

The EORTC QLQ-C30: Past, Present and Future

Neil Aaronson, Ph.D.

The Netherlands Cancer Institute



EORTC modular assessment approach

CORE questionnaire

+

diagnosis-specific or treatment-specific
MODULES

Core questionnaire

- common physical symptoms of cancer and its treatment
- physical and role functioning
- emotional functioning
- social functioning
- global health status and quality of life

Supplementary modules

- disease symptoms
- treatment side-effects
- additional quality-of-life domains

Lung cancer module

- dyspnea
- coughing
- more detailed pain assessment
- wider range of cytotoxic drug toxicity's

Prostate cancer module

- micturation problems
- detailed pain assessment
- intimacy/sexuality
- additional drug toxicities

- **UNIFORMITY** for cross-study comparisons
- **FLEXIBILITY** for adaptation to specific study needs

Desired characteristics of the core EORTC QL questionnaire

- health-related
- patient-based
- multidimensional
- brief and simple to complete
- psychometrically robust
- cross-culturally applicable

Generating HRQL Measures

sequential strategy

- development and testing of instrument in single country/language
- subsequent translation and validation of source instrument in other countries
- examples include:
 - ♦ SIP
 - ♦ NHP
 - ♦ MOS SF-36
 - ♦ FLIC
 - ♦ CARES
 - ♦ RSCL
 - ♦ FACT

Generating HRQL Measures

simultaneous strategy

- **multicultural input at each stage of instrument development process**
- **common protocol for field testing and psychometric analysis of instrument**
- **other examples include:**
 - ♦ **EUROQOL**
 - ♦ **WHOQOL**

Development of EORTC Core QL Questionnaire

- **multidisciplinary and multicultural input**
- **literature reviews and position papers**
- **selection, adaption and/or generation of common set of test items and response categories**
- **standardized procedure for forward-backward translation (into 9 languages)**

Development of core QL questionnaire (contd.)

- **pilot testing (n=15 per country) leading to further revisions**
- **international field-testing of 1st generation questionnaire - QLQ-C36**
- **revision of QLQ-C36 and pilot testing of 2nd generation - QLQ-C30 (version 1.0)**
- **international field-testing of QLQ-C30**

EORTC QLQ-C30 (version 1.0)

<u>Functional Scales</u>	<u>No. of items</u>
▪ physical	5
▪ role	2
▪ emotional	4
▪ cognitive	2
▪ social	2
▪ quality of life	2

EORTC QLQ-C30 (version 1.0)

<u>Symptom Scales/Items</u>	<u>No. of items</u>
▪ fatigue	3
▪ pain	2
▪ nausea and vomiting	2
▪ appetite loss	1
▪ dyspnea	1
▪ sleep disturbance	1
▪ constipation	1
▪ diarrhea	1
▪ financial impact	1

Original international field study

Aaronson NK et al. J Natl Cancer Inst. 1993, 85:365-376.

	Pretreatment n	On-treatment n
TOTAL	346	305
English-Speaking	147	133
Australia	24	19
Canada	58	50
United Kingdom	41	40
United States	24	24
Northern Europe	143	120
Germany	35	32
The Netherlands	30	28
Denmark	11	4
Norway	36	30
Sweden	31	26
Southern Europe	56	52
Belgium	25	22
France	17	17
Italy	14	13

respondent burden

- average time to complete
 - ♦ pretreatment = 12'' (sd = 8'')
 - ♦ on-treatment = 12'' (sd = 7'')
- large majority self-administered
- some preference for interview among:
 - ♦ elderly
 - ♦ poorly educated
 - ♦ physically weak
 - ♦ visually handicapped

Evaluation of the internal structure QLQ-C30 (scale development)

- examination of score dispersion and symmetry
- multitrait scaling
- reliability estimation

Internal consistency of the QLQ-C30 scales across 7 studies

Scale	Coefficient α	
	Range	Average
physical	.64-.76	.71
role	.31-.74	.54
emotional	.73-.85	.80
cognitive	.28-.73	.62
social	.40-.83	.71
global QL	.80-.94	.88
fatigue	.80-.91	.85
pain	.70-.86	.81
emesis	.55-.84	.71

Test-retest reliability of the QLQ-C30 scales

(Hjermstad et al. J Clin Oncol. 1995; 13:1249-54)

