

Jeff A. Sloan  
Marlene H. Frost  
Rick Berzon  
Amylou Dueck  
Gordon Guyatt  
Carol Moinpour  
Mirjam Sprangers  
Carol Ferrans  
David Cella  
**Clinical Significance Consensus  
Meeting Group**

## The clinical significance of quality of life assessments in oncology: a summary for clinicians

Received: 27 January 2006  
Accepted: 2 May 2006  
Published online: 23 June 2006  
© Springer-Verlag 2006

List of Clinical Significance Consensus Meeting Group contributors to this article: Neil Aaronson, Ph.D., Division of Psychosocial Research, Cancer Institute, Amsterdam, The Netherlands; Ivan Barofsky, Ph.D., Quality of Life Institute, East Sandwich, MA, USA; Amy Bonomi, MPH, MacColl Institute for Healthcare Innovation, Seattle, WA, USA; Monika Bullinger, Ph.D., University of Hamburg, Hamburg, Germany; Joseph Cappelleri, Ph.D., Global Research and Development, Pfizer Incorporated, Groton, CT, USA; Diane Fairclough, Ph.D., University of Colorado Health Sciences Center Denver, CO, USA; Larry Gorkin, Ph.D., Pfizer Incorporated, New York, NY, USA; Ron Hays, Ph.D., UCLA Division of GIM and Health Services Research, UCLA Department of Health Sciences, UCLA AIDS Institute, Los Angeles, CA, USA; Patrick Marquis, M.D., MBA, Mapi Values, Boston, MA, USA; Tim Moynihan, M.D., Department of Medical Oncology, Mayo Clinic Rochester, Rochester, MN, USA; Geoff Norman, Ph.D., McMaster University, Hamilton, Ontario, Canada; David Osoba, M.D., QOL Consulting, West Vancouver, BC, Canada; Donald Patrick, Ph.D., Department Health Services, University of Washington, Seattle, WA, USA; Dennis Revicki, Ph.D., MEDTAP International Incorporated, Bethesda, MD, USA; Theresa Rummans, M.D., Department of Psychiatry, Mayo Clinic Rochester, Rochester, MN, USA; Charles Scott, Ph.D., American College of Radiology, Philadelphia, PA, USA; Tara Symonds, Ph.D., Outcomes Research, Pfizer, Global Research and Development, Sandwich, Kent, UK; Claudette Varricchio, Ph.D., RN, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD, USA; Gilbert Wong, M.D., Department of Anesthesiology, Mayo Clinic Rochester,

Rochester, MN, USA; Albert Wu, M.D., Department of Health Policy and Management, School of Hygiene and Public Health, Johns Hopkins University, Baltimore, MD, USA; Kathleen Wyrwich, Ph.D., Saint Louis University, St. Louis, MO, USA.

J. A. Sloan (✉) · A. Dueck  
Department of Health Sciences  
Research, Mayo Clinic  
and Mayo Foundation,  
200 First Street SW,  
Rochester, MN, 55905, USA  
e-mail: jsloan@mayo.edu  
Tel.: +1-507-2849985  
Fax: +1-507-2662478

M. H. Frost  
Women's Cancer Program,  
Mayo Clinic and Mayo Foundation,  
Rochester, MN, USA

R. Berzon  
Corporate Health Economics and  
Market Access, Boehringer Ingelheim  
Pharmaceuticals, Incorporated,  
Ridgefield, CT, 06877, USA

*Present address:*  
R. Berzon  
US Agency for International  
Development, Bureau for Global  
Health, Office of HIV/AIDS,  
Washington, DC 20523, USA

G. Guyatt  
Department of Clinical Epidemiology  
and Biostatistics, McMaster University  
and Health Sciences Center,  
Hamilton, ON, Canada

C. Moinpour  
Southwest Oncology Group Statistical  
Center, Fred Hutchinson Cancer  
Research Center,  
Seattle, WA, USA

M. Sprangers  
Department of Medical Psychology,  
Academic Medical Center,  
University of Amsterdam,  
Amsterdam, The Netherlands

C. Ferrans  
College of Nursing,  
University of Illinois at Chicago,  
Chicago, IL, USA

D. Cella  
Center on Outcomes,  
Research and Education,  
Evanston Northwestern Healthcare  
and Northwestern University,  
Chicago, IL, USA

