Patient-Reported Outcomes in Phase III
Regulatory Requirements
Regulatory issues and best practice

Olivier CHASSANY, MD, PhD
Medical Manager, Clinical Research Dept (institutional sponsor)
Assistance Publique - Hôpitaux de Paris, France
Why there are so few HRQL mention in labelling?

Drug Approval Process
Major biases encountered in reviewing dossiers

- 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

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

**HRQL / PRO objectives**
- Added value of HRQL / PRO
- Choice of the questionnaires
- Hypotheses of HRQL / PRO changes

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

**HRQL / PRO measure**
- Description of the measure (items, domains…)
- Evidence of validity
- Evidence of cultural adaptation

**Statistical analysis plan**
- Primary or secondary endpoint
- Superiority or equivalence trial
- Sample size
- ITT, type I error, missing data

**Reporting of results**
- Participation rate, data completeness
- Distribution of HRQL / PRO scores

**Interpreting the results**
- Effect size
- Minimal Important Difference
- Number needed to treat...

Define the conditions for which the measurement of HRQL/PRO in clinical trial is useful

- Patient’s self-report is the primary or sole indicator of disease activity, e.g. dermatological disorders (psoriasis, acne), erection dysfunction
- No objective marker or several possible markers of disease activity (migraine, osteoarthritis, asthma, menopause, heart failure)
- Disease expressed by many symptoms (IBS)
- To ensure that treatments prolonging survival (AIDS), do not adversely affect patients’ lives due to morbidity, functional or psychological impairments or side effects
- The treatment does not seem to improve survival (cancer, rheumatoid arthritis, Parkinson’s disease), but it could improve HRQL, by reducing pain, anxiety, level of stress or by improving the functional status.
What is not quality of life?

The abuse of the term HRQL in some clinical trials, whereas the questionnaire measured anything else

– A listing of symptoms or of side effects cannot claim to measure HRQL
– The following concepts cannot alone explore all HRQL:
  – physical or intellectual performance scale
  – handicap or functional incapacity scale
  – anxiety or depression scale
  – tiredness or pain scale
  – symptom bother scale.
Choice of instrument - What are we measuring?

Different PROs: example of HIV/AIDS

Fatigue
Items related to intensity, circumstances, and consequences of fatigue.


Cognitive function
Dutch four-item MOS-HIV cognitive functional status subscale.


Treatment satisfaction
Weak correlation between severity of side-effects and score of satisfaction concerning these side-effects (r = 0.18)


Doctor-Patient satisfaction
Satisfaction among HIV-infected patients was not associated with QOL

Importance of various areas of limitations due to asthma among Harlem emergency department users (n = 247) mostly Afro-American patients with a low socio-economic status and a lower compliance.

Asthma-related limitations in sexual functioning: an important but neglected area of quality of life.

Table 1: Analysis of Well-Being between indapamide and captopril.

<table>
<thead>
<tr>
<th></th>
<th>Indapamide before</th>
<th>Indapamide after</th>
<th>Captopril before</th>
<th>Captopril after</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal presentation</td>
<td>0</td>
<td>0</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>Communication</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Muscular or sensory asthenia</td>
<td>1.21</td>
<td>1.07</td>
<td>1.62</td>
<td>1.06</td>
</tr>
<tr>
<td>Functional disturbances</td>
<td>0.36</td>
<td>0.28</td>
<td>0.62</td>
<td>0.25</td>
</tr>
<tr>
<td>Sleep disorders</td>
<td>1.43</td>
<td>1.00</td>
<td>1.37</td>
<td>1.31</td>
</tr>
<tr>
<td>Mental asthenia</td>
<td>0.14</td>
<td>0.43</td>
<td>0.37</td>
<td>0.31</td>
</tr>
<tr>
<td>Emotional lability</td>
<td>0.50</td>
<td>0.28</td>
<td>0.56</td>
<td>0.50</td>
</tr>
<tr>
<td>Anxiety</td>
<td>1.86</td>
<td>1.50</td>
<td>1.24</td>
<td>0.61</td>
</tr>
<tr>
<td>Will</td>
<td>0.28</td>
<td>0.28</td>
<td>0.43</td>
<td>0.31</td>
</tr>
<tr>
<td>Personality</td>
<td>0.50</td>
<td>0.43</td>
<td>0.87</td>
<td>0.50</td>
</tr>
<tr>
<td>Well-being</td>
<td>0.72</td>
<td>0.86</td>
<td>0.87</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Figure 6: General well-being assessment by the investigator. Differences in scoring from baseline to the final visit.

Analysis of Well-Being between indapamide and captopril.
In studies evaluating sexual impairment induced by antihypertensive treatment in male patients, the answers given to nurses, by patients themselves and by their spouses were quite different...

<table>
<thead>
<tr>
<th>Rate of sexual dysfunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurses</td>
</tr>
<tr>
<td>Patients</td>
</tr>
<tr>
<td>Patients (palm pilot)</td>
</tr>
<tr>
<td>Spouses</td>
</tr>
</tbody>
</table>
To follow the rigorous procedures of development of HRQL or PRO questionnaires:

- Item generation
- Scaling
- Item reduction
- Reproductibility
- Content validity
- Construct validity
- Discriminant validity
- Convergent validity
- Responsiveness
- Cultural adaptation

**Multiple causes**
- Lack of rest or exercise
- Improper or inadequate diet
- Psychological stress (depression, anxiety)
- Use of recreational substances
- Anemia
- Abnormalities of the thyroid gland and hypogonadism
- Infections
- Side effects of medications
- Sleep disturbances
- Fever