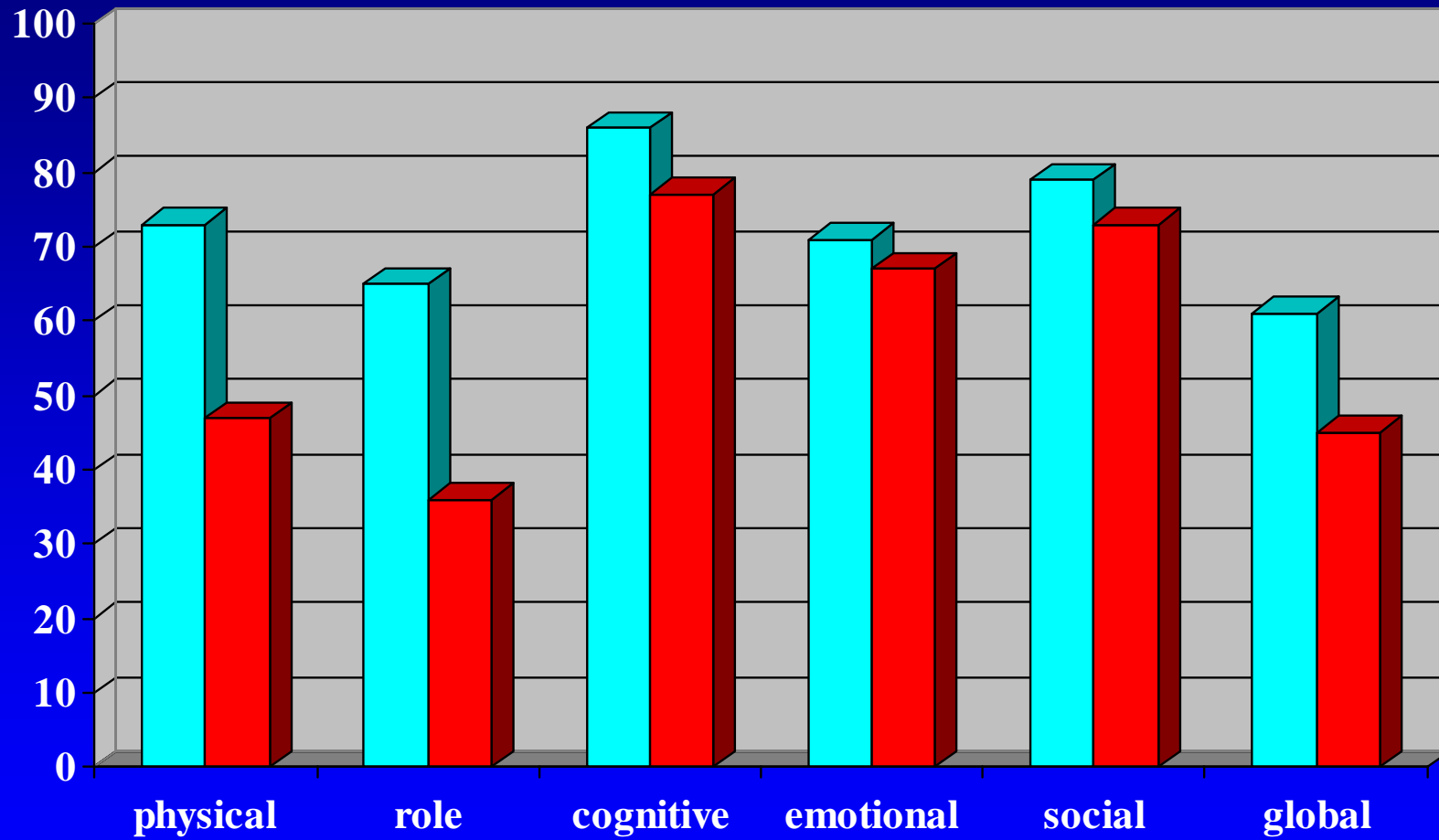
Scale	Spearman (4 day)
physical	.90
role	.85
emotional	.82
cognitive	.75
social	.79
global QL	.85
fatigue	.83
pain	.77
emesis	.70