**Abstract Background:** A series of six manuscripts with an introduction appeared in the Mayo Clinic Proceedings, based upon the collective effort of 30 individuals with an interest and expertise in assessing the clinical significance of quality of life (QOL) assessments. The series of manuscripts described the state of the science of QOL assessments in oncology clinical research and practice and included extensive literature and theoretical justification for the continued inclusion of QOL in oncology clinical research and practice. **Objectives:** The purpose of this paper is to produce a summary of these articles and to supplement these works with additional information that was gleaned from subsequent meetings and discussions of this material. The primary aim of this paper is to present a cogent and concise description for clinicians to facilitate the incorporation of QOL assessments into oncology clinical research and practice. The theoretical discussion is supplemented with an example of how the various ideas can be operationalized in an oncology clinical trial.

**Keywords** QOL · Oncology · Cancer

## Introduction

Quality of life (QOL) was studied within both clinical research and practice settings. In fact, the application of QOL assessments within oncology research has seen a steady and marked increase in recent years [43, 49]. Clinical researchers, for the most part, have been convinced that QOL is a valid and important aspect of the patient's condition and QOL assessment should be included in many aspects of research. However, integration of QOL into the practice setting has been more limited. This is in part due to the fact that practicing clinicians outside the research setting have insufficient knowledge of how to interpret and apply QOL findings. Clinicians who provide clinical care too often do not have access to methodological experts to better comprehend the meaning of QOL data and, perhaps as a consequence, are not aware of the integrative complexities of research and practice. This situation has led to an increasing frustration and skepticism with respect to incorporating QOL into clinical practice, and affected its more extensive use within clinical research [14]. Researchers within the QOL community have come under considerable pressure to make the discipline's methods accessible and understandable to the typical clinician so that additional information regarding the patient can be made available within practice and within research settings. The present initiative was undertaken in an attempt to meet this need.

The Clinical Significance Consensus Meeting Group consisted of 30 QOL research experts who assembled at the Mayo Clinic in Rochester, Minnesota, on October 6 and 7, 2000. This group produced six manuscripts that focused on the clinical significance of QOL assessments within oncology research and practice. The manuscripts were intended to serve as a resource for individuals conducting cancer QOL research, for clinicians considering incorporation of QOL assessments in the treatment of cancer patients, and for stimulating further discussion and research in QOL assessments. The purpose of this paper is to provide, primarily for clinicians, an informative overview of the six manuscripts as an introduction to interpreting QOL scores and understanding their clinical significance. Readers are referred to the original manuscripts for in-depth details and discussion on any particular topic.

In this paper, a variety of issues are addressed surrounding the clinical significance of QOL data. This paper begins with a discussion of approaches for determining clinical significance and for interpreting group vs individual QOL data. This is followed by a discussion of the usefulness of single-item vs multiple-item QOL assessments; the patient, clinician, and population perspectives of clinical significance; and the assessment of changes that occur over time. Finally, practical considerations are reviewed for communicating QOL measurement results to clinicians and patients.

Throughout the paper, an example is given to illustrate the assessment of clinical significance of QOL assessments.

The example employs a hypothetical clinical trial that compares two competing chemotherapy agents for the treatment of metastatic lung cancer. Lung cancer is an appropriate example because it is a disease in which QOL issues are acute and require further study [29]. The methods discussed for this trial vary in their generalizability to cancer-specific disease sites and treatments. Nonetheless, it is the authors' belief that the basic tenets of QOL methodology and, in particular, clinical significance hold within this cancer type, and therefore, lung cancer can serve as an illustrative example for clinicians.

There was much recent activity attempting to overcome the logistics of implementing QOL in oncology clinical practice [15, 18, 19, 26, 59, 60, 62]. The researches, to date, showed that patients appreciate the opportunity to discuss QOL issues with their clinicians [19] and clinicians benefit in knowing more about the QOL of their patients [17, 23, 34]. Despite this seemingly successful implementation of QOL assessment into clinical practice, it was not clear that practice patterns were changed at all [56, 60]. The missing aspect for QOL assessment implementation at present is an indication of the clinical pathways once a notable QOL score was observed [18, 58]. There are ongoing research studies that are addressing this very issue [57]. Ultimately, these barriers will no doubt be overcome and the utility of implementing QOL assessments into clinical practice will be demonstrated. While the examples in this paper focus more on clinical trials rather than clinical practice, the methods involved apply equally to both environments. The emphasis on clinical trials is due to the fact of the evolving role of QOL assessment in clinical practice [4].

