



Trans Tasman Radiation Oncology
Group Limited
ACN 132 672 292

TROG POLICY AND PROCEDURES

Assessing Health Related Quality of Life in Cancer Patients:

A Research Clinicians Guide

TPP E7

Version 2: 21st August 2007

(Always refer to the TROG website to check for the current version of this policy)

TROG Central Operations Office

Level 5, Building 7
Calvary Mater Newcastle
Locked Bag 7 HRMC NSW 2310
Tel: + 61 2 49 211 466
Fax: + 61 2 49 211 465
Email: trog@trog.com.au
Website: www.trog.com.au

Contents

1	Introduction	3
2	Differences between Measures	4
3	Design Issues.....	6
4	When is it appropriate to assess HRQOL?	6
5	Data collection.....	7
6	Choosing the right instrument for your study	10
7	Can Questionnaires be modified?	12
8	Type of response scale	12
9	Cross-cultural application	13
10	The European Organisation for Research and Treatment of Cancer (EORTC) suite 46, 47	13
11	Functional Assessment of Chronic Illness Therapy (FACIT) suite ^{49, 50}	15
12	Analysing, Interpreting and Reporting HRQOL Data.....	16
13	Summary.....	19
14	Acknowledgements	20
15	References.....	20
16	Suggested Reading.....	23
17	Online Resources.....	24

1 Introduction

The benefits of cancer therapies are no longer assessed by solely focusing on traditional biomedical outcomes such as tumour response, survival and treatment related toxicity. When available treatments provide equivalent outcomes, or when a toxic treatment achieves only partial response, Health Related Quality of Life (HRQOL) may be the most appropriate endpoint. The National Health and Medical Research Council's strategic review identified "improving quality of life and the effectiveness, efficiency and equity of the health system" (¹ p. 144) as the most important incentives for Government's investment in health and medical research; measurement of HRQOL contributes to all of these goals. A review of the benefits offered by HRQOL outcome measures to oncology generally and to radiation therapy trials in particular is presented by Siddiqui, Kachnic and Movsas (2006) ². Also informative is the RTOG Model for Outcomes Assessment presented by Bruner et al. (2004) ³, which situates HRQOL within a triad of humanistic, clinical and economic outcomes.

The theoretical framework of HRQOL has arisen from the World Health Organization's definition of health as a state of complete physical, mental and social well being and not merely the absence of disease ⁴. It is generally accepted that HRQOL is a multidimensional construct ⁵ that is strongly influenced by subjective factors such as past experiences, beliefs and expectations ^{6, 7}. The dimensions of HRQOL most commonly included in outcome measures are psychological functioning, physical functioning, disease related and treatment related functioning, emotional functioning and social functioning ^{5, 8, 9}. However, role functioning, cognitive issues, sexual functioning, spirituality, financial concerns, job satisfaction and living conditions may also be considered important to measure ². Instruments also sometimes include a global question such as, "How would you rate your overall quality of life during the past week?" The terms HRQOL, 'Quality of Life (QOL)' and 'patient reported outcomes' are used interchangeably in much of the literature, though there continue to be advocates for distinguishing between them.

In clinical trials, it is suggested that HRQOL may be operationally defined by examining how physical, mental and social well-being are affected by medical treatment ⁹. Despite the acceptance of HRQOL as an important measure, the clinician giving consideration to the addition of this endpoint faces a number of methodological challenges. The measurement of quality rather than quantity of life necessitates the addition of psychosocial parameters into clinical trials so that subjective well being can be measured. Collecting such data requires a different approach from calculating traditional biological endpoints. Quality of life measures have in general been developed by social scientists, and commonly take the form of patient completed questionnaires. Depending on their focus, they may be lengthy and wide ranging, or extremely short and narrowly

focussed. Whilst some have demonstrated sound psychometric properties, others have not. It is important to remember that a tradition of use of a particular measure is not in itself testament to its psychometric properties or its appropriateness for use with a particular group. Instruments should never be said to be valid in themselves; rather, their application is more or less valid with particular populations in particular contexts (O'Connor, 2004, p.46) ¹⁰. The current guidelines will identify the pertinent design issues that research clinicians are advised to consider when including an HRQOL endpoint. A selection of instruments most likely to be appropriate for inclusion in cancer trials will be described. By so doing it is hoped that clinicians embarking on oncological trials may be assisted to make informed choices about appropriate HRQOL measures, and may better understand the strengths and limitations of their choices.

2 Differences between Measures

HRQOL measures have evolved from two main disciplines - psychology and economics ⁶. The psychometric approach has evolved from psychology and produced instruments intended to measure the multidimensional effect of a particular condition or treatment. In particular, it provides information about subjective well-being, another term sometimes used synonymously with HRQOL. The psychometric measures used to assess HRQOL may be arranged hierarchically according to their degree of specificity. Generic instruments have been designed for use in the general population; they are not specific to any particular group or disease. They are best used for the purposes of epidemiological and health policy research, to facilitate comparison between disease states or to measure HRQOL in diseases where no specific measure exists ¹¹. Such lack of specificity makes generic instruments poorly suited to oncology trials; their broadness of scope tends to limit their ability to detect relevant differences in health status or HRQOL over time. This is a general problem; a meta-analysis has shown that of generic measures are less responsive than disease specific measures ¹². Since clinically significant changes in HRQOL following treatment may be relatively small, responsiveness is an essential property to look for when choosing a HRQOL measure. Disease specific instruments focus on the most salient domains of a disease category as well as on specific patient characteristics. A number of such instruments have been designed for use with cancer populations. As cancer at different sites may have widely differing effects on HRQOL, these instruments still represent a rather broad brush approach. Site specific instruments are those which attempt to measure the impact on HRQOL of cancer at a particular anatomical site. Often referred to as modules, these instruments are generally brief, have a narrow focus, and are designed to be administered in addition to a core measure. A module may be developed to gauge side effects associated with a particular treatment (e.g., nausea associated with chemotherapy), disease symptoms related to a particular site (e.g.,

dyspnoea in lung cancer), or additional HRQOL domains affected by the disease or its treatment (e.g., issues of sexuality or body image). Examples of such modules are The Functional Assessment of Cancer Therapy - Lung Cancer Quality of Life Instrument (FACT-L), and the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire specific for breast cancer (EORTC QLQ-BR23). Additional site specific instruments are currently being developed and validated by a number of cooperative research groups, including those with an interest in Radiation Oncology ¹³.

Health economists interested in cost-effectiveness analysis have developed 'utility' measures. By the use of a scale ranging from 0 – 1 (in which 1 represents perfect health and 0 represents death), utility measures generally yield a single summary score. This approach aims to provide information about the relative desirability of different health states that can be used to inform health care policy and planning. Although intrinsically appealing to clinicians or administrators looking for data that can assist them to evaluate trade-offs between quality and length of life, the methods employed to measure the HRQOL benefit of a treatment have some shortcomings. Most commonly, weightings of the relative value of one outcome over another have been derived from the opinions of experts or healthy populations. Implicit in this approach is the belief that these weightings will approximate to patient preferences. However, this assumption is problematic for a range of reasons; to date, there is no evidence that it is valid ¹⁴.

