

Workshop on
Quality of Life Assessment in Cancer
Clinical Research and Clinical Practice:
Facts, Fictions and Future Directions

Neil K. Aaronson
The Netherlands Cancer Institute Amsterdam, NL

Australian Health Outcomes Conference
Canberra, April 29, 2008

A note on terms

- Health outcomes
- Health status
- Quality of life
- Health-related quality of life
- Patient-reported outcomes (PROs)
 - Any report from a patient about a health condition and its treatment (US FDA, 2006)

Patient-reported outcomes (PROs)

- Physical and psychosocial functioning
- Symptoms
- Health-related quality of life
- Satisfaction with care

fact

n. a thing known to be true || a statement about something which has occurred. *he got the facts distorted* || (*law*, in certain phrases only) a crime as a matter of fact, in point of fact, the fact of the matter is... (introductory phrases used to emphasize an explanation or confession) to tell you the truth in fact (usually in contradistinction to some supposed state of affairs) in truth, actually (fr. L. *factum*, a thing done)

The New International Webster Dictionary of the English Language, 1995

fiction

n. A literature consisting of invented narrative, esp. the novel and short story || a falsehood (e.g., that there exists a 'man in the street') conventionally accepted as true because it is useful to make the assumption

The New International Webster Dictionary of the English Language, 1995

fact or fiction?

The term "(health-related) quality of life," is well defined and widely understood.

Fact – if you keep things simple

Fiction – if you dig deeper

"Quality of life is a vague and ethereal entity, something that many people talk about, but which nobody clearly knows what to do about." Campbell et al., 1976

"The idea has become a kind of umbrella under which are placed many different indexes dealing with whatever the user wants to focus on." Feinstein, 1987

"Quality of life is an ill-defined term...it means different things to different people, and takes on different meanings according to the area of application." Fayers & Machin, 2000



4 criteria for evaluating clinical effectiveness of chemotherapeutic agents in lung cancer

D.A. Kamofsky et al., *Cancer* 1:634, 1948

- subjective improvement
- objective improvement
- performance status
- length of survival

Subjective improvement

“The patient’s subjective improvement is measured or described in terms of:

- improvement in his mood and attitude
- his general feeling of well-being,
- his activity, appetite, and the alleviation of distressing symptoms such as pain, weakness, and dyspnea.”

WHO definition of health, 1949

“A state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity.”

Key dimensions of quality of life as defined by ASCO, 1995

Physical	Symptoms commonly caused by cancer and the toxicities of treatment
Psychologic	Effects of cancer and its treatment on cognitive function and emotional state
Social	Effects of cancer and its treatment on interpersonal relationships, school, work and recreation



Attributes of QL definitions

- non-specific vs. health-related
- health states (or status) versus personal evaluation of those states (e.g., expectations, discrepancies, preferences, utilities, satisfaction)
- scope of concerns (e.g., limited to “basics or extended to include e.g., spirituality or existential issues)
- polarity of concerns (well-being vs. dysfunction and its resolution)



Does it matter?

- Yes, because the content of QL questionnaires reflects the underlying definition.
- It may be less important in clinical trials, where group comparisons will be internally valid, regardless of the definition used.
- It is more important in comparing results across trials and in descriptive (e.g., prevalence) studies.

Examples of QL definitions

“The difference between the hopes and expectations of the individual and the individual’s present experience.”

Calman, 1987

“The functional effect of an illness and its consequent therapy upon a patient, as perceived by the patient.”

Schipper et al. 1996

Covinsky et al. Am J Med 1999; 106:435-440

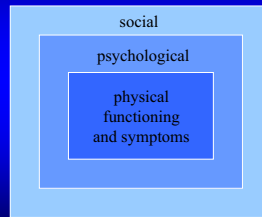
- study of 493 older patients
- QL rated as good/excellent by 43% of those with worst physical functioning and 47% with highest levels of psychological distress
- QL was rated as poor by 15% of those with the best physical functioning and 21% with the lowest levels of psychological distress

fact or fiction?

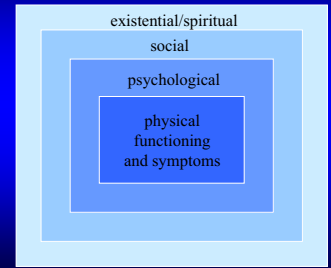
We have (adequate) conceptual models for studying the underlying associations between HRQL domains, and the factors influencing those associations.

reasonably factual

multidimensional HRQL assessment



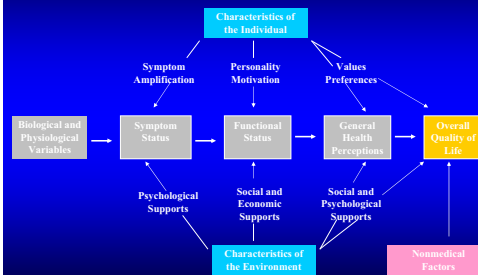
multidimensional HRQL assessment



Advantages of multidimensional approach to QL assessment

- yields specific, clinically relevant information
- permits detection of positive and negative effects
- permits detection of changing patterns over time
- facilitates detection of unexpected effects

Relationships among measures of patient outcome in an HRQL conceptual model (Wilson and Cleary, JAMA 1995; 273(1): 59-65)



Value of conceptual models

- Help clarify issues/concepts/factors in which you are and are *not* interested
- Help make explicit the hypothesized relationship among concepts/factors
- Facilitate hypotheses generation and testing
- Provide structure for statistical analysis strategy (e.g., structural equation modelling)
- Provide leads for possible points of intervention and can help to explain program success or failure

fact or fiction?

The patient is the sole legitimate source of information about his/her HRQL. Other “proxy” raters (e.g., family members, health care providers) are, at best, poor substitutes.

