Quality of Life
a critical endpoint in clinical trials

Corneel Coens, MSc
EORTC HQ, Brussels, BE

Training Course BHS
SATURDAY, 24 MAY 2014
Regulatory view - QoL

• “Quality of Life — A general concept that implies an evaluation of the effect of all aspects of life on general well-being. Because this term implies the evaluation of nonhealth-related aspects of life, and because the term generally is accepted to mean what the patient thinks it is, it is too general and undefined to be considered appropriate for a medical product claim.” FDA 2006

• “Health Related Quality of Life (HRQoL) is a multidomain concept that represents the patient’s general perception of the effect of illness and treatment on physical, psychological, and social aspects of life.” FDA 2006

• “HRQoL is considered to represent a specific type/subset of PROs, distinguished by its multi-dimensionality. Indeed, HRQL is a broad concept which can be defined as the patient’s subjective perception of the impact of his disease and its treatment(s) on his daily life, physical, psychological and social functioning and well-being.” EMEA 2005
• So what does that mean...

• HRQoL data is specific in nature
  • Umbrella concept – “not a single issue”
    • Multiplicity
    • Selection
  • Self-reported – “source = patient”
    • Missing data
    • Bias
  • Subjective – “depends on expectations versus experiences”
    • How to measure?
    • Interpretation?
Multi-dimensional

- COGNITIVE FUNCTIONING
- EMOTIONAL FUNCTIONING
- PHYSICAL FUNCTIONING
- SOCIAL FUNCTIONING
- ROLE FUNCTIONING

- FATIGUE
- PAIN
- APPETITE LOSS
- DYSPNEA
- CONSTIPATION
- INSOMNIA
- NAUSEA/VOMITING
- DIARRHEA

GLOBAL HEALTH STATUS / HRQoL

Utility

Health-Econ: 0→1
Subjective

- HRQoL = reported by patients
  - Depends on patients expectations
  - Depends on patients experiences
  - Different patient = different result
The disability paradox: high quality of life against all odds

Gary L. Albrecht *, Patrick J. Devlieger

University of Illinois at Chicago, School of Public Health, 2035 West Taylor Street, Chicago, IL 60612, USA

Abstract

This paper builds on the work of Sol Levine to examine a disability paradox: Why do many people with serious and persistent disabilities report that they experience a good or excellent quality of life when to most external observers these individuals seem to live an undesirable daily existence? The paper uses a qualitative approach to

The disability paradox:
Many people with serious disabilities report good or excellent (HR)QoL.
Subjective

- HRQoL = reported by patients
  - Depends on patients expectations
  - Depends on patients experiences
  - Different patient = different result
- HRQoL = reported via questionnaires
  - Different wording
  - Different domains
  - Different questionnaire = different result
Subjective

The wording of the question and the response options influences the answers.

Did you have pain?
How much did pain interfere with daily life?
Did you need pain medication?

Are we measuring health?
Or imagination?
Subjective

- HRQoL = reported by patients
  - Depends on patients expectations
  - Depends on patients experiences
  - Different patient = different result
- HRQoL = reported via questionnaires
  - Different wording
  - Different domains
  - Different questionnaire = different result
- HRQoL = no standard methodology
  - Different analysis
  - Different reporting
  - Different method = different result
Measuring a subjective outcome?

Regulatory perspective: content validity

“Evidence that the instrument measures the concept of interest including evidence from qualitative studies that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use.”

- Consequence:
  - Extensive validation needed.
  - Questionnaires need to be developed with an objective in mind.
Example of validated development

Phase 1: Listing of HRQOL issues
Phase 2: Operationalisation
Phase 3: Pre-testing
Phase 4: Field Testing

Years >4
Patients total N = 700

questionnaire
↑
scales
↑
questions
↑
issues
Validation

• **Content validity**
  • Internal validity
    • Questions related to one concept should agree strongly (but not be redundant).
    • Questions from different concepts should agree less.
  • External validity
    • Compare against another ‘accepted’ standard.

• **Reliability**
  • Test-retest: gives the same result (when it should)?

• **Sensitivity/responsiveness**
  • It can discriminate between patients (between groups).
  • It can discriminate within patients (over time).
EORTC QLQ-C30 (version 3)

We are interested in some things about you and your health. Please answer all of the questions yourself by circling the number that best applies to you. There are no "right" or "wrong" answers. The information that you provide will remain strictly confidential.