---

## Methods to explain the clinical significance of quality of life measures [28]

The first issue to be discussed regarding clinical significance is the idea that a significant  $p$  value has intrinsic meaning, clinical or otherwise. It has long been recognized that statistical significance and clinical significance are not the same thing. Many clinicians still determine the importance of change by the level of statistical significance as reported by  $p$  values. However,  $p$  values do not indicate whether a particular finding has clinical implications. A  $p$  value indicates only whether the results of the study are more or less likely than would be found purely by chance.

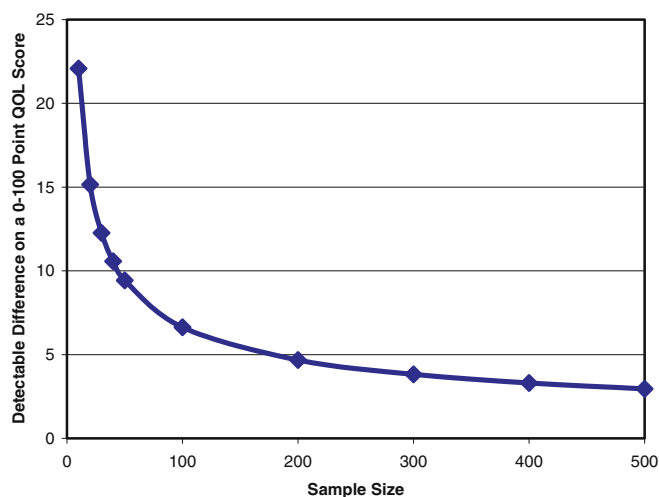
For example, if the proposed clinical trial involved 1,000 patients (500 per arm), the large sample size would provide substantial statistical power to detect very small, perhaps meaningless, differences between the treatments under study (i.e., differences between groups of less than 1 point on a 100-point QOL assessment). In most circumstances, such a difference would have little clinical meaning. Alternatively, if the sample size were 50 per arm instead of 500, then a treatment difference on the order of 15 points on the 100-point scale would attain a significant  $p$  value, which in

all likelihood would have clinical importance. Figure 1 plots the relationship between sample size and the size of the difference that is detectable with 80% power when one is using a two-sample  $t$  test to compare group means. The well-known relationship is displayed wherein as sample size increases, the size of the difference that one can detect decreases. More specifically, with 50 patients per group the  $t$  test can detect a difference of roughly 10 points on a 100-point scale, whereas a sample of 400 patients per group can detect a difference of 3 points on a 100-point scale.

If the researcher begins with a  $p$  value to assess clinical significance, he/she only needs to provide an a priori cutoff point for the assessment under study as to what will be considered clinically meaningful. The source of this definition may be prior research, patient perspective, or clinician experience. Once that cutoff is set, the trial may be sized (number of patients) to have appropriate power and error rates to detect the clinically significant difference.

Clinicians traditionally develop a sense of how to interpret clinical measures by their repeated use and an intuitive sense of how changes in these measures relate to other variables. For instance, physicians experienced with testing lung cancer patients know how a change in forced expired volume of 500 ml will likely impact patient function. Similarly, experienced cardiologists have a good sense of the likely impact of a change in cardiac ejection fraction of 10%. For QOL measures not in day-to-day use in clinical practice, clinicians have not developed these intuitive estimates. Investigators must use other strategies to help the clinician interpret QOL endpoints.

One can frame the problem as an issue of interpretability: What changes in QOL scores correspond to trivial, small, moderate, or large patient benefit? If a person improves by 5 points in a measure of emotional functioning, will it mean that he is perceived as being happier by his family, will miss



**Fig. 1** Size of average difference detectable using a two-sample  $t$  test with 80% power and a 5% significance level between two groups for a 0- to 100-point QOL score as sample size increase

less work, and no longer have to take antidepressant medication? Or will no such changes occur? If a patient with chronic lung disease improves by 5 points in a physical test, will she now be able to climb a flight of stairs comfortably, keep up with her spouse when they go for a walk, and resume playing with her grandchildren? Or will she remain incapacitated by exertional dyspnea?