Clinical validity of the QLQ-C30

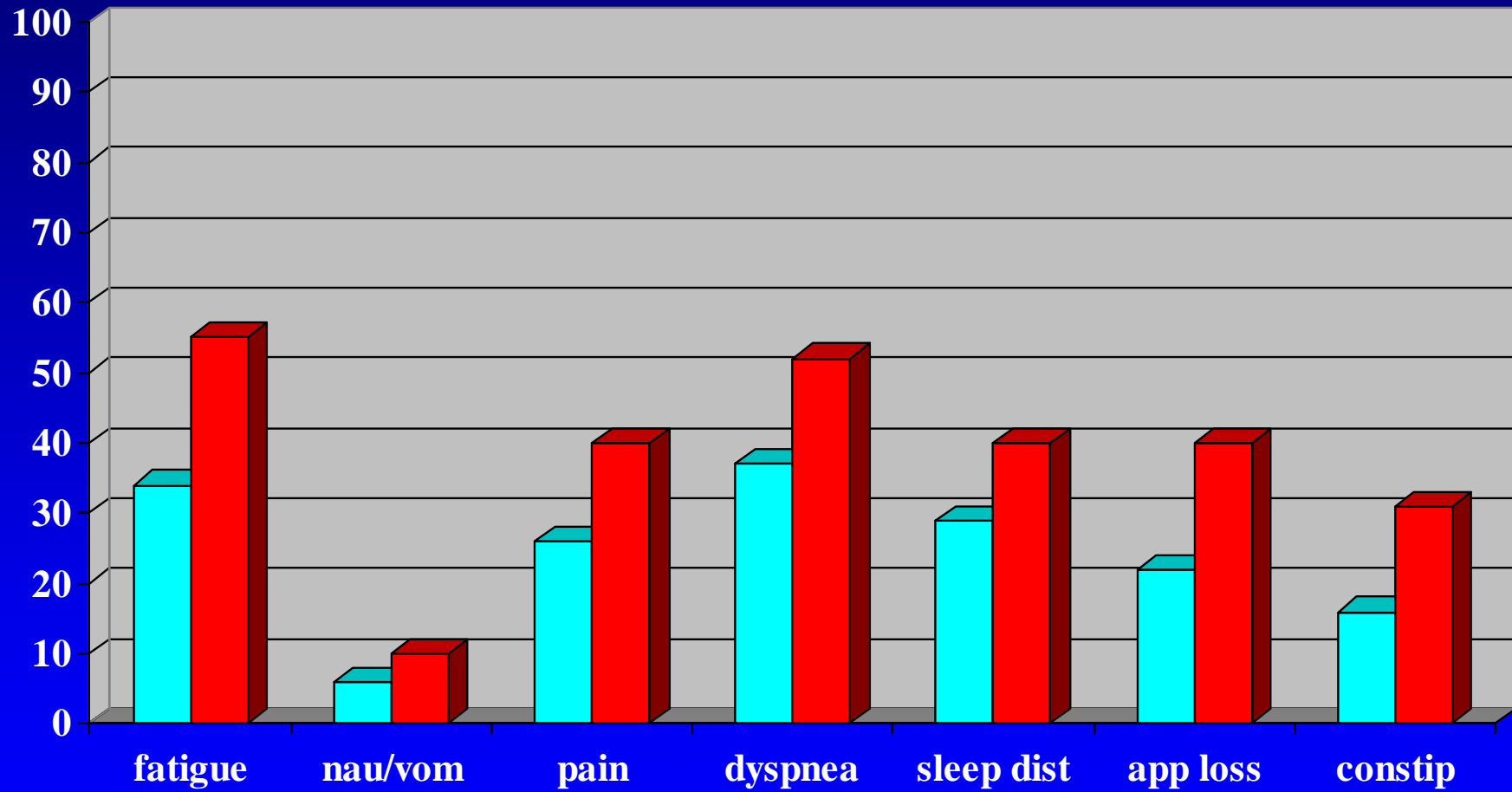
known-group comparisons

- disease stage
- performance status
- weight loss
- treatment modality
- treatment toxicity