Please fill in your initials: ______________________________
Your birthdate (Day, Month, Year): __________________________
Today's date (Day, Month, Year): 31 __________________________

1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?  
   - Not at All 1  
   - A Little 2  
   - Quite a Bit 3  
   - Very Much 4

2. Do you have any trouble taking a long walk?  
   - Not at All 1  
   - A Little 2  
   - Quite a Bit 3  
   - Very Much 4

3. Do you have any trouble taking a short walk outside of the house?  
   - Not at All 1  
   - A Little 2  
   - Quite a Bit 3  
   - Very Much 4

4. Do you need to stay in bed or a chair during the day?  
   - Not at All 1  
   - A Little 2  
   - Quite a Bit 3  
   - Very Much 4

5. Do you need help with eating, dressing, washing yourself or using the toilet?  
   - Not at All 1  
   - A Little 2  
   - Quite a Bit 3  
   - Very Much 4

The Journal of Cancer Therapy
EORTC QLQ-C30 (V3.0)

我们想了解有关您和您的健康的一些情况，请您亲自回答下面所有问题，这里的答案并无“对”与“不对”之分，只要求在最能反映您情况的那个数字上画圈。您所提供的资料我们将会严格保密。

请填上您的代号(编号)：__________

出生日期： ___ 年 ___月 ___日

今天日期： ___ 年 ___月 ___日

1. 您从事一些费力的活动有困难吗，
2. ___________行走对您来说有困难吗？
3. 户外短距离行走对您来说有困难吗？
4. 您白天需要呆在床上或椅子上吗？
5. 您在吃饭、穿衣、洗澡或上厕所时需要他人帮忙吗？
Take home message

Always use a validated instrument!

Data from non-validated or ad-hoc questionnaires might be meaningless.

Regulatory bodies require validation, even too much:
- Translation = new validation
- New class of drugs = new validation
- Change in population = new validation
- ...

EORTC
HRQoL in a cancer clinical trial.
Take home message

There is no single optimal approach for HRQoL data in general!

**Analysis** should depend on

- Trial *objective*!
- Instrument & scales selection.
- Time *schedule* used.
HRQoL objective

- **Adjuvant**
  - Aim is to increase survival
  - can be used as complementary in a benefit risk analysis

- **Advanced**
  - Efficacy gains likely small
  - QoL can inform on the clinical relevance.

- **Palliative**
  - Relief of symptoms, or prevention of symptoms
  - Maintaining or improving QoL
HRQoL objective

“we are interested in the change in HRQoL over time”
NOT good enough.
Not a specific question
Too many possible tests – interpretations

Two possible objectives:

• **Exploratory** – investigate possible trends/differences for future confirmation. No conclusive results.

• **Confirmatory** – specific hypothesis needed: one key question to be solved.
Comparison between ... ?

- Internal
  - Randomized comparison
    - Best safeguard against selection bias
  - Non-randomized comparison
    - What has been done to ensure comparability?

- External controls
  - Historical data: non-randomized comparison
    - Same population
    - Same instrument
    - Same setting
    - ...
Example: EORTC 10921 study

- 448 locally advanced breast cancer (LABC) patients received
  - Cyclophosphamide (75 mg/m\(^2\) orally days 1 to 14)
  - Epirubicin (60 mg/m\(^2\) intravenously [IV] days 1, 8)
  - Fluorouracil (500 mg/m\(^2\) IV days 1, 8)
  - for six cycles every 28 days (six months)
  - Cyclophosphamide (830 mg/m\(^2\) IV day 1)
  - Epirubicin (120 mg/m\(^2\) IV day 1)
  - Filgrastim (5 micro g/kg/d subcutaneously days 2 to 13)
  - for six cycles every 14 days (three months)
- HRQOL was assessed using the EORTC QLQ-C30 at baseline and at months 1, 2, 3, 6, 9, 12, 18, 26, 34, 42, 48 and 54.
- No survival differences were seen between these arms.
Example: EORTC 10921 study

- CEF 6mths: 59.3
- Healthy women: 71.2
- Breast cancer patients: 61.8
- CEF baseline: 65.5
- EC+Gcsf 6mths: 66.3

P-values: P<0.01, P=0.25, P=0.02, P=0.01
Hallmarks of a good objective

What are the expectations?
• What is the baseline state of HRQoL for these patients? Both in scores, as in variation.
• Change over time:
  • What is the nature of the change over time to be expected (disease evolution)?
  • When comparing different treatment arms, where and what difference is expected?
• Compliance issues, long-term vs short-term, salvage/recovery, ...
• These will be used to formulate the assumptions that form the basis for the “number crunching”.