(partial) fiction





Self-report can be limited by:

- age (very young or old)
- cognitive impairment
- symptom distress
- physical disability
- emotional distress
- sociocultural factors affecting communication skills and styles

Exclusion of highly relevant subgroup of patients can result in biased study outcomes



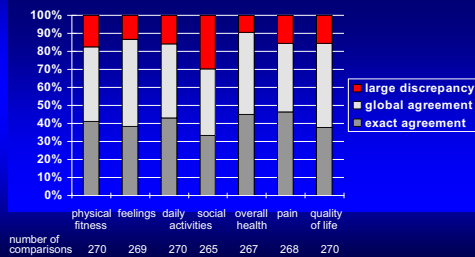


Psychiatrist: How often do you have sex?
 Alvie: Not often enough, I'd say 2 or 3 times a week.
 Annie: Constantly, I'd say 2 or 3 times a week.

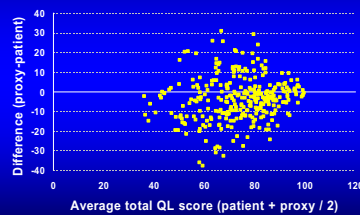
The role of health care providers and significant others in evaluating the QL of patients with chronic disease
 Sneeuw KCA et al. 2002; J Clin Epidemiol 55:1130-43

- 23 studies published between 1991 - 2000
- Moderate/high patient – proxy agreement
- Proxies tended to rate patients as having more problems than did patients themselves
- Magnitude of differences was small (median standardized difference 0.20)

% agreement by HRQL domain (COOP-WONCA charts)



Bland-Altman plot for total HRQL score



Multiple Proxy Perspectives: A conceptual framework

Pickard AS, Knight SJ. Med Care 2005; 43(5):493-9.

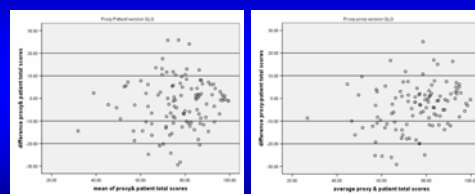
- Proxy-patient viewpoint:
 - Proxy asked to evaluate the patient's HRQL as the patient would assess it himself
- Proxy-proxy viewpoint
 - Proxy asked to evaluate the patient's HRQL from his own perspective
- Judgments from the "proxy-patient" perspective are "...optimally consistent with the patient's view of their HRQL, without embellishment."

Multiple Proxy Perspectives: An empirical test

Gundy C, Aaronson NK. Med Care 2008; 46:209-16

- 224 cancer outpatients undergoing CT and their proxies (typically proxies)
- Both completed the EORTC QLQ-C30
- Proxies randomly assigned to "patient-proxy" or "proxy-proxy" of QLQ-C30
- Bias analysis – group level
- Concordance analysis – individual level

Bland Altman Plots



Proxy-patient version

Proxy-proxy version

Group level bias analysis: QLQ-C30 functional scales

Functional Scales	Proxy-Proxy	Proxy-Patient	Mean between group differences	Effect size
Physical	4.22*	4.78*	-0.56	-0.03
Role	0.00	-2.03	2.03	0.08
Cognitive	-1.07	0.72	-1.79	-0.09
Emotional	8.87*	8.79*	0.08	0.00
Social	5.81*	3.22	2.59	0.07
Global health:QL	6.79*	4.82*	1.97	0.06

Group level bias analysis: QLQ-C30 symptom scales

Symptom scales/items	Proxy-Proxy	Proxy-Patient	Mean between group differences	Effect size
Fatigue	-8.26*	-4.35*	-3.91	-0.20
Emesis	-3.98*	-3.48*	-0.50	-0.03
Pain	-6.57*	-3.62	-2.95	-0.15
Dyspnea	0.93	1.77	-0.84	-0.00
Insomnia	-6.42*	-1.45	-4.97	-0.17
Anorexia	-4.63*	-1.75	-2.88	-0.17
Constipation	-3.81	-3.27	-0.54	-0.12
Diarrhea	0.00	0.59	-0.59	-0.05
Financial	3.06	2.65	0.40	0.09

Concordance analysis at individual level: Intra-class correlations QLQ-C30 functional scales

Scale	Proxy-proxy	Proxy-Patient
Physical	0.80	0.69
Role	0.73	0.59
Cognitive	0.56*	0.34*
Emotional	0.52	0.57
Social	0.44	0.41
Global health/QL	0.55	0.58

Concordance analysis at individual level: Intra-class correlations EORTC QLQ-C30 symptom scales

Scale/item	Proxy-Proxy	Proxy-Patient
Fatigue	0.64	0.69
Emesis	0.67	0.68
Pain	0.73	0.68
Dyspnea	0.45	0.56
Insomnia	0.55	0.49
Anorexia	0.77	0.81
Constipation	0.62	0.64
Diarrhea	0.50	0.67
Financial	0.49	0.47

Multiple Proxy Perspectives: Conclusions

- Despite a “sensible” hypothesis, few empirical differences observed in proxy-patient agreement (at group and individual level) as a function of proxy perspective
- Study needs to be repeated using other HRQL measures and/or patient populations

Discussion

- Is superiority of proxy-patient viewpoint self-evident?
 - Cue visibility
 - Shared schemata
 - Assessing facts vs. Assigning utilities
 - Compounding sources of error

fact or fiction?

Although there are many HRQL questionnaires from which to choose, the dust is settling and a “best bet” can be identified based on a comparison of psychometric characteristics and performance.

fiction

Generic HRQL instruments

- Sickness Impact Profile (SIP)
- Nottingham Health Profile (NHP)
- Spitzer QL Index
- COOP/WONCA Charts
- MOS 36-Item Health Survey (SF-36)
- World Health Organization (WHOQoL)

Cancer-specific HRQL questionnaires

- Functional Living Index – Cancer (FLIC)
- Cancer Rehabilitation Evaluation System (CARES)
- Rotterdam Symptom Checklist (RSCL)
- EORTC QLQ-C30
- Functional Assessment of Cancer Therapy (FACT-G)

