An introduction to Quality of Life in cancer clinical trials

Corneel Coens, MSc
EORTC HQ, Brussels, BE
Voting question 1:

How would you rate your current quality of life?

- 1=Worst imaginable
- 2=Not so good
- 3=Not so bad
- 4=Best imaginable

The future of cancer therapy
Naming convention

- **PRO** = patient reported outcome (HOW you measure)
- **HRQoL** = health-related quality of life (WHAT you measure)

<table>
<thead>
<tr>
<th>HRQoL</th>
<th>PRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

“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
• 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
ROLE FUNCTIONING
SOCIAL FUNCTIONING

FATIGUE
PAIN
APPETITE LOSS
DYSPNEA
CONSTIPATION
INSOMNIA
NAUSEA/VOMITING
DIARRHEA

HRQoL

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?
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
Example of validated development

Phase 1 - Listing of HRQOL issues

Phase 2 - Operationalisation

Phase 3 - Pre-testing

Phase 4 - Field Testing

 questionnaires
  scales
  questions
  issues

Patients total N = 700

Years >4

VALID – RELIABLE - SENSITIVE
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?  1  2  3  4

2. Do you have any trouble taking a long walk?  1  2  3  4

3. Do you have any trouble taking a short walk outside of the house?  1  2  3  4

4. Do you need to stay in bed or a chair during the day?  1  2  3  4

5. Do you need help with eating, dressing, washing yourself or using the toilet?  1  2  3  4
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

<table>
<thead>
<tr>
<th></th>
<th>没有</th>
<th>有点</th>
<th>相当</th>
<th>非常</th>
</tr>
</thead>
<tbody>
<tr>
<td>1．您从事一些费力的活动有困难吗？</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2．长短距离行走对您来说有困难吗？</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3．户外短距离行走对您来说有困难吗？</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4．您白天需要躺在床上或椅子上吗？</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5．您在吃饭、穿衣、洗澡或上厕所时需要他人帮忙吗？</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
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
• ...
Implementing HRQoL in a cancer clinical trial.
Take home message

There is no single optimal analysis technique for HRQoL data in general!

Analysis plan should depend on

- Trial objective!
- Instrument & scales selection.
- Time schedule used.
Why do we measure HRQoL in clinical trials?

• Clinical trials have different objectives
  • Drug development phase: Phase I – II – III – IV
  • Superiority vs non-inferiority vs selection
  • Registrational vs academic
  • ...

→ HRQoL rationale can have different objectives

• Three broad types of HRQoL objectives:
  • Toxicity  – symptoms & AE
  • Efficacy  – functional status
  • Utility    – generic health perception
Example: 18071 trial

- Main eligibility criteria:
  - Complete and adequate resection of stage III melanoma
  - Histologically confirmed melanoma metastatic to lymph node
  - Stage IIIA (if N1a, at least 1 metastasis >1 mm); stage IIIB or IIIC (no in-transit metastasis)
  - No prior systemic therapy for melanoma
  - Documented disease-free following surgery

- Treatment arms:
  - Ipilimumab at 10 mg/kg q3 wks for 4 doses, then q3 months for up to 3 years (Ipi).
  - Placebo q3 wks for 4 doses, then q3 months for up to 3 years (Pbo).

- Main clinical results:
  - Ipi led to a significant improvement in recurrence free survival (RFS).
    - Median RFS
      - Pbo: 17.1 months
      - Ipi: 26.1 months
    - Hazard ratio (95% CI)
      - Pbo: 0.75 (0.64–0.90)

- Safety profile is generally consistent with that observed in advanced melanoma, although the incidence of some immune-related adverse events (irAEs), e.g., endocrinopathies, were higher in this study.