EORTC
The future of cancer therapy
Example: EORTC 10921 study

Global health status / QoL
Means ± 99% CI
Hallmarks of a good objective

What are the **key outcomes**?

HRQoL is a multi-domain concept.

E.g. QLQ-C30 consists of 30 questions that make up 15 scales.

If probability of false positive is 5%, then 15 scales = 53% chance on false positive.

<table>
<thead>
<tr>
<th>Nr of outcomes</th>
<th>% false positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5%</td>
</tr>
<tr>
<td>2</td>
<td>10%</td>
</tr>
<tr>
<td>10</td>
<td>40%</td>
</tr>
<tr>
<td>15</td>
<td>53%</td>
</tr>
<tr>
<td>20</td>
<td>64%</td>
</tr>
<tr>
<td>50</td>
<td>92%</td>
</tr>
</tbody>
</table>

Are all scales relevant?

Select **upfront** your primary endpoints.
Clinical versus statistical significance

When is a difference important?

95% CI around a mean difference of 10 points between two groups, SD=20, ES = 0.5

-20  -10   0   10   20   30   40
Sample size (n1 + n2)

QOL scale (0-100)

p>0.05 ... As n increases, CI decrease, p decreases ... p<0.05 when crosses zero line

The future of cancer therapy
Hallmarks of a good objective

What is the clinical significance?

P-values are a measure of statistical significance, not of clinical relevance. P-values depend on sample size. Was the trial correctly powered for HRQoL?

correctly = not underpowered & not overpowered

Statistical significance should not be the only reference value.

Clinical significance is as important: minimal important difference

“smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management”

Example: EORTC 10921 study

CEF 6mths  | Healthy women  | Breast cancer patients  | CEF baseline  | EC+Gcsf 6mths
---|---|---|---|---
71.2 | 61.8 | 65.5 | 66.3 |

Δ = 11.9  | Δ = 2.5  | Δ = 6.2  | Δ = 7.0 |
P<0.01    | P=0.25   | P=0.02   | P=0.01   |

MID = 5-10
Further design

Instrument selection

- Need for a validated instrument
- Instrument should cover the required domains (objective)
- Minimal clinical significance is specified?
- Practical: administration, cultural validity, previously reported, ...

Time schedule used

- 2 approaches
  - Time-driven: fixed interval from clearly defined timepoint.
    E.g. 6 months from randomization, ...
  - Event-driven: events trigger the assessment
    E.g. end of treatment, start of 4th cycle, ...

- Comparability is the main issue.
- Allow for time windows.
- Should be clearly described in the protocol.
Analysis of HRQoL data.
**Analysis techniques**

- Classical data (survival, progression, ...)
  - One patient = one outcome
- HRQoL data
  - One patient = multiple outcomes over time

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Survival Time</th>
<th>Summarized Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>122 (D)</td>
<td>62</td>
</tr>
<tr>
<td>2</td>
<td>148 (D)</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>67 (D)</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>223 (A)</td>
<td>74</td>
</tr>
</tbody>
</table>
Analysis techniques

- Classical data (survival, progression, ...)
  - One patient = one outcome
- HRQoL data
  - One patient = multiple outcomes over time

Two main approaches for the actual analysis:

- Summary statistics
- Longitudinal modelling
HRQoL analysis: Summary statistics

1. **Summary statistics** = summarize data per patient into single number (~ clinical data)
   - Average HRQoL (duration)
   - Highest/lowest reported score over time period. (extremes)
   - AUC (Area Under the Curve). (combines duration + magnitude)
   - Proportion of patients with 10-point de/increase. (responders)
   - Time until certain event. (survival-type approach)
   - ...

Choice of method depends on **objective** and **instrument**.