Key psychometric attributes of HRQL instruments

- measurement model
- reliability
- validity
- responsiveness
- interpretability
- cultural adaptability
- burden

attributes of 5 cancer-specific QL measures

	FLIC	RSCL	CARES	EORTC	FACT
n of studies	55	30	21	55	19
measurement model					
conceptual	•	•	•	•	•
empirical	•	•	•	•	•
burden	•	•	•	•	•
reliability					
internal consistency	•	•	•	•	•
test-retest	•	•	•	•	•

attributes of 5 cancer-specific QL measures

	FLIC	RSCL	CARES	EORTC	FACT
validity					
content	•	•	•	•	•
construct	•	•	•	•	•
criterion	•	•	•	•	•
predictive	•	•	•	•	•
responsiveness	•	•	•	•	•
interpretability	•	•	•	•	•
cultural adaptations					
translations	•	•	•	•	•
psychometrics	•	•	•	•	•

Choice of HRQL instrument should be driven by:

- the research question(s) to be addressed
- the population under study
- the conceptual basis of candidate questionnaires
- the research context and tradition (if any) in which the study is being conducted
- the specific content and wording of candidate questionnaires

Assessing validity of HRQL instruments: classical approaches (SAC/MOT 2001)

Content-related validity

- evidence that the content domain of an instrument is appropriate relative to its intended use
- the use of lay and expert panel judgments
- complete the questionnaire(s) yourself!

Negative affect items

SF-36	“Have you felt so down in the dumps that nothing could cheer you up?”
	“Have you felt downhearted and blue?”
FACT-G	“I feel sad”
QLQ-C30	“Did you feel depressed?”

Future perspective items

SF-36	“I expect my health to get worse.”
FACT-G	“I worry about dying.”
CARES-SF	“I worry about whether the cancer will progress.”
QLQ-C30	--

Translations

Involvement of native speakers is mandatory, otherwise spelling or other (minor) mistakes may go unnoticed:

“We take your bags and send them in all directions”
(sign in a Copenhagen airline ticket office)

“Specialist in women and other diseases”
(sign in a physician's office in Rome)

“It is strictly forbidden on black forest camping sites that people of different sex, for instance men and women, live together in one tent unless they are married with each other for that purpose.”
(sign in posted in Germany's Black Forest)

fact or fiction?

Given the plethora of HRQL questionnaires currently available, there is little or no need for continued efforts at instrument development.

fiction

- Condition-specific questionnaires tend to be more sensitive to group differences and responsive to inter- and intra-individual changes over time

supplemental modules/scales

- combine “core” instrument with condition-specific modules/scales
 - EORTC “modules”
 - FACT subscales
 - NCIC symptom checklists

EORTC QL modules

- body image
- breast
- bladder
- brain
- colorectal
- esophageal
- high dose chemo
- leukemia (adult)
- lung
- ophthalmic
- ovarian
- peripheral neuropathy
- prostate
- supportive care

FACT subscales

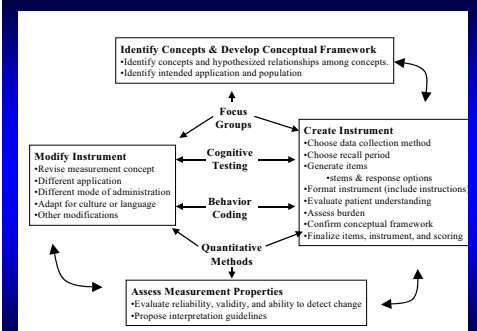
- breast
- bladder
- brain
- cervical
- colorectal
- esophageal
- head & neck
- HIV
- lung
- ovarian
- pancreatic
- prostate
- anorexia
- BMT
- fatigue
- fecal incontinence
- neurotoxicity
- taxane toxicity
- urinary incontinence
- spirituality

Advantages of core + module approach to HRQL assessment

- facilitates comparison of results across studies
- provides sufficient flexibility to address questions specific to a given patient population or treatment

Develop a new instrument at your own risk and only as a last resort

- Labor intensive
- Time consuming
- Long “probationary” period before psychometrics are assessed and verified
- Translation and cultural adaptation process for use in international settings adds layer of complexity



“Modern” Psychometrics

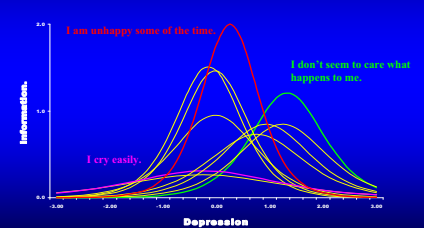
- Item response theory (IRT)
- Item banking
- Dynamic or computer-adaptive testing

Item Response Theory Information Curve



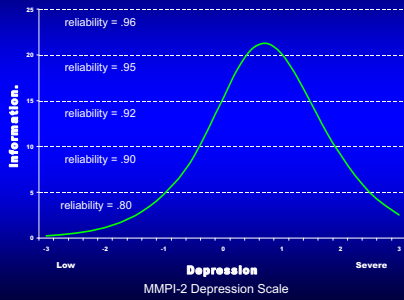
Indicates the range over the measured construct where an item is best at discriminating among individuals. Higher information denotes more precision (or reliability) for measuring a person's trait level.

Item Response Theory (IRT) Modeling Item Information Curves



10 Items from the MMPI-2 Depression Scale

Item Response Theory (IRT) Modeling Scale Information Curve



Item banking and “dynamic” or computer-adaptive testing

- Ask a question from an item bank
- Use the response to this question to estimate the ‘level’
- Select next question based on this knowledge
- Use the response to this question and the previous one to re-estimate the ‘level’ - and so on...
- Stop when the desired level of precision is reached

Implications of IRT for HRQL and other PRO assessments

- Increase measurement precision
- Can vary level of precision dependent on task at hand (group versus individual level)
- Can calibrate scores across existing measures
- Can generate a number of “versions” of a questionnaire
- May eliminate (or at least reduce) issues of proprietary interest

U.S. NIH-NCI Patient-Reported Outcomes Measurement Information System (PROMIS)

- Born out of U.S. National Institutes of Health (NIH) Roadmap Initiative
 - PROMIS I: 2004-2009
 - PROMIS II: 2009-2013
- Goal: To improve assessment of self-reported symptoms and other HRQL domains across many chronic diseases.