Utility measures can be used to derive quality adjusted life years (QALYs), which can then used as the unit of health benefit in cost-utility analyses. A related approach is to adjust survival for quality of life. Examples include the Q-Twist ¹⁵ and Q-Time (the latter was developed by RTOG) ¹⁶ This approach leads to analysis of quality-adjusted survival.

Currently there are challenges in implementing both psychometric and utility approaches to HRQOL measurement. Instruments developed by social scientists provide a detailed descriptive account of the patients' perspective and are responsive to change over time. However, they do not anchor responses to a common standard of health, nor do they typically require patients to indicate the personal importance of difficulties occurring in different domains ⁸ (but see footnote ⁱ). Consequently, it is hard to determine with any degree of certainty the magnitude of impact of a particular outcome, especially at the population level. Instruments developed by economists on the other hand, whilst addressing these issues, do so in a manner which is less than satisfactory.

ⁱ An alternative approach to HRQOL assessment, commonly called individualised measures, allows people not only to weight domains but also to specify the domains that are most important to them (e.g. the SEIQoL). However this raises issues of comparability between people and across trials.

Clinical researchers are best advised to use instruments that have demonstrated sound psychometric properties with the groups that form the focus of the trial, and to tailor their protocols with regard for the specific requirements of the collection of psychosocial data. Specific factors for clinical consideration will be addressed in the next section.

3 Design Issues

The decision to measure HRQOL has an impact upon the design and conduct of clinical research. The experiences of clinical trials groups in this area is well reported elsewhere ¹⁷⁻²⁰ (see Movsas, 2003 ²¹, for a perspective from RTOG). The current guidelines will provide a brief description of the salient points identified from the literature.

4 When is it appropriate to assess HRQOL?

Measurement of HRQOL is almost always relevant, although it will often be secondary to other outcomes such as survival. Most trials will fall into one of the following categories for which measurement of HRQOL is indicated:

1. Different treatment modalities are compared;
2. Treatments of different intensity or duration are compared;
3. Treatments are expected to be of similar efficacy (e.g., survival);
4. The disease site is associated with very poor prognosis, and therefore HRQOL may be seen as the primary endpoint;
5. Adjuvant therapy for patients at low risk of recurrence of disease is compared;
6. Treatment is recognised as very burdensome to patients;
7. Treatments differ in short-term efficacy but the overall failure rate is high
8. A new (invasive) treatment is being evaluated ^{9, 19, 22}.

Although the majority of trials to include quality of life measures have been Phase III, there are a number of reasons why quality of life might also usefully be measured at Phase I and II. First, measurement of quality of life may be important in establishing the feasibility of a Phase III trial and in determining effect size and calculating sample size. In particular, it is important to obtain patient perspectives on new treatments that include information about the frequency, severity and burden of side-effects. Also, while Phase I and II trials do not offer the opportunity to compare quality of life data between groups, there may be important opportunities to correlate patient reports with toxicity data. The selection of an appropriate quality of life measure may also offer a means of quantifying response to treatment where a disease does not lend itself to measurement of biomedical outcomes. Phase I and II trials may provide useful data for validation of a given instrument ²³. Finally, the physical and psychological burden that Phase I and II trials place on patients may demand measures of their wellbeing to ensure that participants' interests are not being overshadowed by those of science or society ²⁴.

5 Data collection

5.1 Planned data collection

Because of the nature of the required information, and because time points for collection of information are specific, retrospective collection of HRQOL data is not feasible; rather, data collection must be planned prospectively. This requires a change from a retrieval mechanism to an interactive approach requiring direct patient contact. Many multicentre studies identify a key person at each site, such as a research nurse, to coordinate the collection of HRQOL data.

5.2 Integration of HRQOL endpoints into trial

In 1994, the RTOG reported that accrual rates into studies with an integrated HRQOL component were much higher (60 -90 %) than accrual rates into studies requiring separate registration for entry into a companion trial measuring HRQOL (20 - 40 %) ²⁵. In 1995, the EORTC Genito Urinary study group initiated a study to assess HRQOL issues amongst patients with urological cancer ²⁶. Recognising a high level of clinician scepticism regarding the potential benefits of HRQOL measurement, this endpoint was optional. As a consequence only seven of the 22 participating institutions entered HRQOL data. The authors noted that as all patients who were approached to participate did so, the low total accrual rate was a result of unwillingness on the part of clinicians not patients. When information is requested that is viewed as separate from that needed for treatment

purposes it is often met by resistance from patients and staff alike ²⁷. With this in mind it is recommended that HRQOL endpoints be integrated into study design ^{20, 28}.

5.3 Who completes the Questionnaire

The subjectivity of HRQOL data clearly indicates that HRQOL should ideally be assessed by self report methods. The inappropriateness of relying on a physicians' report is illustrated by a urological study which compared physicians and patients pre-treatment HRQOL assessments ²⁶. In this study, only half of the patients reporting moderate or severe urological symptoms were so recognised by their physician. Self report questionnaires are typically designed to be brief and self explanatory. It is expected that they may be completed by patients without assistance. To ensure the least inconvenience to patients and staff they are often best completed in the waiting room before the patient sees the doctor. In situations where patients cannot themselves report HRQOL, it may be necessary to rely on a proxy. A review by Sneeuw et al. (2002) ²⁹ generally found moderate to high agreement between patients and proxies regarding ratings of patient HRQOL, although proxies tended to report more problems with HRQOL than did patients themselves. Agreement also tended to be higher when proxies were family members than when they were health providers.

5.4 Length of measure

Compliance is improved if time taken to complete a measure is brief. Ten to twelve minutes appears to be an acceptable time for most patients ³⁰. If total completion time exceeds 15 minutes, very sick patients in particular may fail to complete them, resulting in a biased sample. Due to differences in disease progression, questionnaires will invariably include items that are poorly targeted at any given individual. Three approaches have been taken to reducing unnecessary burden on patients. The modular approach referred to earlier supplements a standard scale with a subscale of questions relevant to patients with cancer at a particular site or experiencing certain symptoms. Some measures are also offered in abbreviated versions that include only items most relevant to patients in specific groups (see, for example, the EORTC QLQ-C15-PAL below). Finally, and most promising, is work currently in progress to develop computerized adaptive testing (CAT) that will be able to select items from a large item bank based on individual's previous responses ³¹.

5.5 When should baseline measurement be made?

All studies should include a baseline measurement to enable comparison with subsequent measurement of HRQOL. A small number of studies have found effects for altering the timing of baseline assessment, most of which perhaps unsurprisingly found that scores

were affected when baseline measurement was delayed until after treatment started (see Hakamies-Blomqvist et al., 2001, ³² for a review). Hurny et al (1994) ³³ reported that baseline HRQOL scores taken on first day of treatment, from a sample of patients having adjuvant chemotherapy for breast cancer, appeared inflated. The researchers concluded that this may have represented a coping response. This suggests that day one of treatment may not be the optimal time to take a baseline measure of HRQOL as it may yield a falsely positive estimate. It may be that by administering a baseline measure before randomisation, researchers may best capture current HRQOL, without the added noise of the patients' hopes and fears about the future treatment. However, researchers should always follow the guidelines given by instrument developers whenever this is available.