Uses only partial information & often subject to handle missing data!
### Example: EORTC 10921 study

<table>
<thead>
<tr>
<th>Summary statistic</th>
<th>CEF</th>
<th>EC+G-CSF</th>
<th>Treatment difference P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference at 6m</td>
<td>58.1 (SE=1.83)</td>
<td>66.3 (SE=1.83)</td>
<td>0.002</td>
</tr>
<tr>
<td>Difference at 12m</td>
<td>65.6 (SE=2.04)</td>
<td>62.6 (SE=1.97)</td>
<td>0.301</td>
</tr>
<tr>
<td>Average until 12 m</td>
<td>54.5 (SE=1.26)</td>
<td>52.6 (SE=1.14)</td>
<td>0.275</td>
</tr>
<tr>
<td>Minimum until 12m</td>
<td>38.6 (SE=1.40)</td>
<td>32.9 (SE=1.36)</td>
<td>0.003</td>
</tr>
<tr>
<td>Recovery (%)</td>
<td>64.8% (SE=4.6)</td>
<td>64.0% (SE=4.5)</td>
<td>0.911</td>
</tr>
</tbody>
</table>

* Recovery = at least 10 decrease during the first 12 months compared to baseline but less than 10 points decreased at 12 months.

**Optimal method?**  
Depends on objective!

* Use MID as part of the summary.
The ‘time’ relationship

HRQoL score

Time
Possibility 1

HRQoL score

Time
The ‘time’ relationship

HRQoL score vs. Time
Longitudinal modeling

What is modeling?

• Describe an outcome as a formula of known factors.
• Example:

\[
[\degree C] = ([\degree F] - 32) \times \frac{5}{9}.
\]

• But formulas are not perfect – add error term.
  • Systolic blood pressure = 0.006 age\(^2\) - 0.02 age + 120.
  • But not all 50 years old have 0.006*2500 – 0.02*50 +120 = 134 SBP
  • SBP = 0.006 age\(^2\) - 0.02 age + 120 + \(\varepsilon\)
Longitudinal modeling (2)

• HRQoL can be expressed in a formula

\[ QL = x_0 + x_1 \cdot \text{age} + x_2 \cdot \text{inflammatory} + x_3 \cdot \text{time} + x_9 \cdot \text{Trt} + \varepsilon \]

If applied to EORTC 10921:

<table>
<thead>
<tr>
<th>( x_0 ) = 73.9</th>
<th>( x_1 ) = -0.11</th>
<th>( x_2 ) = -2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_3 ) = 1.98</td>
<td>( x_4 ) = -11.9</td>
<td>( x_5 ) = -17.5</td>
</tr>
<tr>
<td>( x_6 ) = -17.7</td>
<td>( x_7 ) = -0.95</td>
<td>( x_8 ) = 2.33</td>
</tr>
</tbody>
</table>

Specific HRQoL assessment:

Patient (51 years old, no inflammatory) on EC+G arm scored 50 at 3 months
Longitudinal modeling (2)

- HRQoL can be expressed in a formula

\[ QL = x_0 + x_1 \cdot \text{age} + x_2 \cdot \text{inflammatory} + x_3 \cdot \text{time} + x_9 \cdot \text{Trt} + \epsilon \]

If applied to EORTC 10921:

<table>
<thead>
<tr>
<th>( x_0 )</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</th>
<th>( x_4 )</th>
<th>( x_5 )</th>
<th>( x_6 )</th>
<th>( x_7 )</th>
<th>( x_8 )</th>
<th>( x_9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>73.9</td>
<td>-0.11</td>
<td>-2.5</td>
<td>1.98</td>
<td>-11.9</td>
<td>-17.5</td>
<td>-17.7</td>
<td>-0.95</td>
<td>2.33</td>
<td>-2.1</td>
</tr>
</tbody>
</table>

Specific HRQoL assessment:

\[ 50 = 73.9 - 0.11 \times 51 + 0 - 17.7 - 2.1 (\text{+1.62}) \]
And further...

Modeling is flexible!

- If something is deemed of importance, it can be added to the model.
- Example: HRQoL before or after progression.

\[ QL = x_0 + x_1.\text{age} + x_2.\text{inflammatory} + x_{3-8}.\text{time} + x_9.\text{Trt} + x_{10}.I_p + \epsilon \]

where \( I_p \) is a progression status indicator (time dependent).

But there is a price to pay...

- Extra parameters to estimate.
- INTERPRETATION?
Missing data

• Specific characteristic of HRQoL data is ...
  
  **MISSING DATA**

• Self-reported outcome
  • Patient cannot be forced to reply
  • Patient cannot always be reached
  • Retrospective data gathering is not feasible
    High(er) proportion of missing data.