PROMIS

- Core domains: pain, fatigue, depression, anxiety, physical function, social function, and overall general health.
- Additional domains: sexual function, cognitive function, sleep/wake function, and illness impact
- Pediatric PROMIS: same core domains as above

PROMIS

Current status, plans and contact

- Core domains currently ready in short form and CAT versions
- Additional domains will be available by late 2008/early 2009
- Resources: <http://www.NIHpromis.org/>
 - PROMIS Assessment Center
 - PROMIS Short Forms
 - Documentation and Supporting Materials

fact or fiction?

The major methodological challenges in HRQL analysis – missing data, multiple comparisons, and clinical interpretation of statistical results – have been resolved or are well on their way to being resolved.

Reasonably factual
(2 out of 3 ain't bad)

Missing data: Items from questionnaires

- Relatively minor problem (less than 5%)
- For multi-item scales, missing responses can be estimated/replaced
- High level of missing values for a given item may signal problem of appropriateness or acceptability

Missing data: Questionnaires

- Missing at random (e.g., administrative failure)
 - Largely avoidable
- Systematic loss to follow-up due to illness or death (“informative censoring”)
 - Often unavoidable (e.g., in advanced disease trials)
 - Complex problem with imperfect but workable solutions
 - Mixed effects ANOVA
 - Growth curve analysis

Multiple comparisons

- Inherent problem with multidimensional HRQL measures (health profiles)
- Results in inflated p values
- Three primary solutions
 - Use summary scores, where available
 - Focus on a few “cardinal” (primary) outcomes
 - Apply statistical adjustments

Defining clinical vs. statistical significance in HRQL scores

“The 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”

(Juniper et al, 1994)

Nature of the problem

- Health-related quality of life (HRQL) questionnaire scores are only sometimes intrinsically interpretable and meaningful
- Typically this is the case with single-item measures that can be scored in terms of percentages

Example 1: pain prevalence

Have you had pain during the past 3 days?

- | | | |
|---------------|-----|--|
| • Not at all | 27% | } 50% of the patients reported having had moderate to severe pain during the past 3 days |
| • A little | 23% | |
| • Quite a bit | 35% | |
| • Very much | 15% | |

Example 2: change in pain over time

- Percentage of patients with moderate to severe pain
 - Baseline: 50%
 - Post-treatment: 20%
 - 6 month follow-up: 35%
- These results are intuitively/intrinsically interpretable to clinicians and patients

Multi-item HRQL outcome measures: EORTC QLQ-C30 fatigue scale

1. Did you need to rest?
2. Have you felt weak?
3. Were you tired?
 - Response categories (1) not at all (2) a little; (3) quite a bit; (4) very much
 - Scale score = linear conversion of the sum of these 3 items to a 0-100 scale (higher score = more symptoms)

QLQ-C30 fatigue scale means and (standard deviations)

- | | |
|----------------------|-------------|
| • Baseline: | 63.8 (20.6) |
| • Post-treatment: | 74.3 (21.4) |
| • 6 month follow-up: | 67.8 (20.9) |
- P values can all be highly statistically significant
 - Neither the baseline score nor the post-treatment or follow-up scores have intrinsic or easily interpretable meaning or significance

Statistical versus clinical significance

- The Achilles heel of statistical significance is that p values are influenced heavily by sample size
- With a small sample, meaningful differences may not be statistically significant
- With a large enough sample, trivial differences can be statistically significant
- Mean of 74.3 vs 63.8 (s.d. = 20)
 - N = 50 per group p > 0.10
 - N = 75 per group p < 0.05
 - N = 100 per group p < 0.005

What is a clinically meaningful HRQL score? c/of Jeff Sloan



Two strategies for estimating the clinical significance HRQL scores

- Distribution-based approach
- Anchor-based approach

Distribution-based approach: effect size

- Reflects the magnitude of difference in scores (over time within a group or between groups) in relation to the degree of variability in the data
- $\text{mean}_1 - \text{mean}_2 / \text{s.d.}$
 - $74.3 - 63.8 / 20 = 0.52$
 - $67.8 - 63.8 / 20 = 0.20$
- Cohen's rule of thumb:
 - 0.2 = small effect
 - 0.5 = moderate effect
 - 0.8 = large effect
- Allows one to compare results across studies

Variations on a theme: Other distribution-based formulas

- Effect size
- Standardized response mean
- Responsiveness statistic
- Paired T statistic
- Relative change
- Standard error of measurement
- $M_1 - M_2 / \text{sd}_{\text{total}}$
- $M_1 - M_2 / \text{sd}_{\text{difference}}$
- $M_1 - M_2 / \text{sd}_{\text{stable group}}$
- $M_1 - M_2 / \text{se}_{\text{difference}}$
- $M_1 - M_2 / M_1$
- $\text{sd}_{\text{test}} \times \sqrt{1 - \text{reliability}_{\text{test}}}$

It looks more complicated than it is. You can always consult your friendly, local statistician.

Anchor-based approaches

- **Population-focused** (multiple) anchors express differences in HRQL scores in terms of their association with and impact on population characteristics
- **Individual-focused** (single) anchors express differences in HRQL scores in terms of patients' perceptions of minimally important changes

Population-focused anchors

- Express differences in terms of their impact on populations, e.g., related to:
 - Functional level
 - Symptoms
 - Disease severity
 - Diagnosis
 - Health care utilization
 - Job loss

Example of population-focused anchors SF-36 physical functioning scale (0-100)

- 32% of those who score 40 on the SF-36 physical function scale can walk 100 meters, as compared with 49% of those who score 50
- 10 point difference represents an absolute change of 17%, a relative increase of 53%; a relative decrease of 35%

Example of population-focused anchors Symptoms

- Scores on the 5-item SF-36 mental health index (MHI-5) in relation to suicidal ideation
 - 29% of those who score 20 on the MHI-5 have suicidal ideation, versus 14% of those who score 40
 - absolute change of 15%; relative increase as MHI-5 score drops of 105%; relative decrease as MHI-5 score rises of 51%

Example of population-focused anchors Diagnosis

- Mean QLQ-C30 Global HRQL scores (0 - 100):
 - heterogeneous localized disease = 74
 - heterogeneous metastatic disease = 54
 - recurrent high-grade gliomas = 60
 - advanced prostate cancer = 45
 - lung cancer, >10% weight loss = 43