5.6 When and how often should additional measurement be collected?

The timing of administration should be gauged against expected changes in health status. In many cases this will mean deteriorating health, and whilst it may seem better to administer a HRQOL measure whenever the patient is seen by the clinician, better quality data and better compliance rates may be achieved if measurement points are limited ²⁸. Data collection schedules must also have regard to clinician and institutional burden: A survey of researchers working in the area of psychosocial oncology revealed that the most common difficulty encountered whilst conducting quality of life studies was obtaining the cooperation of medical staff ³⁴.

Care also needs to be taken when measuring HRQOL in patients whose treatment (e.g., chemotherapy) may produce cyclic short term effects on HRQOL which will impact measurement at different points of the cycle. Klee et al. (2000) ³⁵ have provided evidence for a clinical model of the cyclic changes in physical symptoms and side-effects that may prove useful in planning the timing of HRQOL measurements. Bernard (2005) ³⁶ has advocated the use of patient diaries and event-related triggered data collection (e.g., using palm-pilots) as a partial solution to the problems encountered by timing.

Because HRQOL data cannot be collected retrospectively and missing data poses a problem for analysis may bias findings, a real-time tracking system could be considered to monitor compliance at each of the data collection time points ²¹. Compliance is defined as the number of questionnaires returned as a proportion of those anticipated. One way to improve compliance, then, is to define these parameters in such a way as to maximise compliance – for example, setting realistic 'windows' during which questionnaires should be returned and defining cut-off points after which there is no anticipation that dying

patients will complete their questionnaires ³⁷. Resources permitting, it is ideal if reminders can be given to patients to complete their questionnaires at appropriate times.

5.7 What timeframe should the patient consider when responding to items?

Most HRQOL measures inform the respondent what timeframe to consider when responding to items. Timeframes tend to seek a balance between the higher reliability and lower recall bias offered by shorter timeframes and the more representative time sample offered by longer timeframes. A timeframe of one week may represent an acceptable choice in most cases and is adopted by both of the most commonly used measurement suites designed for use with cancer patients, the EORTC QLQ and FACIT suites (see below for more details).

6 Choosing the right instrument for your study

- a) HRQOL instruments most commonly take the form of questionnaires. Many of the components of HRQOL cannot be directly observed but must be indirectly evaluated by measuring the responses to a collection of questions known to reflect the construct of interest. The instrument must demonstrate adequate coverage, reliability, validity, and responsiveness to change with the populations of interest in order to be of value in clinical research. A discussion of psychometric properties to consider when choosing an instrument can be found in Hays et al (1993) ³⁸.
- b) Coverage - the instrument should address each objective and subjective component of the domains that are relevant to the patient population, and which are likely to be affected by the treatment or the disease. For the purposes of clinical trials this will mean that a disease or site specific instrument will be more appropriate than a generic instrument.
- c) Reliability - Scores on HRQOL instruments contain some random error and may also contain some bias, as well as a component which reflects the patient's HRQOL. Test-retest and internal consistency are techniques commonly applied to determine the consistency with which an instrument is measuring a given construct. Condensing scale items into a single score is only justified if the scale has adequate internal consistency ²⁶ and, even then, should only be undertaken if the instrument's developers have recommended this. The statistic generally used to quantify internal consistency is Cronbach's α (alpha). Generally accepted standards for research dictate that α should exceed 0.7 if an application of an instrument is to be regarded as acceptably reliable. A high level of caution

is needed in interpreting reliability coefficients. The size of the coefficient will tend to be proportional to the variability between individual between-patient scores and the number of items. There may also be a trade-off between reliability and validity in that homogenous items give high reliability but may give poor coverage of the construct being measured (O'Connor, 2004, p.82) ¹⁰.

- d) Validity – describes the range of interpretations that can reasonably be placed on a measurement score – i.e. what does a person's score tell us about him or her? (McDowell & Newell, 1996, p. 29 – 37) ³⁹ Three types of validity are commonly evaluated to provide indirect evidence of validity. Content validity is an assessment of the completeness and appropriateness of a proposed scale. Most often, this will take the form of opinion by patients and clinicians as to the appropriate inclusion and exclusion of items that they consider will contribute to comprehensive and sufficient measurement of the construct.
- e) Criterion validity considers the degree to which scores on an instrument correlate with a 'gold standard'. In the case of HRQOL measures, however, there is typically no gold standard for ready immediate comparison. Instead, tests of predictive validity may be employed by assessing the degree to which scores on a given instrument predict future clinical outcomes of interest, such as disease progression and survival.
- f) Because of the absence of a gold standard, development of HRQOL measures typically involves an ongoing process of construct validation. Construct validity is an overarching process in which a number of indicators of validity – each of which maybe insufficient and problematic on its own - are accumulated to explore the robustness of the measure in question. Convergent validity, for example, might concern the degree to which scores on a new scale designed to measure a particular aspect(s) of HRQOL correlate with those from other, well established, HRQOL measures of the same aspect(s) of HRQOL. Divergent validity is conversely concerned with testing the degree to which scores correlate poorly or inversely with measures of a construct believed not to be related to the target construct. Tests of discriminative validity are used to see whether an instrument is able to distinguish between clinical groups which differ by established clinical criteria. Analyses of factorial validity may also be employed to examine the underlying conceptual structure of the scale by focussing on the degree to which items group together in measuring different themes within the overall construct.
- g) Sensitivity and responsiveness – these two related properties refer to an instrument's ability to identify clinically significant differences between groups (at a point in time) and

change within individual patients within a group (followed over time) respectively (Fayers & Machin, 2000, p. 66 - 71) ⁵. It is vital that an instrument is chosen that appropriately targets the range of HRQOL likely to be characteristic of the groups of interest. Floor and ceiling effects occur where the lowest and highest values on a scale fail to express the true range of patient experience. Such effects will significantly reduce an instrument's capacity to discriminate individuals (and hence groups) at the floor or ceiling, and to detect change (downward from floor value or upwards from the ceiling value) in individuals (and hence groups) at the floor or ceiling at baseline. These attributes are particularly important for instruments selected to evaluate the effectiveness of interventions.

7 Can Questionnaires be modified?

Instrument development is a complex process. Whilst it may be tempting to modify existing measures, perhaps by the inclusion or deletion of items, such modification must always be accompanied by additional analysis to ensure that applications of the modified instrument continue to be reliable and valid. Clinicians using instruments without proven reliability, validity, coverage, and responsiveness to change may be frustrated to discover that they are able to draw few if any valid conclusions from their findings.