• Is this a problem?
Missing data

Missing data leads to

• **Loss of power**
  • Less patients = less chance to detect difference
  • Sample size issue
  • Solution: add more patients

• **Bias**
  • Characteristics of patients who do not reply ≠ those who do reply.
  • Interpretation issue
  • Solution: ? (adding extra patients will not help)
Example: EORTC 10921

Global health status / QoL
Means + SD

EC+G arm only

Extra dropout added related to the actual value.
Missing data

How can missing data be handled?

• Ignore it
• Adapt the model
  • ... but we don’t have the data.
  • Extended techniques do exist (Pattern mixture models, Marginal models, ...) – assume MAR, not MNAR.
• Replace missing data
  • Imputational techniques
• Choice of endpoint
  • Summary statistic (?)
Imputational techniques

Imputation of missing data.

imputation = replace missing value with ‘best guess’ and then run an analysis on the full dataset.

- Use information from
  - other “similar” patients
  - values from other assessments by the same patient
  - or a mixture of both

- Imputation of missing data can be acceptable but one should understand its limitations.
- Imputational methods assume the data is missing at random (often violated assumption).
- Most imputational methods lead to underestimation of variance.
<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## LOCF

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
Example: EORTC 10921

Compliance in 10921: 90% at baseline; dropping to 60% during follow-up

**High**: impute maximum score (100)

**Low**: impute minimum score (0)
Missing data

So ... how can missing data be handled?

**SENSITIVITY ANALYSES**

- Check if the results are robust under different scenarios:
  - Alternative endpoints
  - Alternative populations (subgroups)
  - Alternative models/methodology
  - Imputational techniques – “what if scenarios”
- **Sensitivity analyses support** the main analyses, they never replace it.
- Main analysis should **always** be pre-specified. Sensitivity analyses **as much as possible**.
Missing data

Conclusion

• ‘Best’ method will depend on ...
  • Causes of missing data
  • Patterns of missing data
  • Availability of ancillary information
  • Available methods (field changing rapidly)

• Choice should depend on desirable characteristics of missing data analyses with regard to the objective.

• Prevention is best
Interpretation

So what does it all mean?
Interpretation

Problems with interpretation

• Missing data can cause bias.
  • Very difficult to assess the extent.
  • Cause of missing data?

• Mean QoL profiles over time.
  • As “worst patients” tend to drop out earlier, average goes up.
  • If two groups (responders + non-responders) = average is meaningless.
Interpretation
Interpretation
Interpretation
Interpretation: response shift

Severe complication causes better QoL?

Selection effect

Complications = better QoL

The future of cancer therapy
Interpretation

• HRQoL = reported by patients
  • Depends on patients expectations
  • Depends on patients experiences
• But... expectations and experiences change during a trial.
• Response shift: change in outcome due to change of reference.
• Can be problematic if different during two arms.
  • Wait-and-see vs immediate treatment
Interpretation

- Minimal important difference
  - Not always well established
  - Often ‘rule of thumb’ (half standard effect size)
  - May differ:
    - Population
    - Treatment
    - Prevalence / severity.

- HRQoL domains are not independent
  - Symptom clusters: synergistic.
    - Fatigue: depression, social functioning, concentration, ...
  - Masking effect: antagonistic.
    - “When your in pain, you’re in pain and nothing else.”
Interpretation

• Blinding
  • Blinding or not changes the expectations of the patient!
  • Blinding erases “hope vs disappointment” but replaces it with anxiety.
  • FDA = blinding.

• Generalizibility
  • HRQoL results = outside clinical trials?
  • Clinical trials = controlled environment
    • Selected patients
    • Selected interventions
    • Expectations!
  • Measuring HRQoL = changing HRQoL?
HRQoL as endpoint

The ‘value’ of HRQoL
HRQoL as efficacy endpoint

Efficacy

- HRQoL as surrogate for patient benefit.
  - Useful if
    - OS benefit small or negligible.
    - PFS or response not reliable/assessable.
- FDA and EMA guidelines:
  - HRQoL only acceptable if first the efficacy (and toxicity) is documented.
  - So in practice: HRQoL = secondary endpoint.
HRQoL as efficacy endpoint

- QALY:
  - QoL Adjusted Life Years.
  - Combine quantity with quality of life.
  - 1 year 100% HRQoL = 2 years 50% HRQoL.
  - Derived from health-economics: utilities.

- QTWiST:
  - QoL Time Without Symptoms/Toxicity.
  - Time until event: death, toxicity, HRQoL deterioration.
  - Flexible definition: aims to emulate ‘treatment benefit’.