Strengths and weaknesses of population-focused anchors

- Strengths
 - Retains underlying complexity
 - Acknowledges that small, medium and large differences may change with populations, even when using the same instrument
 - Availability of multiple anchors allows one to use external standards that are most relevant
- Weaknesses
 - Retains underlying complexity and thus violates the KISS principle

Individual-focused anchors: Within-patient transition ratings

- Patients rate the magnitude of the change that they have experienced over time (e.g., in fatigue)

Much improved Somewhat improved Unchanged Somewhat worse Much worse

- These are compared with pretest and posttest mean scores on the HRQL scale of interest (e.g., QLQ-C30 fatigue scale)
- The mean change in the scale scores of those who report a small but perceptible change over time = minimally important difference (MID)
- A change of 10 or more points on the QLQ-C30 scales = a subjectively significant change = MID

Osoha et al. J Clin Oncol 1998; 16: 139-44

Defining a HRQL “response” and “time to HRQL or symptom progression”

- Use individual-focused anchor (i.e., within-patient transition data) to define:
 - HRQL response as those who improve ≥ 10 points on QLQ-C30 scales
 - Symptom progression as those who deteriorate ≥ 10 points on QLQ-C30 (symptom) scales
- Can vary cutoffs (sensitivity analysis)

Temozolomide vs Procarbazine in recurrent glioblastoma multiforme

Osoba D et al. J Clin Oncol 2000; 18:1481-91

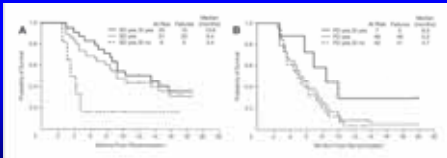
Percentage with ≥ 10 point improvement:

QLQ-C30 and BN-20	TMZ %	PCB %
Role functioning	27	18
Social functioning	32	18
Global QOL	15	18
Visual disorder	34	24
Motor dysfunction	32	18
Communication deficit	40	19
Drowsiness	19	14

Survival of NSCL patients as a function of tumor status and symptom improvement

Cella D et al. J Clin Oncol 2005; 23: 2946-54

Patients were treated with gefitinib (epidermal growth factor receptor tyrosine kinase inhibitor)



Symptom improvement = ≥ 2 point increase in symptom index score for ≥ 28 days without worsening at any interim weekly assessment

Good news:

Converging results using different methods

- Distribution-based** methods indicate that a $\frac{1}{2}$ standard deviation is typically clinically relevant
- Anchor-based** methods suggest the same (e.g., EORTC QLQ-C30; FACT-G; SF-36)
- A 10 point difference on a 100-point scale (a $\frac{1}{2}$ SD) is almost always going to be clinically significant

Summary re: clinical significance

- Statistical significance \neq clinical significance
- Distributional- and anchor-based methods converge to suggest that a $\frac{1}{2}$ SD difference (10 points on a 100 point HRQL scale) is clinically meaningful

fact or fiction?

- Having cutoffs for clinically important differences facilitates defining HRQL “responders”
- These types of data can be used for calculating sample size requirements
- Yield results that are intuitively meaningful and easily interpretable

HRQL data have contributed significantly to the clinical trial process in oncology and other fields of medicine.

Fact, but we can do better

Uses of HRQL information

- Understand the burden of disease
- Screen and stratify patients for treatment
- Demonstrate treatment efficacy
- Facilitate regulatory approval
- Substantiate marketing claims
- Raise consumer and clinician awareness
- Improve patient-provider communication and patient compliance

Uses of HRQL information

- Understand the burden of disease
- Screen and stratify patients for treatment
- Demonstrate treatment efficacy
- Facilitate regulatory approval
- Substantiate marketing claims
- Raise consumer and clinician awareness
- Improve patient-provider communication and patient compliance

HRQL data can yield unanticipated results

Sugarbaker et al. *Surgery* 91:17-23, 1982

- Small RCT (n = 26) in soft-tissue sarcoma:
 - amputation + CT vs. limb-sparing surgery + RT + CT
- HRQL assessed post-surgery
 - SIP, PAIS, Katz ADL, Barthel Index, clinical assessment of mobility, pain, sexuality
- No significant differences between treatment arms, with exception of sexual functioning, which favored amputation group
- Led to improvement in limb-sparing procedure (e.g., better RT shielding)

Clinical trial-based HRQL data in drug approval process

Tannock IF et al. *J Clin Oncol* 1996;14:1756-64.
Osoba D et al. *J Clin Oncol* 1999;17:1654-63

- RCT of daily prednisone ± mitoxantrone (q 3 wks) in metastatic, hormone-resistant prostate cancer (n = 161)
- QLQ-C30, PROSQOLI at baseline and q 3 wks
- Pain improved in 29% of men treated with P+M vs 12 % treated with P alone, and lasted 43 vs 18 wks
- Survival the same in both groups
- P+M ⇒ improvement over time in physical and role function, fatigue, insomnia, drowsiness
- QOL improvements lasted longer in P+M group

Making tradeoffs explicit

Treatment of early stage prostate cancer

Holmberg L et al. *NEJM* 2002; 347:781-789

- RCT of radical prostatectomy vs watchful waiting in early stage prostate cancer (N = 695)
 - 100% follow-up survival (6 yrs), 87% follow-up HRQL (4 yrs)
- Prostatectomy – trend for reduced all-cause mortality (18% versus 15%; p = 0.31)
- Prostatectomy - decrease in prostate-specific death rates (9% versus 5%; RR 0.50, 0.27 to 0.91, p = 0.02)

Making tradeoffs explicit

Steinck G et al. *NEJM* 2002; 347: 7980-796

- Sexual dysfunction
 - 45% waiting; 80% prostatectomy
- Urinary leakage
 - 21% waiting; 49% prostatectomy
- Urinary obstruction (weak stream)
 - 44% waiting; 28% prostatectomy
- Bowel function, anxiety, depression, well-being did not differ