<table>
<thead>
<tr>
<th></th>
<th>Pbo</th>
<th>Ipi</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade 3-4 IrAE</strong></td>
<td>2.5%.</td>
<td>42.0%</td>
</tr>
<tr>
<td><strong>Grade 3-4 Gastrointestinal</strong></td>
<td>0.8%</td>
<td>15.9%</td>
</tr>
<tr>
<td><strong>Grade 3-4 Hepatic</strong></td>
<td>0.2%</td>
<td>10.6%</td>
</tr>
</tbody>
</table>

Voting question 2:

In this trial, the best use of quality of life is:

• 1 = Toxicity endpoint (because Ipi causes severe AEs)
• 2 = Efficacy endpoint (Ipi impacts on health/functioning)
• 3 = Utility endpoint (HTA: Ipi is expensive)
• 4 = do not include (no added value)
HRQoL as toxicity endpoint

• 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</td>
</tr>
</tbody>
</table>

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

• Clinicians tend to underestimate.

• Patients reporting more accurate?

The future of cancer therapy
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
- **Research Database**
  - Manual data entry
Reporting flow: PRO

Patient Experiences
Symptom

Patient direct reporting

Research Database
HRQoL as toxicity endpoint

- Interest = patient reported symptoms to complement AE.
- Analysis:
  - Similar to CTC reporting: worst grade, % serious symptoms
  - No formal testing
- Limitations:
  - Less frequent than CTC
  - Unsolicited symptoms?
  - Worst ↔ overall experience
  - Missing data ?!
- Specific tools exist: PROMIS, PRO-CTCAE, ...
## HRQoL as toxicity endpoint

<table>
<thead>
<tr>
<th></th>
<th>CTC</th>
<th></th>
<th>QLQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ipilimumab (N = 471)</td>
<td>Placebo (N = 474)</td>
<td>Ipilimumab (N = 471)</td>
<td>Placebo (N = 474)</td>
</tr>
<tr>
<td>Grade 3</td>
<td>Grade 4</td>
<td>Grade 3</td>
<td>Grade 4</td>
<td>50-75</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>9.8</td>
<td>0</td>
<td>1.9</td>
<td>0</td>
</tr>
<tr>
<td>Nausea</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vomiting</td>
<td>0.4</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>Fatigue</td>
<td>1.9</td>
<td>0.2</td>
<td>1.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Headache</td>
<td>0.8</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>App. loss</td>
<td>0.2</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
</tr>
</tbody>
</table>

- CTC Nausea & CTC Vomiting compared with QLQ Nausea/vomiting scale
- CTC Headache compared with QLQ Pain (general)
HRQoL as toxicity endpoint

A problem is a problem when it is problematic!

Clinician: “Does it interfere with the treatment”
  • Eg. hematological values

Patient: “Does it interfere with my life”
  • Eg. alopecia
HRQoL as efficacy endpoint

• HRQoL as measure of patient benefit.
• Focus on patient functioning and clinical relevance
  • Requires selection of scales
  • Requires good MIDs
• Especially useful when:
  • Efficacy benefit small or not reliable.
  • Toxicity profile difference.

• Statistical approach = mimic efficacy analysis
• Requires a specific objective (similar to efficacy endpoint)

“we are interested in the change in HRQoL over time”
NOT good enough.

Two possible objectives:
• Exploratory – investigate possible trends/differences for future confirmation. No conclusive results.
• Confirmatory – specific hypothesis needed: one key question to be solved
HRQoL objective

• HRQoL objective(s) must be stated in specific terms.

• It needs to specify:
  • Type of effect:
    • Treatment effect: randomized comparison, controls
    • Time effect: cohort, single arm
  • Size and scope of the effect
    • Magnitude and timing/duration
    • Clinical relevance
  • Key domains (issues) – scale selection?
    • Remember: multi-dimensional
Example: EORTC 18071 study

• EORTC QLQ-C30 was administered at baseline, week 4, week 7, week 10, week 24, and every 12 weeks thereafter up to week 108.

• Comparisons:
  • (within arm) Baseline (Ipi) vs week 24 (Ipi)
    • Context?
  • (between arms) week 24 (Ipi) vs week 24 (Pbo)
    • Randomized comparison?
  • (external) week 24 (Ipi) vs normative data
    • Comparable?
## Example: EORTC 18071 study

<table>
<thead>
<tr>
<th></th>
<th>Week 10</th>
<th>95% CI Diff</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ipi – baseline</td>
<td>78.4 (18.8)</td>
<td>-8.9 [-11.6 to -6.2]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Pbo - week 10</td>
<td>77.4 (19.7)</td>
<td>-7.9 [-10.6 to -5.1]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>General melanoma</td>
<td>68.7 (21.5)</td>
<td>0.8 [-2.1 to 3.6]</td>
<td>0.582</td>
</tr>
<tr>
<td>Healthy</td>
<td>75.0 (19.6)</td>
<td>-5.5 [-8.2 to -2.7]</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
HRQoL 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.