**Fatigue description**
- Lack of energy
- Sleepiness
- Tiredness
- Exhaustion
- Inability to get enough rest
- Weakness

Specific fatigue questionnaire

HRQL questionnaire: must have items related to fatigue

Various factors involved in the multidimensional HRQL construct

- Control of disease
- 2nd illness
- Social support
- Personality traits

Diabetes

Understanding
Diet
Cause
Diabetes burden
Positive attitude
Coping with disease
Negative impact

Items about DIET can express different concepts
Input of patients in item generation is critical

**Diabetes** --> Cause --> **Food** --> consequence --> **DIET**

I am able to keep my diet regimen under control

**Control of disease / self-management**

My diabetes and its treatment (e.g. diet) keeps me going out with friends / to restaurant / as much as I want

**Interference with social and personal relationships**

I find it hard to do all the things (e.g. diet) I have to do for my diabetes

**Coping with disease**

Discriminant validity of the Functional Digestive Disorders Quality of Life questionnaire (FDDQL)

International study: France, Germany, Great Britain
391 IBS and dyspeptic patients

Score values (m ± SD) of the global SSc HAQ and HAQ-DI, according to the number of the following organ involvements (n=6):
- Raynaud’s phenomenon
- Digital ulcers
- Gastro-intestinal
- Pulmonary
- Musculoskeletal
- Hand contracture

Comparison using ANOVA (p < 0.0001 for both scores) (n=100 patients)

HAQ-DI: Health Assessment Questionnaire – Disability Index;
Global SSc HAQ = (8 HAQ-DI domains + 5 VAS)/13.

Validation of French version of the scleroderma health assessment questionnaire (SSc HAQ).
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 »

- Canada (US)  Shoveling the snow
- Japan  Beat futons
- Norwegian  Going fishing
Responsiveness - generic questionnaires

Psychological General Well-Being (PGWB) & GERD
HRQL is not improved by gastro-esophageal reflux disease drugs?

24% difference in pyrosis relief

No difference in PGWB score

• Specific CFQ-14 developed in France
• Translated in German
• Studies in n = 197 and n = 103 adolescents/adults
• **Construct validity**: same 9 HRQL domains as in the French original CFQ-14
• **Internal consistency**: ranged from 0.71 to 0.94
• **Clinical validity**: supported by severely ill patients reporting lower HRQL than less ill patients

Study Design: specific issues related to HRQL / PRO measure

- **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

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 generally lead to a higher rate of positive results. No causal link can be established between the therapeutic intervention and the HRQL change.

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 a 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?

French Drug Approval (200)
• HSQ (Health Status Questionnaire)
• before / after scores on 1300 patients
• All p values < 0.0001
• Conclusion: all HRQL domains were significantly different across treatment groups

• Problem: 1300 provide 80% power to detect a change of 1 unit on a 0-100 point scale

JCO 2001 (anonymous)
Statistical analysis plan: Estimating the adequate sample size

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

<table>
<thead>
<tr>
<th>SF-36</th>
<th>Ran</th>
<th>Pla</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>82.6</td>
<td>80.0</td>
<td>2.6</td>
</tr>
<tr>
<td>RP</td>
<td>77.0</td>
<td>74.6</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>73.8</td>
<td>69.1</td>
<td>4.7</td>
</tr>
<tr>
<td>GH</td>
<td>69.7</td>
<td>68.7</td>
<td></td>
</tr>
<tr>
<td>VT</td>
<td>58.0</td>
<td>54.4</td>
<td>3.6</td>
</tr>
<tr>
<td>SF</td>
<td>85.5</td>
<td>83.7</td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>81.9</td>
<td>78.2</td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td>72.5</td>
<td>71.7</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 6. QUALITY OF LIFE CHANGE SCORES (MEAN ± SE) FOR PATIENTS COMPLETING 20 WEEKS OF ACTIVE THERAPY AND PATIENTS WITHDRAWING FROM STUDY

<table>
<thead>
<tr>
<th>Quality-of-Life Scale</th>
<th>Nifedipine GITS Complete (n = 119)</th>
<th>Nifedipine GITS Withdrawn (n = 68)</th>
<th>Atenolol Complete (n = 131)</th>
<th>Atenolol Withdrawn (n = 47)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Psychosocial Score</td>
<td>+0.149 ± 0.03*</td>
<td>-0.156 ± 0.06</td>
<td>+0.020 ± 0.03</td>
<td>-0.063 ± 0.06</td>
</tr>
<tr>
<td>Mental Health Index</td>
<td>+0.134 ± 0.04*</td>
<td>-0.151 ± 0.06</td>
<td>+0.029 ± 0.03</td>
<td>-0.079 ± 0.09</td>
</tr>
<tr>
<td>Psychological Well-Being</td>
<td>+0.195 ± 0.04*</td>
<td>-0.117 ± 0.07</td>
<td>+0.041 ± 0.05</td>
<td>-0.018 ± 0.08</td>
</tr>
<tr>
<td>Psychological Distress</td>
<td>+0.098 ± 0.04*</td>
<td>-0.173 ± 0.06</td>
<td>+0.022 ± 0.03</td>
<td>-0.113 ± 0.10</td>
</tr>
<tr>
<td>General Perceived Health</td>
<td>+0.104 ± 0.04*</td>
<td>-0.148 ± 0.06</td>
<td>+0.016 ± 0.04</td>
<td>-0.084 ± 0.06</td>
</tr>
</tbody>
</table>

* P < .001, significant difference between completers and withdrawals.

Positive (+) sign denotes improvement, negative (−) sign worsening.

N = 365 (394 randomized)

→ When the meaning of range score of different questionnaires is opposite.

<table>
<thead>
<tr>
<th>Table 1. Comparison of geometric mean scores (range limits)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>CRQ</td>
</tr>
<tr>
<td>BPQ</td>
</tr>
</tbody>
</table>

**BPQ** low scores = good QoL; **CRQ** low scores = severely disabled.