Baseline HRQL predicts survival

Gotay CG et al. *J Clin Oncol* 2008; 26:1355-1363

- Review of 39 studies including ~ 14,000 patients with diverse cancer diagnoses
- Studies employed a range of HRQL measures (single symptom, multidimensional, global QL)
- Analyses typically accounted first for known sociodemographic and clinical prognostic factors (e.g. performance status, weight loss)

Baseline HRQL predicts survival

- In 36 of 39 studies, HRQL proved to be a significant predictor of survival, above and beyond known prognostic factors
- Effect sizes (hazards ratios) tended to be small to moderate

How would you rate your overall health? excellent good fair poor

In general population studies, self-rated health is one of the most consistent, independent predictors of:

- use of medical and mental health services
- morbidity
- 5 and 10 year mortality

HRQL as prognostic factor: clinical trial applications

- stratification prior to randomization
 - help ensure pretreatment group equivalence
 - increase efficiency of trial
 - facilitate planned subgroup analyses

fact or fiction?

Drug regulatory agencies are increasingly open to and supportive of the use of QL outcomes in clinical trials.

**In theory, factual –
In practice, fiction?**

U.S. FDA 1985

“...In the past, new anti-cancer drugs were approved solely on the basis of objective tumor response, but this is no longer the case....Survival and quality of life are the key efficacy parameters.”

Johnson and Temple, *Cancer Trt Rep*, 1985

U.S. FDA 1996

“The Oncologic Drugs Advisory Committee has recommended that beneficial effects on QoL and/or survival be the basis for approval of new anticancer drugs...”

Beitz, Gnecco & Justice, *JNCI Monographs*, 1996

Endpoints in U.S. F.D.A. approval of oncology drugs: 1990 - 2002

- Marketing approval given to 57 drugs via standard procedures
- Basis of approval:
 - Survival – 32%
 - Tumor response – 46%
 - Tumor-specific symptoms – 23%
 - Other – 16%
 - Quality of life – 0%

• Johnson RJ et al. *J Clin Oncol* 2003; 21:1404-11

FDA drug approval based (in part) on symptom relief

Drug	Indication	Endpoint
Gemcitabine	Pancreatic cancer	Clinical benefit response (pain, PF, weight gain)
Paclitaxel	Karposi's sarcoma	Cosmesis
Liposomal daunorubicin	Karposi's sarcoma	Cosmesis
Methoxsalen	T-cell lymphoma	Edema, scaling
Mitoxantrone	Metastatic prostate cancer	Pain
Topotecan	SCLC	Lung cancer symptoms

“Although QOL assessments have been submitted in oncology drug applications, this aspect of the clinical trials has generally not been well conducted. Problems have usually included unblinded assessment, large amounts of missing data, and poorly defined prospective analytic plans...”

Johnson RJ et al. *J Clin Oncol* 2003; 21:1404-11

A question of miscommunication?

Williams G, Pazdur R, Temple R (FDA Division of Oncology Drug Products), *J Biopharm Stat* 2004

- “...in the late 1970's and early 1980's, FDA determined that acceptable endpoints for cancer drug approval were survival or an improvement in the quality of a patient's life, e.g., an improvement in tumor-related symptoms.”

A question of miscommunication?

Williams G, Pazdur R, Temple R (FDA Division of Oncology Drug Products), *J Biopharm Stat* 2004

- “...in the late 1970's and early 1980's, FDA determined that acceptable endpoints for cancer drug approval were survival or an improvement in the quality of a patient's life, i.e., an improvement in tumor-related symptoms.”

FDA PRO Guidance Document (draft 2006)

- Intended to inform/guide industry on use of PROs in medical product development to support labeling claims
- Developed “in-house” (coordinated by Laurie Burke) with advice from external expert (Donald Patrick)
- Although it has a relatively narrow focus, there is a real risk that it will impact significantly on HRQL research in a much broader context

FDA PRO Guidance Document: The pros (vs. cons)

- Generally well written, well-informed document reflecting the state-of-the-art in PRO assessment
- Provides good overview and concrete (non-binding) advice on:
 - Developing and evaluating PRO instruments
 - Designing and implementing PRO studies within trial context
 - PRO data analysis

FDA PRO Guidance Document The cons

- Risk of creating an uneven playing field, with higher standards required of PRO measures than of other clinical measures
- Risk of overly rigid (and thus counterproductive) application of guidelines in FDA practice
- Risk of the tail (FDA) wagging the dog (health outcomes research community)

FDA PRO Guidance Document Examples of (possible) misapplication

- Widely used instruments (e.g., SF-36, QLQ-C30, FACT) may be inappropriate because of insufficient patient input in their development
- Use of subscales from parent instruments inappropriate without extensive psychometric data supporting their use as “stand alone” measures
- Need to fully validate cultural translations/adaptations of existing measures before they can be used in trials

FDA PRO Guidance Document Examples of (possible) misapplication

- PRO measurement inappropriate in non-blinded clinical trials
- Anything longer than 24 hour recall as time frame for PRO assessment is problematic
- Global HRQL not sufficiently understood, defined, operationalized to allow for use in trials

How to improve the situation?

- Provide FDA with non-proprietary feedback (e.g., via ISOQOL, ISPOR, etc)
- Encourage FDA to train up their staff to better understand the theory *and* practice of health outcomes research
- Resist generalizing FDA position on diverse issues to health outcomes field, in general.

EMEA to the rescue?

- In 2006 published “reflection paper” on HRQL in regulatory setting
- More descriptive; less prescriptive
- More open/sympathetic to HRQL outcomes
- Possibility for middle-ground?

fact or fiction?

HRQL assessment is ready for prime time as a tool in daily clinical practice.

“faction”

HRQL in clinical research vs. practice

- HRQL outcomes are now widely accepted as relevant, if not essential to the clinical trial process
- In clinical practice HRQL issues also play a role, albeit informally, in decision-making
- However.....

