- Primary or secondary endpoint
- Efficacy or equivalence trial
- Sample size
- ITT, type I error, missing data

## 5- Reporting of results

- Participation rate, data completeness
- Distribution of HRQL scores

## 6- Attempt to interpret the results

- Effect size
- Comparisons of scores with other scores
- Comparison with external criteria
- Number needed to treat...

O Chassany, et al for the European Regulatory Issues on Quality of Life Assessment (ERIQA) group. European guidance document. **Submitted to DIA journal 2001.**