Composite endpoint:
  - Are all events equal?
  - Averages: effects can be hidden.
HRQoL as toxicity endpoint

Toxicity

• HRQoL as a measure of adverse events.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>EXAMPLE</th>
<th>DATA SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laboratory</td>
<td>Neutropenia</td>
<td>Laboratory report</td>
</tr>
<tr>
<td>Observable</td>
<td>Retinal tear</td>
<td>Clinical staff</td>
</tr>
<tr>
<td>Symptoms</td>
<td>Nausea</td>
<td>Clinical staff vs. patients</td>
</tr>
</tbody>
</table>

• CTCAE not necessarily reliable.
  • Low interrater agreement (Atkinson & Basch: SBM, 2010)

• Clinicians tend to underestimate.
• Patients reporting more accurate?
Clinicians tend to systematically underestimate CTCAE symptoms

Basch: NEJM, 2010
Reporting flow: CTCAE

Patient Experiences Symptom

Clinician Interprets Symptom

Clinician interviews patient at visit

Clinician writes in chart

Chart Representation of Symptom

Data Manager Interpretation of Symptom

Data manager abstracts chart

Manual data entry

Research Database

EORTC
Reporting flow: PRO

Patient Experiences
Symptom

Patient direct reporting

Research Database
# HRQoL as toxicity endpoint

<table>
<thead>
<tr>
<th></th>
<th>Toxicities/AEs</th>
<th>HRQoL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame of reference for magnitude</td>
<td>Worst magnitude</td>
<td>Average magnitude</td>
</tr>
<tr>
<td>Breadth of assessed symptoms</td>
<td>Broader, to allow for reporting of any potential adverse symptom</td>
<td>Narrower, encompassing common/important symptoms</td>
</tr>
<tr>
<td>Frequency of administration</td>
<td>Frequent, to cover all time points (frequency = recall)</td>
<td>Periodic, at key time points with recall period.</td>
</tr>
<tr>
<td>Ability to enter unsolicited symptoms</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Clinical trial outcomes being measured</td>
<td>Tolerability; safety</td>
<td>Treatment benefit</td>
</tr>
</tbody>
</table>
HRQoL as toxicity endpoint

• Contexts
  • Early-phase trials to identify AE signals/DLTs; dose-finding
  • Controlled trials to compare tolerability profiles
  • Late toxicity detection
  • Routine cancer care for toxicity monitoring/management

• Benefits
  • More efficient and comprehensive AE data collection
  • Better understanding of patient toxicity experience

• Drawbacks
  • No unsolicited symptoms
  • Missing data
  • Comparability and generalizibility
HRQoL as stand-alone endpoint

• HRQoL as direct measure of treatment benefit
• Needs context of
  • Efficacy
  • Toxicity
• Advised approach:
  • Categorize patients as responders.
  • *a priori responder definition*: the individual patient PRO score change over a predetermined time period that should be interpreted as a treatment benefit. Can incorporate non-PRO elements (survival, compliance,...)
• FDA: one step further
  • Comparison of Cumulative Distribution Functions of responder categories
Conclusion

Use of HRQoL as an endpoint:

Registrational trials:
- Replace efficacy: NO
- Replace toxicity: NO
- Additional endpoint: YES
  - 2006-2010: FDA 93 products that included PRO endpoints (of 432); 8 with documented HRQoL treatment benefits.

Academic trials:
- HRQoL feasible as endpoint: YES, but...
  - Missing data.
  - Analytical issues.
  - Interpretational issues.
- HRQoL needs clear objective (a priori design)

HRQoL needs the same approach as ‘hard’ endpoints
The future of HRQoL

Web Server

Study management

Database Server

Backup Server

Web-based Monitoring

EORTC

The future of cancer therapy
We are interested in some things about you and your health. Please answer all of the questions yourself by clicking on the answer that best applies to you. There are no "right" or "wrong" answers. The information that you provide will remain strictly confidential.