Functioning scales x ECOG

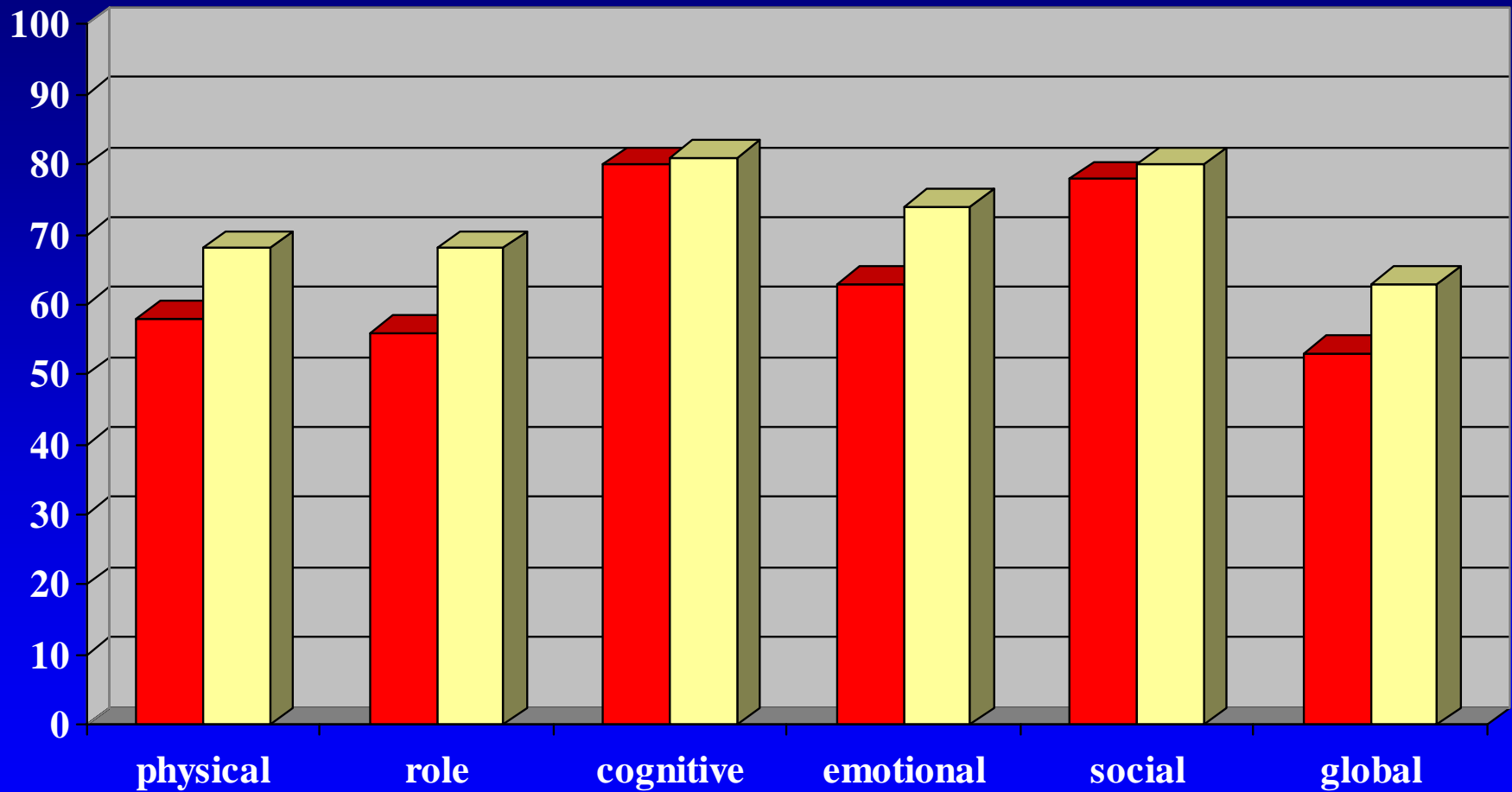


Symptom scores x ECOG

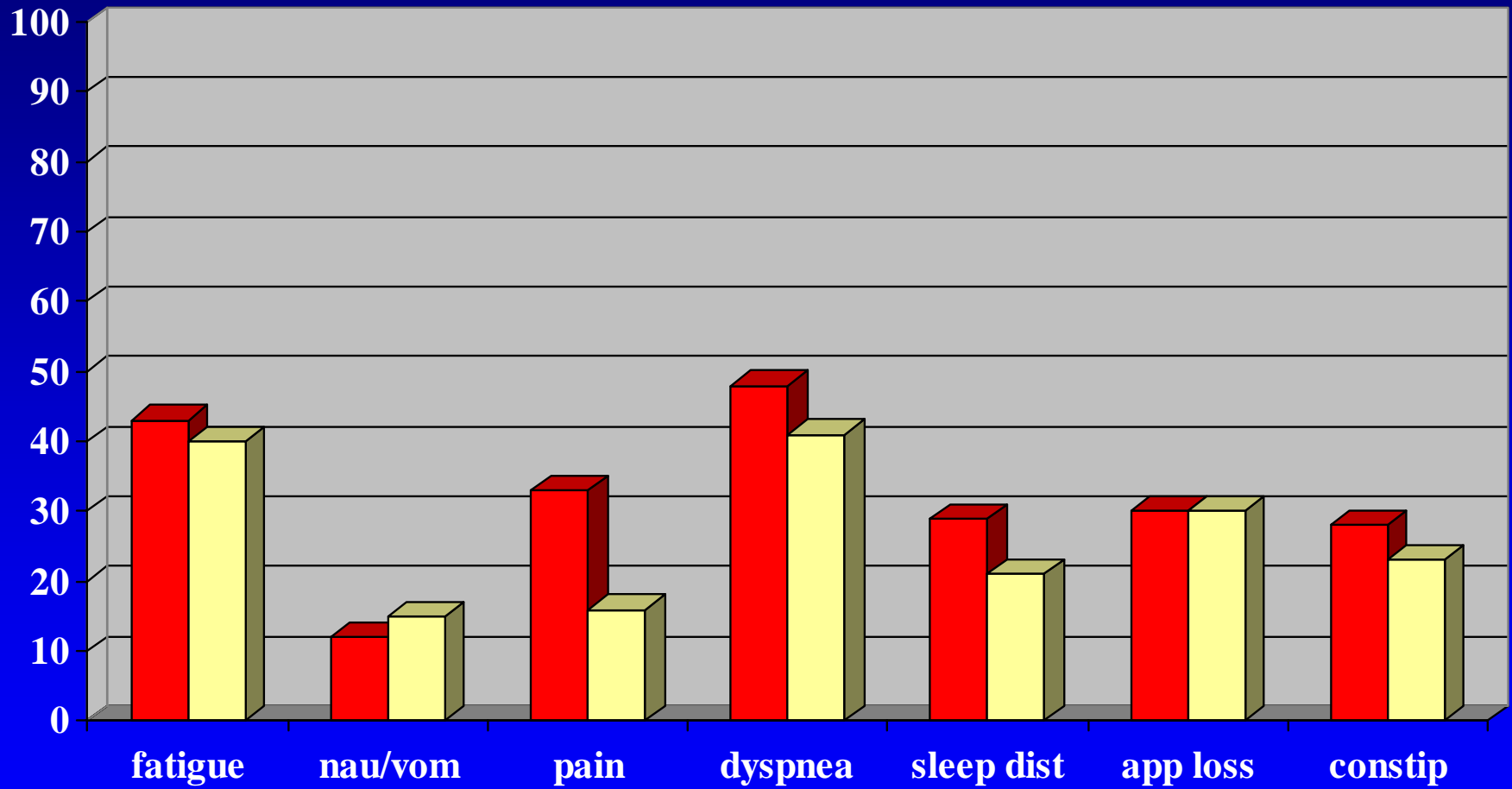


Responsiveness to change in
health status over time

Functioning scales: improved ECOG



Symptom Scores: improved ECOG



Summary of findings

The QLQ-C30 (version 1.0)

- could be quickly and easily completed by the large majority of patients
- exhibited satisfactory to excellent scaling properties
- exhibited validity and responsiveness to change in clinical status over time
- Performed similarly across broad language-cultural groups

Fine-tuning the QLQ-C30

versions 30(+3), 2.0 and 3.0

- role functioning scale
- global quality of life scale
- physical functioning scale

Modifications of the QLQ-C30

Role Functioning Scale

Original Wording

Are you limited in any way in doing either your work or doing household jobs?

Are you completely unable to work at a job or to do household jobs?

Yes or no

Revised Wording

Were you limited in doing either your work or other daily activities?

Were you limited pursuing your hobbies or other leisure time activities?