8 Type of response scale

Most HRQOL instruments employ either a categorical or a linear analogue scale. The categorical scale has a limited number of labelled choices from which the patient must choose. The simplest categorical scale will consist of a 'yes' / 'no' choice. Whilst four or five response categories on a categorical scale achieve acceptable reliability ⁴⁰, the addition of more response choices improves scale precision, and may improve sensitivity and responsiveness (notwithstanding other key considerations about these two characteristics, noted above). Some categorical scales offer as many as 10 labelled choices ⁴¹. A linear analogue scale is usually a 10cm long line with descriptive anchors at either end but no categories marked along the length. The respondent is required to make a mark at any point along the line. Whilst arguably more precise than categorical scales, linear analogue scales have proven difficult for some patients to understand, resulting in missed or invalid responses ⁴². Preparing linear analogue data for computer entry is also time consuming and laborious ⁴⁰. These practical considerations have tended to outweigh the potential psychometric superiority of linear analogue scales.

As well as including multiple items that are each believed to contribute to an overall construct, scales can also include a global HRQOL question (e.g., “How would you rate your overall quality of life over the past week?” [EORTC QLQ-C30]), or even consist of just a single item⁴³. Single item measurement is clearly attractive from the perspective of time required to complete and minimal patient burden, especially when patients are very sick or HRQOL is not the primary endpoint. However, single item scales have certain psychometric properties that may or may not be appropriate to your research objectives and setting. (see Sloan et al., 2002⁴⁴, for a discussion).

9 Cross-cultural application

The subjective assessment required to measure quality of life renders it particularly vulnerable to differing interpretation across cultures. Culture is known to influence the way people define, recognise, experience and report their illness⁴⁵. In order that respondents from different cultures may be admissible to the trial, it is important that chosen instruments are as culturally unbiased as possible. Instruments which employ a weighted scoring system whereby some items are attributed more importance than other items are inevitably culturally biased as the weightings placed on items are influenced by cultural norms and values. Many instruments have been developed by people from a single country and then translated into other languages. In these instances, it is important that translators have ensured that wording is unambiguous and free from colloquialisms. Culture specific illustrations and explanations must be avoided. The EORTC core measure was developed by a group of people representing a range of countries and is possibly the least culture bound core measure. Even in this instance, the group showed a bias towards western cultures.

In the following section, two suites of instruments commonly used in cancer trials will be described. Whilst the instruments differ with regard to format, content, emphasis and scaling properties, they are both brief, patient-completed, multidimensional measures, and have established psychometric properties.

10 The European Organisation for Research and Treatment of Cancer (EORTC) suite 46, 47

The EORTC QLQ suite is designed for use across the full spectrum of cancer diagnoses in conjunction with supplementary modules, and has been used in over 3,000 studies worldwide. The QLQ-C30 is the core component of the EORTC’s modular approach to QOL assessment. It represents QOL domains relevant across a wide range of cancer sites and treatment types. It is complemented by modules specific to particular cancers, such as lung cancer and breast cancer.

The core module facilitates comparison across the diversity of trials administered by the EORTC, and the supplementary modules provide sensitivity for particular trials by providing more detailed information relevant to specific cancer sites and treatments.

The current EORTC QLQ-C30 has undergone a number of revisions since its inception as the QLQ-C36 in 1986. Subsequently modified to improve its psychometric properties and renamed the QLQ-C30, the most recent upgrade of the instrument (1998; Version 3.0) features 30 items. These are contained within five functional scales (measuring physical, role, emotional, social and cognitive functioning), three symptom scales (measuring pain, fatigue and nausea and vomiting), a global health status/ quality of life scale, and 6 single items measuring dyspnea, insomnia, appetite loss, constipation, diarrhea and financial impact. Scales and single items use four point response categories and refer to perceptions during the past week. The developers caution against using any cumulative global score, recommending that reference instead be made to patients' response to two global questions – “how would you rate your overall health during the past week?” and “how would you rate your overall quality of life during the past week?” The psychometric properties of the scale have been validated in a variety of cancer populations ⁴⁸. In order that the instrument would have cross-cultural application the questionnaire was developed by individuals representing a broad range of mostly European cultures. Their input influenced the choice of item content and item wording. The QLQ-C30 is available in 80 languages. The choice of an unweighted scoring system increases the possibility that the instrument will be applicable across cultures.

The EORTC QLQ-C15-PAL is a 15 item version of the QLQ-C30 that has been developed using expert and patient opinion and item response theory to target items most relevant to patients receiving palliative care.

The QLQ-C30 is supplemented by disease specific modules for Breast, Lung, Head and Neck, Oesophageal, Ovarian, Gastric, Cervical cancer and Multiple Myeloma. Other disease specific modules are under development but have yet to be validated. There is also a ‘satisfaction with care’ module. Like the QLQ-C30, these modules are usually developed for a multicultural research setting and are available in many languages.

The EORTC Quality of Life Group provide reference data to help enable researchers to determine sample size requirements for future trials. The Group also provides a scoring manual that includes advice on interpreting scores. These resources and electronic versions of the questionnaires and modules are available for free for academic users via the EORTC website ⁴⁷. These however are copyrighted instruments for which a new User's Agreement form must be completed and submitted to the Quality of Life Unit for each trial prior to use.

11 Functional Assessment of Chronic Illness Therapy (FACIT) suite

49, 50

Like the EORTC suite, the Functional Assessment of Chronic Illness Therapy Measurement System (FACIT) is comprised of a core instrument - the FACT-G – and subscales. The FACT-G was systematically developed by social scientists over a five year period. First trialled in 1992, it has undergone a number of refinements. The current version (version 4; 1997) comprises 27 items. Items are contained within four validated subscales (Physical Well-Being, Social/Family Well-Being, Emotional Well-Being, and Functional Well-Being). As well as domain scores, the instrument also yields a total HQL score. A five point response scale is used and refers to patient perception over the previous seven days. Completion time is approximately five minutes. Whilst the FACT-G is a multidimensional instrument, the method of patient review used in developing the FACT-G has resulted in an instrument with less emphasis on physical and functional domains than other HRQOL measures. The FACT-G is available in 45 languages. Normative data is available from the US from samples of patients with a range of cancers and for the general population ⁵¹. Unlike the QLQ-C30, the FACT-G also provides an overall HRQOL score.

In addition to the FACT-G, the FACIT suite includes specific subscales for breast, bladder, brain, colon, central nervous system, cervical, esophageal, endometrial, head and neck, hepatobiliary, leukemia, lung, lymphoma, ovarian and prostate cancers. There are also five treatment specific subscales available for patients with neurotoxicity from systemic chemotherapy, those undergoing bone marrow transplant, those receiving biologic response modifiers and/or retinoid therapy and those receiving taxane therapy. Further symptom-specific subscales refer to anorexia/cachexia, diarrhea, anemia/fatigue, endocrine symptoms, lymphedema, neutropenia, thrombocytopenia, fecal incontinence and urinary incontinence. The FACIT approach differs slightly from the EORTC modular system, where stand-alone modules are used in conjunction with the QLQ-C30. In the FACIT system, each of these disease, treatment and symptom-specific instruments implicitly includes the FACT-G as its core of questions. For example, the FACT-B contains all 27 questions from the FACT-G plus an additional 23 questions that relate specifically to breast cancer.