Statement of the problem

- Functional and psychosocial health problems experienced by patients are often not discussed, and thus remain undetected and untreated:
 - Fatigue
 - Depression
 - Role functioning
 - Sexuality
 - Social isolation
 - Cognitive decline

This holds true for both primary and specialty care

Communication about specific HRQL topics (N = 240 oncology consultations with patients receiving palliative chemo)

	yes	physician initiated	physician asked open question
daily activities	64%	73%	40%
pain	72%	52%	28%
fatigue	46%	32%	7%
emotional	35%	21%	17%

Detmar et al. *JAMA* 2001; 285: 1351-7

Example of closed questions

Doctor: "So, at home you're able to do everything you want?"

Patient: "Well, no, not really. I do what I can manage."

Doctor: Mhmm. Did you have problems with a sore mouth this time as well?"

Let's change the subject

- Patient: "Well, I think it's a side-effect of the chemotherapy; that's probably why I'm physically and mentally exhausted?"
- Doctor: "Yes, probably, so, how is your pain?"

Patients' preferences for discussing HRQL issues (%)

	No	Yes, if the doctor initiates	Yes
Daily activities	10.4	25.4	62.1
Symptoms	1.3	9.2	87.9
Emotional issues	6.3	26.1	66.9

- No statistically significant differences in communication as a function of preferences

Detmar et al. *JAMA* 2001; 285: 1351-7

Roots of the problem – The doc

- Lack of time ("opening a can of worms")
- Lack of interest/low priority
- Lack of training and/or skills in eliciting problems
- Perceived paucity of effective interventions (e.g., fatigue). "Don't diagnose what you can't treat."

Roots of the problem – The patient

- Too many problems to discuss – comorbidity
- Belief that problems come with the territory (particularly if chronic)
- Reluctance to "burden" doc with problems
- Limited "vocabulary of distress" (e.g., children, poorly educated, ethnic minorities with language issues, cognitively challenged patients)
- Culturally-determined willingness to raise issues and express emotions

Extremes of expressiveness

Southern Europe



Northern Europe



Roots of the problem – The system

- Concerns about reimbursement for staff time and effort
- Concerns with liability (responsibility to act on information)
- Absence of well-coordinated, multidisciplinary care
- Limited institutional mission statement ("center of excellence" = high cure rate)



Conspiracy of silence + doorknob phenomenon



Possible solutions

- Communication skills training
- Patient empowerment initiatives
- Development of effective medical and psychosocial interventions
- Introduction of standardized, routine assessment of patients' functional health and symptom experience

Possible solutions

- Communication skills training
- Patient empowerment initiatives
- Development of effective medical and psychosocial interventions
- Introduction of standardized, routine assessment of patients' functional health and symptom experience

Making the problem go away (a first step)

- Ensure that key physical, functional and psychosocial problems are assessed and reported to clinicians, nurses, and other caregivers on a regular basis

How?

- By means of routine, standardized assessments using patient self-report questionnaires that are:
 - brief and simple to complete
 - summarized in a simple, easily digestible format
 - easy to interpret

Patient-reported outcomes in clinical practice

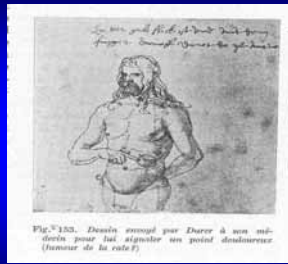


Fig. 153. Dessin envoyé par Durer à son médecin pour lui signaler un point de douleur (douleur de la rate).

Albrecht Durer, 1471-1528
German renaissance artist and mathematician

Brodman K. et al. The Cornell Medical Index: An adjunct to medical interview JAMA 1949; 140:531-4

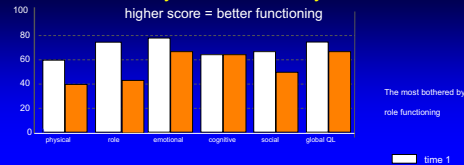
- 195 item self-administered questionnaire on physical and psychological symptoms and medical history
- completed prior to office visit in 10-30 minutes; high compliance rates
- Elicited information not found in medical records

Clinic-based HRQL data capture



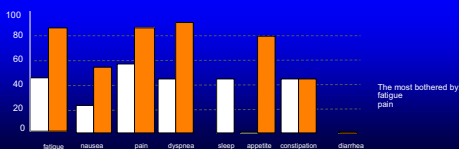
Quality of life Summary

higher score = better functioning



The most bothered by: role functioning

higher score = more symptoms

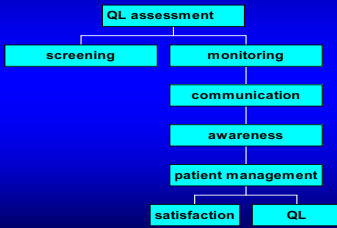


The most bothered by: fatigue pain

HRQL assessment in daily clinical practice: Feasibility

- Self-administered questionnaires can be completed quickly in office-based practice
- Computer-assisted (e.g., touchscreen) administration is acceptable and efficient
- No evidence that collection of standardized QL data interferes with normal clinic routine or lengthens average visit time

What can you expect to achieve? A cascade of effects



HRQL assessment in clinical practice

16 controlled studies published 1987-2004

(Taenzer et al. 2000; McLachlan et al. 2001; Detmar et al. 2002; Velikova et al. 2003)

- communication +
- awareness +
- patient management +/-
- satisfaction +/-
- HRQL +/-

HRQL assessment in daily clinical practice

Systematic review of RCT's

(Valderas JM et al. Qual Life Res 2008; 17:179-93)

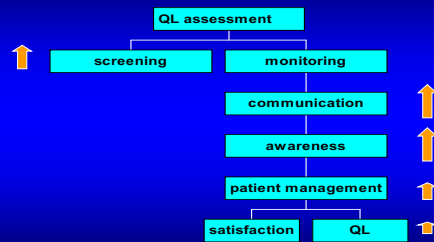
- 28 RCT's published between 1978-2007
- 19 in primary care; 9 in specialist care
- 54% of interventions were single point-in-time PRO assessments only
- 50% assessed psychiatric problems only
- 70% provided feedback to clinicians in real time

PRO assessment in daily clinical practice

Systematic review of RCT's

(Valderas JM et al. Qual Life Res 2008; 17:179-93)