Not at all, a little, quite a bit, very much

Role functioning scale

Internal consistency and test-retest reliability

		<u>Original</u>	<u>Revised</u>
<u>Internal Consistency(α)</u>			
Canadian data (N = 696)	T ¹	0.61	0.82
	T ²	0.67	0.85
	T ³	0.59	0.88
Dutch data (N = 485)	T ¹	0.26	0.78
	T ²	0.35	0.83
	T ³	0.29	0.86
<u>Test-Retest (ICC)</u>			
Dutch data (N = 86)		0.59	0.82

Global quality of life scale

Original items

How would you rate your overall **physical condition** during the past week?

How would you rate your overall **quality of life** during the past week?

Additional (Replacement) item

How would you rate your **overall health** during the past week?

Physical Function Scale

Items

Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?

Do you have any trouble taking a long walk?

Do you have any trouble taking a short walk outside of the house?

Do you have to stay in bed or a chair for most of the day?

Do you need help with eating, dressing, washing yourself or using the toilet?

Response Choices

Dichotomous: scale range = 5 - 10

4 point Likert-type scale range = 5 - 20

Physical function scale: internal consistency reliability

(Canadian data N = 696)

	<u>Original</u>	<u>Revised</u>
T ¹	0.70	0.81
T ²	0.76	0.86
T ³	0.72	0.84

Ongoing and Future Research Activities

- Define clinically meaningful change
- Use modern (item-response) test theory and techniques to:
 - **identify differences in item “performance” across languages, age, gender, diagnosis, disease stage**
 - **generate abbreviated versions of the QLQ-C30**
 - **develop computer-adaptive version(s) for use in clinical research and practice**

Defining clinically meaningful change

e.g., Osoba D et al. J Clin Oncol. 1998;16:139-44.

- **Compared QLQ-C30 change scores with patients' retrospective rating of change (SSQ)**
- **Mean change scores of 5-10 points = small**
- **Mean change scores of 10+ points = moderate to large**
- **Report proportion of patients (per group) that has experienced the (10 point) threshold level of change for a predefined time period (e.g., over 2 cycles of therapy)**

Differential Item Function (DIF) Analysis: Assessing the equivalence of translations

Petersen MA et al. Quality Life Res. 2003; 12:373-385

- Tested whether QLQ-C30 EF scale items “functions” similarly across 10 languages (11 countries)
- Do patients with the same “level” of (emotional) functioning respond similarly to individual items, regardless of language

Differential Item Function (DIF) Analysis: Assessing the equivalence of translations

- Very large data set ($> 10,000$ cases)
- Found some evidence of DIF in 7 of 9 translations; may be due to:
 - **translation effect**
 - **cultural differences**
 - **confounding factors (case mix issue)**

Why is it important?

- **QLQ-C30 is currently available in 45 languages (from Arabic to Zulu), with more on the way**
- **Despite best efforts, errors can creep into translations**
- **DIF represents a robust, technical approach to evaluating the quality of translations**

Creating a brief(er) version of the QLQ-C30 using IRT and interview methods

Petersen MA, Groenvold M, Bjorner J et al.

- **QLQ-C30 may be too long for use in some settings; e.g. palliative care**
- **Use IRT methods to test whether multi-item scales can be shortened without significant loss of information**
- **Some scales or items may be inappropriate and thus might be eliminated**

Creating a brief(er) version of the QLQ-C30 using IRT and interview methods

Petersen MA, Groenvold M, Bjorner J et al.

- **Statistical analyses based on between 2,300 and 10,000 cases from a range of available datasets**
- **Interviews with 41 patients and 66 health care professionals from 6 countries**

Creating a brief(er) version of the QLQ-C30 using IRT and interview methods

Petersen MA, Groenvold M, Bjorner J et al.

- **5 scales shortened:**
 - **Emotional functioning:** 4 → 2 items
 - **Physical functioning:** 5 → 3 items
 - **Fatigue** 3 → 2 items
 - **Emesis** 2 → 1 item
 - **QL** 2 → 1 item
- **Role, social and cognitive functioning scales deleted (relatively unimportant or too complex)**
- **2 single items (diarrhea and finances) deleted**

QLQ-C15-PAL

The Future

Computer Adaptive Testing

- Dynamic assessment vs. fixed-length questionnaires
- IRT-based algorithms
- Potential for highly flexible, efficient and reliable assessment of core HRQL issues
- Combine with “branching” algorithm (stem + contingency questions)

CAT + contingency approach to HRQL assessment in daily clinical practice