A Trial Outcome Index can be calculated for any of the disease, treatment and symptom specific instruments. This is the sum of the Physical Well-Being (PWB), Functional Well-Being (FWB), and "additional concerns" subscales, and has been found to be particularly responsive to change in HRQOL.

Two summary variations of the FACT-G are available: The first (the FANLTC), for patients with non-life threatening conditions, excludes the item “I worry about dying”; the second (the FACT-GP) is for use with the general population and includes only 21 of the 27 FACT-G items.

Non-cancer-specific subscales in the FACIT suite include those relating to palliative care, spiritual well-being, HIV infection, multiple sclerosis and treatment satisfaction.

12 Analysing, Interpreting and Reporting HRQOL Data

A number of guidelines are available to help authors ensure that they have adequately analysed, interpreted and reported findings from HRQOL measurement in clinical trials^{37, 52-58}. That data is of high quality to begin with is assumed as an essential prerequisite⁵².

12.1 Analysis

Readers are referred to the TROG statistical guidelines for information on sample calculation, interim analyses, sequential trials and stopping rules, and meta-analysis.

Estimation of sample size is based on the test size significance, the power and expected difference (effect size) in HRQOL over time or between groups⁵. For guidelines on calculating sample sizes in two group studies, readers are referred to Campbell et al. (1995)⁵⁹.

Only those HRQOL endpoints defined before the trial began should properly form the focus of analysis (Fayers & Machin, 2000, p.309 – 317)⁵. Confidence intervals, p-values, and effect sizes should be reported. Extent of missing data should be reported, and patterns of missing data should be examined to determine the extent to which the missing data are random or non-random (Fairclough, 2002,⁶⁰ Chapter 4). Methods for analysis, assumptions about the data distribution and missing data mechanisms, and definitions of clinically important differences should be made clear. The ‘intention to treat’ status or otherwise of analyses should be made explicit. The summary statistics required will vary according to how many possible values the HRQOL scale has (few or many) and the distribution of observed scores. Means and standard deviations commonly reported for scales which have a relatively large number of possible values (10 or more). Medians and interquartile ranges may be used when data are not normally distributed. Proportions are often used for scales which have 4 or less scale values. When choosing which statistical procedure to use for examining between-group differences or within-patient changes over time, attention should be paid to whether the data meets the normality assumptions of

parametric (e.g., z- or t-tests). Non-parametric (e.g., Mann-Whitney or Wilcoxon signed ranks) tests may be used when data are not normally distributed. Multiple regression is commonly used to adjust for patient characteristics such as age. Where there is a multiplicity of outcomes or a large number of repeated measurements are to be compared, the researcher should adopt conservative p-values as a precaution against false positives. Graphic exploration of data is recommended as an informal means of summarising complex data, identifying patterns of missing data and other unexpected patterns.

12.2 Interpretation

Because statistical significance alone is not a sufficient means of interpreting changes in HRQOL scores observed in clinical trials, other methods have had to be found⁶¹. Two different approaches have been taken to interpreting HRQOL data: anchor-based and distribution based. Anchor-based approaches are concerned with establishing whether changes in HRQOL scores following treatment are clinically significant by comparing these changes with some independently measured criterion. These methods focus on the threshold that distinguishes trivial differences from the Minimum (Clinically) Important Difference (MCID/MID), defined as “the smallest difference...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”⁶². Anchors can be patient-based, clinician-based or based on some other external criterion (O’Connor, 2004, p.52)¹⁰. Patient based anchors most often concern patients’ global judgment concerning outcome – for example, whether they have noticed a change in their condition and the severity or importance they attach to this change (see Osoba et al., 1998,⁵⁸ for an example). A less popular approach involves asking experienced clinicians to make judgments regarding the different impacts on patients’ daily lives they would associate with different scores on a given measure⁶². Problems with this approach relate to the fact that doctors may fail to recognise symptoms that are important to patients and, in any case, may simply defer to patients’ own opinions. More widely used anchors relate to clinically significant events such as performance status or toxicity⁶³. Once the MID has been identified, one can compare the proportions of patients in each group who have achieved it. This converts the MID to another useful measure – the number needed to treat (NNT)

²¹.

Fayers and Machin (2000; p.336 - 337)⁵ contrast “within-persons” methods where patients are asked to compare their current and previous HRQOL levels with a “between-person” method in which patients are asked to compare their present state to that of others with the same condition⁶⁴; results from the two methods have suggested similar estimates of minimal clinical importance, at least in a sample of patients with pulmonary disease. Still another approach involves making changes in HRQOL brought about through illness easier to relate to by comparing them to

changes brought about by familiar events, such as loss of job or divorce (see Testa and Simmons, 1996, ⁶ for an example).

Unlike anchor-based methods, distribution-based methods interpret the effect of a given intervention by examining the underlying distribution of the results. The measure of variability can be either within-patient (standard deviation of change over time) or between-patient (standard deviation of patient scores at a point in time, e.g., baseline). Cohen (1988; p.12, 25) ⁶⁵ has operationally defined effect sizes of 0.2, 0.5 and 0.8 as 'small', 'moderate' and 'large' effect sizes respectively. Cohen's arbitrary operational definitions of effect sizes have been found to approximate to levels of clinical importance in some ⁶⁶ though not all ⁵⁸ studies of HRQOL in cancer. A combined approach to deriving MID that uses both anchor-based and distribution-based methods has become more popular in recent years, especially in relation to the FACIT system ⁶⁷. In their summary of key issues related to methods for determining clinically significant differences on HRQOL measures, Wyrwich et al. ⁶⁸ urge authors to use and report multiple strategies both to enhance the interpretability of the HRQOL measure in question and to contribute more generally to understanding the clinical significance of change.

It will often be of interest to compare HRQOL scores observed in groups of patients in trials with those observed in a reference group such as the general population at large or general population excluding people with chronic illness. Obvious advantages exist here for selecting a measure that has been used extensively with populations of interest and/or for which normative data has been reported (e.g., FACT-G and EORTC QLQ-C30). However, when comparing between groups, care needs to be taken to adjust for differing age and gender distributions where this is appropriate ⁶⁹. Care is also advised in assuming that patients and members of the general population have responded to questions about HRQOL in comparable ways. Research suggests that patients may show a selective reporting bias in which they discount problems that they perceive are not related to their cancer or its treatment. Bias of this kind may result in higher than expected scores for patients compared with members of the wider population who are likely to respond to questions from a more general perspective ⁵.

Patients' responses to repeated administrations of HRQOL measures at the beginning, during and end of a trial may also undergo response shift due to adaptation to illness or changes in experience. Potentially confounding response shifts can take one of two forms. In the first (beta change), patients recalibrate their perceptions along the same scale, for example changing their perception of what constitutes 'severe' nausea after exposure to a new treatment which causes significant nausea, or changing their perception of what constitutes 'severe' pain after pain slowly but progressively worsens with disease progression. The second and more worrying form of

response shift (gamma change) is when patients reconceptualise the constructs being measured – in the worst case, fundamentally changing their perception of what is meant by the terms used in a global quality of life question ⁷⁰.