1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?  
   - Not at all  
   - A little  
   - Quite a bit  
   - Very much  

2. Do you have any trouble taking a long walk?  
   - Not at all  
   - A little  
   - Quite a bit  
   - Very much  

3. Do you have any trouble taking a short walk outside of the house?  
   - Not at all  
   - A little  
   - Quite a bit  
   - Very much  

4. Do you need to stay in bed or a chair during the day?  
   - Not at all  
   - A little  
   - Quite a bit  
   - Very much  

5. Do you need help with eating, dressing, washing yourself or using the toilet?  
   - Not at all  
   - A little  
   - Quite a bit  
   - Very much
### Example QLQ-C30: Physical functioning scale

<table>
<thead>
<tr>
<th></th>
<th>Question</th>
<th>Not at All</th>
<th>A Little</th>
<th>Quite a Bit</th>
<th>Very Much</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Do you have any trouble taking a <em>long</em> walk?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Do you have any trouble taking a <em>short</em> walk outside of the house?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Do you need to stay in bed or a chair during the day?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Do you need help with eating, dressing, washing yourself or using the toilet?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Do you have any trouble taking a short walk?

- Not at all
- Very much

Do you have any trouble taking a long walk?

Do you need help dressing?

- List of 56 possible items
- Next question automatically selected based on previous replies.
- Each step maximizes the information gained
- Stop when predefined degree of precision is reached
The future of HRQoL

Targeted therapies = new challenges.

- Specific issues:
  - New symptoms such as rash.
  - Issues with treatment compliance

- Requires new questionnaires:
  - Validation process under new guidelines.

- Long term medication
  - Long term frequent follow-up: not feasible.
  - Recall period.
  - Interference with other diseases.
Challenges

• Validation of new instruments
  • Need for specific instruments
    • Cancer classifications
    • Targeted therapies
  • Validation process: need for faster
    • Recycling: item bank
• Computer era
  • conversion of existing tools
  • New possibilities
    • CAT
    • Cognitive tests
Challenges

• **Translations**
  • Online questionnaires
    • Feedback
  • Cultural adaptations
    • “Where do we draw the line?”

• **Proxy assessments**
  • Under investigation
  • Online adaptive format
Challenges

• **Blinding**
  - Blinding or not changes the expectations of the patient!
  - Blinding erases “hope vs disappointment” but replaces it with anxiety.
  - FDA = blinding.

• **Generalizability**
  - HRQoL results = outside clinical trials?
  - Clinical trials = controlled environment
    - Selected patients
    - Selected interventions
    - Expectations!
  - Measuring HRQoL = changing HRQoL?
Challenges

• **Publication**

Level of evidence and methodology required for QoL assessment is currently unclear.

No uniform guidelines for QoL publications.
• Required vs encouraged vs optional.
• Level of HRQoL/PRO review differs.
• Simultaneous publication QoL and clinical results not always feasible.

Difficult to obtain ‘assumptions’ from literature.
... the end.

- Thank you for your attention

- Further reading
Conclusion

HRQoL in clinical trials:

• Difficult ... but rewarding if you use a validated instrument.
• There is no standard approach to HRQoL design, collection, analysis and interpretation in clinical trials.
• Just as with clinical endpoints, HRQoL objective needs to be a justified and realistic reflection of medical expectations.
... the end.

• Thank you for your attention

• Further reading
Recommended articles


• Patient reported outcomes as endpoints in medical research. Fairclough D. Statistical Methods in Medical Research 2004; 13: 115-138
Imputation

Some examples of possible techniques
<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## LOCF

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
## Mean per patient

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>7.33</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
## Mean per time

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>8.5</td>
<td>6</td>
<td>4.5</td>
</tr>
</tbody>
</table>
**Closest neighbour**

= mean of adjacent categories

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
Closest profile

= calculate the difference at each time point, impute from most similar patient(s)

eg. Difference between patient 1 and 2 = 1 + 1 + 1 = 3; between patient 1 and 4 = 0 + 2 + 0 = 2; Hence patient 4 resembles patient 2 best.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>
External regression

= use non-HRQoL data to model the HRQoL outcomes. Use this model to predict the missing values.

One can use baseline characteristics such as gender, age, performance status, staging, ... but also time-varying characteristics such as occurrence of AEs, progression date, cumulative CT dose, ...

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WHO 2</td>
<td>7</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Local</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>WHO 3</td>
<td>8</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Local</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>WHO 3</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Metast</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Terminology
PRO vs HRQoL

- **PRO** = patient reported outcome.
- **HRQoL** = health-related quality of life

<table>
<thead>
<tr>
<th>HRQoL</th>
<th>PRO</th>
</tr>
</thead>
</table>
| Yes   | Yes: Questionnaires  
| No    | Yes: Treatment compliance diaries  
|       | No: Clinical data       |