Are all scales relevant?

Select upfront your primary endpoints.
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 18071 study

- †: clinical relevant treatment difference
- ○: clinical relevant change from baseline
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 18071 study

Voting question 3:

What is the best summary statistic for global QoL in this study?

1 = average during induction
2 = average after induction
3 = change from baseline at week 24
4 = % of patients who experience a 10 point or worse deterioration
5 = maximum during induction
6 = minimum during induction
7 = Time until QoL deterioration (score < 50)
<table>
<thead>
<tr>
<th>Global Health/QoL</th>
<th>Treatment arm</th>
<th>Diff. in Means (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pbo (N=476)</td>
<td>Ipi (N=475)</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADI</td>
<td>0.24</td>
<td>-5.44</td>
<td>5.6 (3.7 – 7.7)</td>
</tr>
<tr>
<td>AAI</td>
<td>-0.47</td>
<td>-6.30</td>
<td>5.8 (3.2 – 8.4)</td>
</tr>
<tr>
<td>Change to week 24</td>
<td>1.37</td>
<td>-4.33</td>
<td>5.7 (2.7 – 8.7)</td>
</tr>
<tr>
<td>% &gt;10 point ↓</td>
<td>45.8%</td>
<td>63.6%</td>
<td>17.8% (11.2 - 24.4)</td>
</tr>
<tr>
<td>Max DI</td>
<td>82.6</td>
<td>79.9</td>
<td>2.9 (0.8 - 5.1)</td>
</tr>
<tr>
<td>Min DI</td>
<td>71.1</td>
<td>65.6</td>
<td>5.5 (2.7 - 8.3)</td>
</tr>
<tr>
<td>Time to QL&lt;50</td>
<td>98.8 wks</td>
<td>97.1 wks</td>
<td>HR= 1.01 (0.87 – 1.18)</td>
</tr>
</tbody>
</table>

* ADI = average change during induction; AAI = average change after induction; SD = standard deviation.

Optimal method? Depends on objective!
The ‘time’ relationship

HRQoL score

Time

The future of cancer therapy
The ‘time’ relationship

HRQoL score vs. Time
Longitudinal modeling

What is modeling?

• Describe an outcome as a formula of known factors.
  
• Example:
  
  \[
  \text{[°C]} = (\text{[°F]} - 32) \times \frac{5}{9}.
  \]

• 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\(\times\)2500 – 0.02\(\times\)50 +120 = 134 SBP
  
  • SBP = 0.006 age\(^2\) - 0.02 age + 120 + \(\varepsilon\)
Longitudinal modeling (2)

- HRQoL can be expressed in a formula

\[ QL = x_0 + x_1 \cdot \text{gender} + x_2 \cdot \text{stage} + x_{3-8} \cdot \text{time} + x_9 \cdot \text{Trt} + \varepsilon \]

If applied to EORTC 18071:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_0)</td>
<td>62.1</td>
</tr>
<tr>
<td>(x_1)</td>
<td>4.1</td>
</tr>
<tr>
<td>(x_2)</td>
<td>-1.4</td>
</tr>
<tr>
<td>(x_3)</td>
<td>3.0</td>
</tr>
<tr>
<td>(x_4)</td>
<td>2.5</td>
</tr>
<tr>
<td>(x_5)</td>
<td>-0.7</td>
</tr>
<tr>
<td>(x_6)</td>
<td>-1.5</td>
</tr>
<tr>
<td>(x_7)</td>
<td>1.2</td>
</tr>
<tr>
<td>(x_8)</td>
<td>1.0</td>
</tr>
<tr>
<td>(x_9)</td>
<td>-2.1</td>
</tr>
</tbody>
</table>

Specific HRQoL assessment:

Patient (male, stage II) on Ipi arm scored 60 at week 24
Longitudinal modeling (2)