(%)	Studies (n)	Significant results
Process of care		
• Advice, education, counseling	7	43%
• Target diagnoses and notations	14	50%
• Referrals, consultations	11	18%
Outcomes of care		
• General functional status	6	50%
• Satisfaction with care	12	42%
• Physician-rated utility of intervention	6	66% (28-97%)



Possible strategies to increase impact on patient management and health outcomes

- Get more concrete – Supplement or replace generic HRQL measures with condition-specific measures
- Combine quantitative, questionnaire-based HRQL data with qualitative, interview-based information
- Link HRQL information to treatment guidelines and clinical pathways

Possible strategies to increase impact on patient management and health outcomes

- Get more concrete – Supplement or replace generic HRQL measures with condition-specific measures
- Combine quantitative, questionnaire-based HRQL data with qualitative, interview-based information
- Link HRQL information to treatment guidelines and clinical pathways

The use of HRQL assessments in daily clinical oncology nursing practice:

A community hospital-based intervention study

Doranne L. Hilarius, Paul Kloeg, Chad M. Gundy, Neil K. Aaronson

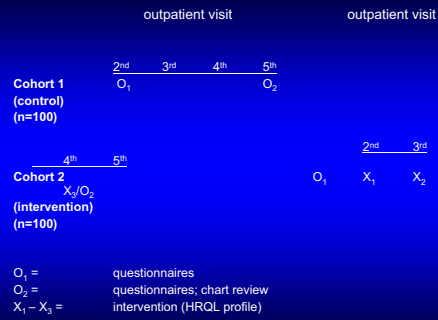
Cancer (in press)

Study participants

- 219 cancer patients receiving adjuvant or palliative chemotherapy in the outpatient clinic of a large community hospital in North Holland.
- 11 oncology nurses responsible for the delivery of the chemotherapy

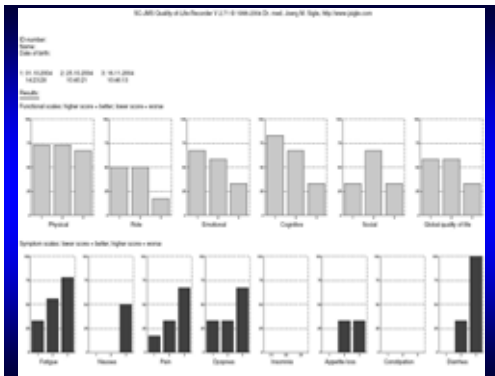
Research design

- Classical randomized study was contra-indicated due to risk of contamination effect
- Chose for **sequential cohort design**
 - 1st cohort of 100 patients = usual care control group
 - 2nd cohort of 100 patients = intervention group



Intervention

- Completion via computer touch screen of:
 - a "core" HRQL questionnaire (the EORTC QLQ-C30)
 - EORTC HRQL modules for breast, colon and lung cancer
- Graphic summary provided to both patient and nurse immediately before encounter
- Cumulative scores presented at each subsequent visit



Key results

- Significant increase in:
 - frequency with which HRQL issues were discussed (both generic and condition-specific)
 - nurses' awareness of patients' HRQL
 - number of HRQL-related notations in the medical records
 - HRQL-related counseling behavior
- No significant effect on patients' satisfaction or HRQL over time

Possible strategies to increase impact on patient management and health outcomes

- Get concrete – Supplement or replace generic HRQL measures with condition-specific measures
- Combine quantitative, questionnaire-based HRQL data with qualitative, interview-based information
- Link HRQL information to treatment guidelines and clinical pathways

Assessment is not enough: A randomized controlled trial of the effects of HRQL assessment on quality of life and satisfaction in oncology clinical practice

Rosenbloom SK, Victorson DE, Hahn E, Peterman AH, Cella D

Psycho-oncology 2007; 16:1069-79

Study participants

- 213 cancer patients receiving palliative chemotherapy for breast, lung or colorectal cancer
- oncology nurses responsible for the delivery of the chemotherapy (*n* not reported)

Study design and intervention

- 3-arm randomized clinical trial
 - Completion of FACT-G, with summary given to treating nurse prior to clinical encounter
 - Completion of FACT-G + personal interview, with summary given to treating nurse
 - Usual care control group

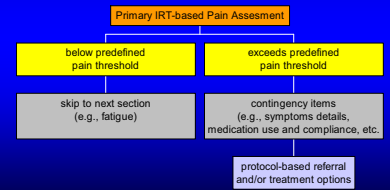
Key results

- No significant impact on:
 - Patient management
 - Patient satisfaction over time
 - Patient HRQL over time
- Nurse-patient communication and nurses' awareness not assessed

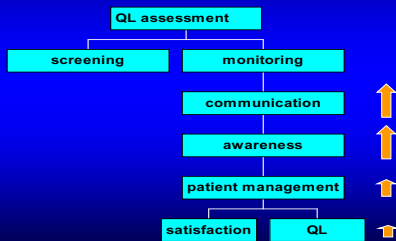
Possible strategies to increase impact on patient management and health outcomes

- Get concrete – Supplement or replace generic HRQL measures with condition-specific measures
- Combine quantitative, questionnaire-based HRQL data with qualitative, interview-based information
- Link HRQL information to treatment guidelines and clinical pathways

CAT + Contingency approach to HRQL assessment in daily clinical practice



Conclusions: Results to date



Future directions

- Develop more efficient questionnaires using computer adaptive (dynamic) testing
- Identify critical thresholds for symptoms and functional impairment that trigger more specific probes
- Link clinically relevant HRQL outcomes to treatment guidelines
- Develop tailored health education feedback to patients based on their HRQL responses (clinical pathways)

“It is likely that in the early years of the 21st century, the completion of a quality of life questionnaire at a patient visit will be as routine as the taking of vital signs.”

9:61-5

Ganz PA. *Oncology* 1995;

“In theory there is no difference between theory and practice. In practice there is.”
Yogi Berra

“The future ain’t what it used to be”
Casey Stengel

“The best way to predict the future is to invent it.”

Alan Kay