12.3 Reporting and presentation of HRQOL results

Efficace et al. (2003) ⁵⁶ have offered a useful checklist for evaluating reports of HRQOL data. The checklist groups 11 issues into four categories – conceptual, measurement, methodology and interpretation (see Appendix 1). The authors define a paper with high quality HRQOL outcomes as one that meets at least 8 out of the 11 criteria, including three high-priority concerns as defined by a panel of 30 international experts - "baseline compliance reported", "psychometric properties reported" and "missing data documented". Osoba et al. (2005) ⁵² provide a step-by-step process for the analysis and interpretation of HRQOL data that includes advice on how to systematically account for missing data. Justification of why a particular measure has been used is universally regarded as important ⁵⁷ – the fact that the instrument has been used repeatedly in previous trials is not adequate justification in itself. Fayers et al. (1997) ³⁷ and Wiklund (2004) ⁵³ have gone further to emphasise the importance of justifying why HRQOL has been measured at all.

13 Summary

The breadth of clinical research aims in oncology trials suggests that it is unlikely that a single instrument will ever be appropriate for all trials. The EORTC and FACIT suites offer general instruments with a range of additional site and treatment and symptom specific modules. They are psychometrically sound, and are available in a large and increasing number of languages. Importantly, they are subject to ongoing development and improvement in terms of standards of measurement and interpretation. These suites provide the most widely used HRQOL instruments in cancer internationally. These instruments increase their utility by providing a wealth of comparative and reference data to facilitate the planning and interpretation of future studies, including calculating sample sizes and defining effect sizes. This is an important consideration for clinicians wanting to select an instrument. By having regard to the issues of design and measurement outlined in this paper, clinicians may be confident that the collection of HRQOL data will be a worthwhile exercise that will further their evaluation of cancer treatments.

14 Acknowledgements

These guidelines were originally authored for TROG by Dr. Nigel Spry in January, 1998. The current, updated version was prepared by Drs. Madeleine King and Tim Luckett on behalf of the Quality of Life Office of the Psych-oncology Co-operative Research Group (PoCoG) in June, 2007.

15 References

1. National Health and Medical Research Council. *The virtuous cycle: Working together for health and medical research. Health and medical research strategic review discussion document* December 1998. Available online at <http://www.health.gov.au/internet/wcms/publishing.nsf/content/hmrslr.htm>.
2. Siddiqui F, Kachnic LA, Movsas B. Quality-of-life outcomes in oncology. *Hematology - Oncology Clinics of North America*. Feb 2006;20(1):165-185.
3. Bruner DW, Movsas B, Konski A, et al. Outcomes research in cancer clinical trial cooperative groups: the RTOG model. *Quality of Life Research*. Aug 2004;13(6):1025-1041.
4. World Health Organization. *Constitution of the World Health Organization*. Geneva: WHO Basic Documents; 1948.
5. Fayers PM, Machin D. *Quality of life: Assessment, analysis and interpretation*. Chichester, UK: John Wiley and Sons, Ltd; 2000.
6. Testa MA, Simonson DC. Assessment of quality of life outcomes. *New England Journal of Medicine*. 1996;334:835-840.
7. Patrick DL, Bush JW, Chen MM. Toward an operational definition of health. *Journal of Health & Social Behavior*. Mar 1973;14(1):6-23.
8. Cella DF, Tulsky DS. Measuring quality of life today: methodological aspects. *Oncology (Huntington)*. May 1990;4(5):29-38; discussion 69.
9. Moinpour CM, Feigl P, Metch B, Hayden KA, Meyskens FL, Jr., Crowley J. Quality of life end points in cancer clinical trials: review and recommendations. *Journal of the National Cancer Institute*. Apr 5 1989;81(7):485-495.
10. O'Connor R. *Measuring quality of life in health*. Elsevier/Churchill Livingstone; 2004.
11. Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care I: Applications and issues in assessment. *British Medical Journal*. Oct 31 1992;305(6861):1074-1077.
12. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *Journal of Clinical Epidemiology*. Jan 2003;56(1):52-60.
13. EORTC Radiation Oncology Group. European Organisation for Research and Treatment of Cancer Group for Research in Radiation Oncology. <http://groups.eortc.be/radio/>. Accessed June 5th, 2007.
14. Carr A, Higginson IJ. Measuring quality of life: Are quality of life measures patient centred? *BMJ: British Medical Journal*. Jun 2001;322(7298):1357-1360.
15. Goldhirsch A, Gelber RD, Simes RJ, Glasziou P, Coates AS. Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis.[see comment]. *Journal of Clinical Oncology*. Jan 1989;7(1):36-44.
16. Murray KJ, Nelson DF, Scott C, et al. Quality adjusted survival analysis of malignant glioma. Patients treated with twice-daily radiation (RT) and carmustine: A report of Radiation Therapy Oncology Group (RTOG) 83 - 02. *Int J Radiat Oncol Biol Phys*. 1995;31(3):453-459.

17. Kornblith AB, Anderson J, Cella DF, et al. Quality of life assessment of Hodgkin's disease survivors: a model for cooperative clinical trials. *Oncology (Huntington)*. May 1990;4(5):93-101; discussion 104.
18. Moinpour CM, Hayden KA, Thompson IM, Feigl P, Metch B. Quality of life assessment in Southwest Oncology Group trials. *Oncology (Huntington)*. 1990;4(5):79-84.
19. Kiebert GM, Kaasa S. Quality of life in clinical cancer trials: experience and perspective of the European Organization for Research and Treatment of Cancer. *Journal of the National Cancer Institute*. 1996;Monographs.(20):91-95.
20. Osoba D, Dancey J, Zee B, Myles J, Pater J. Health-related quality-of-life studies of the National Cancer Institute of Canada Clinical Trials Group. *Journal of the National Cancer Institute*. 1996;Monographs.(20):107-111.
21. Movsas B. Quality of life in oncology trials: a clinical guide. *Seminars in Radiation Oncology*. Jul 2003;13(3):235-247.
22. Gotay CC, Moore TD. Assessing quality of life in head and neck cancer. *Quality of Life Research*. Feb 1992;1(1):5-17.
23. Cella D. Discussant: Phase I and Phase II Cancer Clinical Trials: What Can Patient-Reported Outcomes Do for You? Paper presented at: Patient Reported Outcomes Assessment in Cancer Trials: Evaluating and Enhancing the Payoff to Decision Making; September 20th - 21st, 2006; Bethesda, Maryland, USA.
24. Hope-Stone LD, Napier MP, Begent RH, Cushen N, O'Malley D. The importance of measuring quality of life in phase I/II trials of cancer therapy - the effects of antibody targeted therapy: part I. *European Journal of Cancer Care*. 1997;6(4):267-272.
25. Scott CB, Stetz J, Bruner DW, Wasserman TH. Radiation Therapy Oncology Group quality of life assessment: design, analysis, and data management issues. *Quality of Life Research*. Jun 1994;3(3):199-206.
26. da Silva FC, Fossa SD, Aaronson NK, et al. The quality of life of patients with newly diagnosed M1 prostate cancer: experience with EORTC clinical trial 30853. *European Journal of Cancer*. Jan 1996;32A(1):72-77.
27. Yancik R, Edwards BK, Yates JW. Assessing the quality of life of cancer patients: Practical issues in study implementation. *Journal of Psychosocial Oncology*. 1989;7:59-74.
28. Gotay CC, Korn EL, McCabe MS, Moore TD, Cheson BD. Building quality of life assessment into cancer treatment studies. *Oncology*. 1992;6:25-37.
29. Sneeuw KCA, Sprangers MAG, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. *Journal of Clinical Epidemiology*. Nov 2002;55(11):1130-1143.
30. Aaronson NK, Cull A, Kaasa S, Sprangers MAG. The EORTC modular approach to assessment of quality of life in oncology. *International Journal of Mental Health*. 1994;23:75-96.
31. Lai J-s, Cella D, Chang C-H, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: An illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*. Aug 2003;12(5):485-501.
32. Hakamies-Blomqvist L, Luoma ML, Sjostrom J, et al. Timing of quality of life (QoL) assessments as a source of error in oncological trials. *Journal of Advanced Nursing*. Sep 2001;35(5):709-716.
33. Hurny C, Bernhard J, Coates A, et al. Timing of baseline quality of life assessment in an international adjuvant breast cancer trial: its effect on patient self-estimation. The International Breast Cancer Study Group. *Annals of Oncology*. Jan 1994;5(1):65-74.
34. Aaronson NK, Van Dam FS, Polak CE, et al. Prospects and problems in European psychosocial oncology: a survey of the EORTC study group on quality of life. *Journal of Psychosocial Oncology*. 1989;4:43-53.
35. Klee MC, King MT, Machin D, Hansen HH. A clinical model for quality of life assessment in cancer patients receiving chemotherapy. *Annals of Oncology*. 2000;11(1):23-30.
36. Bernhard J. Timing of quality of life assessment in cancer clinical trials: fine tuning remains a challenge.[comment]. *Annals of Oncology*. Apr 2005;16(4):523-524.