- HRQoL can be expressed in a formula

$$QL = x_0 + x_1 \cdot \text{gender} + x_2 \cdot \text{stage} + x_3 \cdot \text{time} + x_9 \cdot \text{Trt} + \varepsilon$$

If applied to EORTC 18071:

<table>
<thead>
<tr>
<th>$x_0$</th>
<th>$x_1$</th>
<th>$x_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.1</td>
<td>4.1</td>
<td>-1.4</td>
</tr>
<tr>
<td>$x_3$</td>
<td>$x_4$</td>
<td>$x_5$</td>
</tr>
<tr>
<td>3.0</td>
<td>2.5</td>
<td>-0.7</td>
</tr>
<tr>
<td>$x_6$</td>
<td>$x_7$</td>
<td>$x_8$</td>
</tr>
<tr>
<td>-1.5</td>
<td>1.2</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Specific HRQoL assessment:

$$60 = 62.1 + 4.1 - 1.4 - 1.2 - 2.1 (+ 1.5)$$
QoL modeling

How to practically model HRQoL outcomes:
There are 2 main factors that influence a score
• Patient characteristics (between-level)
• Evolution over time (within-level)

• You need a model that can handle ‘repeated’ measures.
  • Correlations within patients
  • Become very complicated fast

All models are wrong; but some are useful!
HRQoL as efficacy endpoint

- Statistical approaches: mimic efficacy analysis
  - Summary statistics:
    - Categorize patients as responders.
    - Time-to-QoL event
    - Integrate in Clinical Benefit Rate
  - Longitudinal modelling

- Problem: content validity and MID.
  - Validated instruments
  - Are established MIDs reliable?
    - Cumulative Distribution Functions (CDF) of responder categories
HRQoL as utility endpoint

- HRQoL as a single outcome – perception of health or HRQoL

- Context
  - Easy metric that is applicable across trials
  - Used in HTA & HE
  - Useful as ‘screening’ endpoint or if prior info is lacking

- Statistical approaches: descriptive OR borrow from HE
  - QALY, QTWIST, ICER, ....

- Instruments:
  - Most classical questionnaires (QLQ-C30, FACT, ...) have an ‘overall’ scale/outcome.
    - Not always very sensitive...
  - Dedicated utility questionnaires: EQ-5D
HRQoL as utility endpoint

- QALY = Quality Adjusted Life Years
  - Combine: (Utility) x (Duration)
  - Eg. 20 mths @ 75% QL + 40 mths @ 25% QL = 15 + 10 = 25 QALmths

Where should the next course location be?
Zeist (NL) vs Honolulu (USA, Hawaii)

<table>
<thead>
<tr>
<th>Location</th>
<th>Duration</th>
<th>Temperature</th>
<th>QALY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeist</td>
<td>6 days</td>
<td>18°C</td>
<td>108</td>
</tr>
<tr>
<td>Honolulu</td>
<td>4 days</td>
<td>29°C</td>
<td>116</td>
</tr>
</tbody>
</table>
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.
HRQoL objective

- HRQoL objective(s) must be prespecified.
  - State your HRQoL objective as specific as possible.
  - Same considerations as clinical endpoints:
    - Choice of endpoint
    - Choice of treatment effect
    - Existing body of evidence

- Further design (instrument, time schedule, administration, analysis, ...) follows from objective(s).

- Practical issues:
  - Is the instrument easily administrated? Is it translated?
  - Multiple questionnaires: avoid overlap. All relevant, each time?
  - Choose assessment times that are feasible?
  - Monitor the QoL data collection during your trial.
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)
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 understands 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>
<tr>
<td>Patient</td>
<td>Time 1</td>
<td>Time 2</td>
<td>Time 3</td>
<td>Time 4</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<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 18071 study
Example: EORTC 18071 - low
Example: EORTC 18071 - high
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**.
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 your objective.

• Prevention is your best bet - Always aim for 100% compliance!
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
BACKUP SLIDES
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

The future of cancer therapy
Example: EORTC 10921 study
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.

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

Some examples of possible techniques
## Original

<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>

The future of cancer therapy
## 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></td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>0</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>