37. Fayers PM, Hopwood P, Harvey A, Girling DJ, Machin D, Stephens R. Quality of life assessment in clinical trials--guidelines and a checklist for protocol writers: the U.K. Medical Research Council experience. MRC Cancer Trials Office. *European Journal of Cancer*. Jan 1997;33(1):20-28.
38. Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. *Quality of Life Research*. 1993;2:441-449.
39. McDowell I, Newell C. *Measuring health: A guide to rating scales and questionnaires*. Oxford, UK: Oxford University Press; 1996.
40. Fayers PM, Jones DR. Measuring and analysing quality of life in cancer clinical trials: a review. *Statistics in Medicine*. 1983;2:429-446.
41. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. *CMAJ Canadian Medical Association Journal*. Apr 15 1986;134(8):889-895.
42. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. *Journal of Chronic Diseases*. 1987;40(12):1129-1133.
43. Spitzer WO, Dobson AJ, Hall J, et al. Measuring the quality of life of cancer patients: a concise QL-index for use by physicians. *J Chronic Dis*. 1981;34:585-597.
44. Sloan JA, Aaronson N, Cappelleri JC, Fairclough DL, Varricchio C, the Clinical Significance Consensus Meeting Group. Assessing the Clinical Significance of Single Items Relative to Summated Scores. *Mayo Clinic Proceedings*. 2002;77(5):479-487.
45. Prayer L. *Medicine and culture*. New York: Henry Holt and Company, Inc; 1988.
46. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*. 1993;85:365-376.
47. European Organisation for Research and Treatment in Cancer. Welcome to the EORTC Quality of Life Group. <http://www.eortc.be/home/gol/>. Accessed 6th June, 2007.
48. Aaronson NK, Cull A, Kaasa S, Sprangers MAG. The European Organisation for Research and Treatment of Cancer (EORTC) Modular Approach to Quality of Life Assessment in Oncology: An Update. In: Spilker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials*. 2 ed. Philadelphia: Lippincott-Raven Publishers; 1996:179-189.
49. Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *Journal of Clinical Oncology*. 1993;11(570-579).
50. Functional Assessment of Chronic Illness Therapy. FACIT. <http://www.facit.org/>. Accessed 31st May, 2007.
51. Brucker PS, Yost K, Cashy J, Webster K, Cella D. General population and cancer patients norms for The Functional Assessment of Cancer Therapy-General (FACT-G). *Evaluation & the Health Professions*. Jun 2005;28(2):192-211.
52. Osoba D, Bezjak A, Brundage M, et al. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of The National Cancer Institute of Canada Clinical Trials Group. *European Journal of Cancer*. Jan 2005;41(2):280-287.
53. Wiklund I. Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life. *Fundamental and Clinical Pharmacology*. 2004;18(3):351-363.
54. Fletcher A, Gore S, Jones A, et al. Quality of life measures in health care: II. Design, analysis, and interpretation. *BMJ* 1992;305:1145-1148.
55. Hopwood P, Stephens RJ, Machin D. Approaches to the analysis of quality of life data: Experiences gained from a Medical Research Council Lung Cancer Working Party palliative chemotherapy trial. *Quality of Life Research*. 1994;3:339-352.
56. Efficace F, Bottomley A, Osoba D, et al. Beyond the development of health-related quality-of-life (HRQOL) measures: a checklist for evaluating HRQOL outcomes in cancer clinical trials--does HRQOL evaluation in prostate cancer research inform clinical decision making? *Journal of Clinical Oncology*. Sep 15 2003;21(18):3502-3511.
57. Staquet M, Berzon R, Osoba D, Machin D. Guidelines for reporting results of quality of life assessments in clinical trials. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*. 1996;5(5):496-502.

58. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology*. Jan 1998;16(1):139-144.
59. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *British Medical Journal*. 1995;311:1145-1148.
60. Fairclough D. *Design and Analysis of Quality of Life Studies in Clinical Trials*. New York: Chapman & Hall; 2002.
61. Osoba D, King MT. Interpreting quality of life in individuals and groups: meaningful differences. In: Fayers PM, Hays RD, eds. *Assessing quality of life in clinical trials: Methods and practice*. 2nd ed. Oxford, UK: Oxford University Press; 2005:243-257.
62. Jaeschke RJ, Singer J, Guyatt G. Measurement of health status: ascertaining the minimally clinically important difference. *Control Clin Trials*. 1989;10:407-415.
63. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Quality of Life Research*. 1996;5(6):555-567.
64. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *Journal of Clinical Epidemiology*. 1996;49:1215-1219.
65. Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. 2 ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
66. King MT. Cohen confirmed? Empirical effect sizes for the QLQ-C30. *Quality of Life Research*. 2001;10(3):278.
67. Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. *Evaluation & the Health Professions*. Jun 2005;28(2):172-191.
68. Wyrwich KW, Bullinger M, Aaronson N, et al. Estimating clinically significant differences in quality of life outcomes. *Quality of Life Research*. Mar 2005;14(2):285-295.
69. Hjermstad MJ, Fayers PM, Bjordal K, Kaasa S. Using reference data on quality of life--the importance of adjusting for age and gender, exemplified by the EORTC QLQ-C30 (+3). *European Journal of Cancer*. 1998;34(9):1381-1389.
70. Rapley M. *Quality of life research: A critical introduction*. London, Thousand Oaks California: Sage Publications; 2003.

16 Suggested Reading

General texts

- Fayers PM, Machin D. *Quality of life: Assessment, analysis and interpretation*. Chichester, UK: John Wiley and Sons, Ltd; 2000.
- O'Connor R. *Measuring quality of life in health*. Elsevier/Churchill Livingstone; 2004.
- Spliker B (Ed). *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia: Lippincott-Raven Publishers; 1996.

HRQOL in radiation oncology

- Bruner DW, Movsas B, Konski A, et al. Outcomes research in cancer clinical trial cooperative groups: the RTOG model. *Quality of Life Research*. Aug 2004;13(6):1025-1041.

Choosing a HRQOL survey

- Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. *Quality of Life Research*. 1993;2:441-449.
- Robinson PG, Carr AJ, Higginson IJ. How to choose a quality of life measure. In: Carr AJ, Higginson IJ, Robinson PG, eds. *Quality of Life*. London: BMJ Books; 2003:88-100.
- Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*. 2002;11:193-205.

Analysis and interpretation of HRQOL data

- Efficace F, Bottomley A, Osoba D, et al. Beyond the development of health-related quality-of-life (HRQOL) measures: a checklist for evaluating HRQOL outcomes in cancer clinical trials--does HRQOL evaluation in prostate cancer research inform clinical decision making? *Journal of Clinical Oncology*. Sep 15 2003;21(18):3502-3511. Available online at <http://jco.ascopubs.org/cgi/content/full/21/18/3502>
- Fayers PM, Hays RD (Eds). *Assessing quality of life in clinical trials: Methods and practice*. New York, NY: Oxford University Press; 2005.
- Fayers PM, Hopwood P, Harvey A, Girling DJ, Machin D, Stephens R. Quality of life assessment in clinical trials--guidelines and a checklist for protocol writers: the U.K. Medical Research Council experience. MRC Cancer Trials Office. *European Journal of Cancer*. Jan 1997;33(1):20-28.
- Fletcher A, Gore S, Jones A, et al. Quality of life measures in health care: II. Design, analysis, and interpretation. *BMJ*. 1992;305:1145-1148.
- Movsas B. Quality of life in oncology trials: a clinical guide. *Seminars in Radiation Oncology*. 2003; 13(3): 235-47.
- Osoba D, Bezjak A, Brundage M, et al. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of The National Cancer Institute of Canada Clinical Trials Group. *European Journal of Cancer*. Jan 2005;41(2):280-287. Available online at http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T68-4DV1FFF-3&_user=115085&_coverDate=01%2F01%2F2005&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000008818&_version=1&_urlVersion=0&_userid=115085&md5=044d3c787cdd68dd1146608a94967756
- Staquet M, Berzon R, Osoba D, Machin D. Guidelines for reporting results of quality of life assessments in clinical trials. *Quality of Life Research*. 1996;5(5):496-502.
- Wiklund I. Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life. *Fundamental and Clinical Pharmacology*. 2004;18(3):351-363.
- Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T, Sloan JA. Estimating clinically significant differences in quality of life outcomes. *Quality of Life Research*. 2005, 14(2):285-295.

17 Online Resources

Databases of HRQOL measures

Behavioral Measures Core <http://www.behavioralmc.org>

Freely accessible database that includes limited information about a range of measures together with electronic samples of the instruments themselves and abstracts of key references where permission has been obtained.

PROQOLID <http://www.proqolid.org/>

The most comprehensive online database of QOL measures. Limited information about each measure is available free of charge; more detailed information and samples of each instrument is available via subscription.

Key measures

EORTC Quality of Life Group <http://www.eortc.be/home/qol/>

Includes samples of instruments from the most widely used HRQOL measurement suite in cancer research. Also includes scoring manual, reference values, guidelines on assessing QOL in EORTC clinical trials, and information about how to become involved in the activities of the EORTC QOL Group.

FACIT <http://www.facit.org/>

Samples of instruments from the FACIT measurement system and information about administration and psychometric properties.

Other resources

International Society for Quality of Life Research (ISOQOL) <http://www.isoqol.org/>
ISOQOL provides educational outreach and collaborative support for HRQOL initiatives. It offers scientific publications (including the reviewed journal, Quality of Life Research) and an international conference programme.

Psycho-oncology Co-operative Research Group (PoCoG) <http://www.pocog.org.au/>
A member of the Cooperative Clinical Trials Groups of the Clinical Oncological Society of Australia (COSA), PoCoG is developing a Quality of Life Office to serve as a resource for Australian and New Zealand clinical trials groups. The Quality of Life Office will develop a library of instruments and relevant references together with detailed guidelines to aid trial managers in choosing an instrument and analysing and interpreting HRQOL data. There are also plans for the Quality of Life Office to engage in collaborative research with clinical trials groups in the future.

Appendix 1: Taken from Efficace et al, 2003, p. 3504⁵⁶ and reprinted with permission from the American Society of Clinical Oncology

Table 1. Minimum Standard Checklist for Evaluating HRQOL Outcomes in Cancer Clinical Trials

HRQOL Issue	Answer			Description
Conceptual				
A priori hypothesis stated	Yes	No	N/A*	Assessed whether authors had a predefined HRQOL end point and/or stated expected changes because of the specific treatment.
Rationale for instrument reported	Yes	No		Assessed whether authors gave a rationale for using a specific HRQOL measure.
Measurement				
Psychometric properties reported	Yes	No		Assessed whether a previously validated measure was used or psychometric properties were reported or referenced in the article.
Cultural validity verified	Yes	No	N/A†	Assessed whether the measure was validated for the specific study population.
Adequacy of domains covered	Yes	No		Assessed whether the measure covered, at least, the main HRQOL dimensions relevant for a generic cancer population and/or according to the specific research question.
Methodology				
Instrument administration reported	Yes	No		Assessed whether authors specified who and/or in which clinical setting the HRQOL instrument was administered.
Baseline compliance reported	Yes	No		Assessed whether authors reported the number of patients providing an HRQOL assessment before the start of treatment.
Timing of assessments documented	Yes	No		Assessed whether authors specified the HRQOL timing of assessment during the trial.
Missing data documented	Yes	No		Assessed whether authors gave some details on HRQOL missing data during the trial.
Interpretation				
Clinical significance addressed	Yes	No		This refers to the discussion of HRQOL data being clinically significant from a patient's perspective and not simply statistically significant.
Presentation of results in general	Yes	No		Assessed whether authors discussed the HRQOL outcomes, giving any comments regardless of the results (either expected or not).

Abbreviation: HRQOL, health-related quality of life.

*If a study explicitly states an exploratory HRQOL evaluation.

†If the HRQOL measure is validated in the same population as the one of the trial.