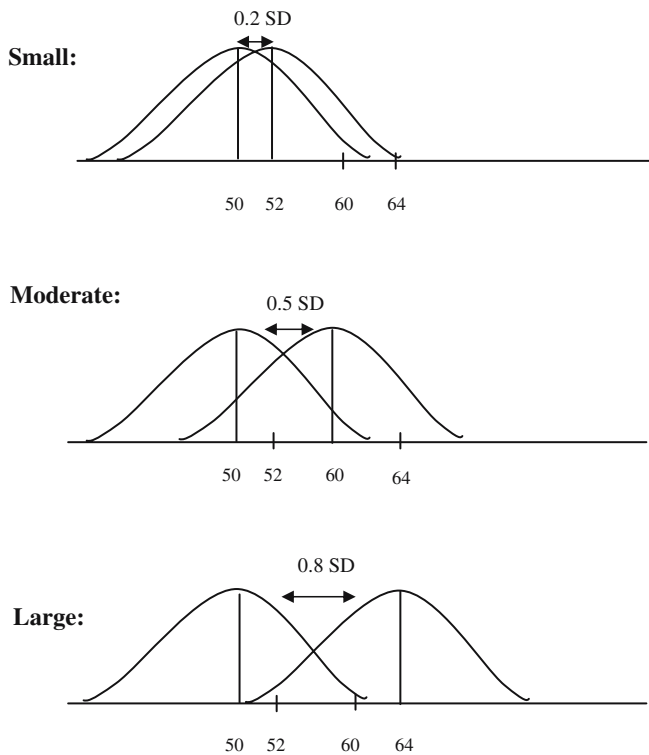
Strategies for establishing the interpretability of QOL measures can be classified as *anchor-based* or *distribution-based* [35]. Anchor-based measures require an independent standard (or anchor) that is itself interpretable and that is at least moderately correlated (i.e., have at least half of its variance in common) with the instrument being explored. For example, one anchor for interpreting QOL measures is how the QOL score relates to patients' reports of whether they are improved or unimproved and, if improved, whether the magnitude of the change is small, medium, or large. Such global ratings of change are likely to be at least moderately correlated with QOL measures intended to tap similar domains. Other anchors can include toxicity assessments and performance status.

Distribution-based methods relate observed differences to the underlying distribution of scores. Hence, distribution-based methods rely heavily on probability and statistical theory. Investigators may express differences in terms of statistical entities such as between-group SD units [32], within-person SD units, and the SEM [63]. One suggestion is to consider a between-group SD of 0.2 as a small difference, 0.5 as a moderate difference, and 0.8 as a large difference [12, 44, 46] (see Fig. 2 for a conceptualization).

The advantage of distribution-based methods is that the values are easy to generate. Whatever the study, there will always be one or more measures of variability available. This contrasts with the work needed to generate an anchor-based interpretation that can require additional patient-reported data. A limitation is that clinicians struggle with the meaning of distribution-based methods. For instance, if the magnitude of effect is 0.3 SD units, this means that the difference between treatment groups is 30% of the variability inherent in the distribution.

No approach to interpretability is perfect. Attempts to link anchor-based and distribution-based methods may prove helpful [9, 40]. Use of multiple statistical strategies will likely enhance the interpretability of any particular instrument because the reliance on a single approach and its inherent biases and assumptions is removed.

The lung cancer trial example can illustrate concepts in the previous paragraph. Consider the steps involved in the challenge of obtaining an estimate for a clinically meaningful difference in QOL scores in the hypothetical clinical trial comparing two chemotherapy regimens for patients with lung cancer. Suppose the clinician wants to use the lung cancer-specific Functional Assessment of Cancer Therapy–Lung (FACT-L) [6] measure to assess QOL. Based on their experience, the assessment's developers, Cella et al. [8], suggest that a five-point difference in the



**Fig. 2** Use of a SD to determine small, moderate, and large differences between group means

trial outcomes index (TOI), a FACT-L subscale, is clinically meaningful. If the FACT-L TOI score is a feasible primary measure of patient status from the clinical standpoint, then sample size need not be estimated for additional outcomes. The study can be powered to detect the defined difference of 5 points. Meaningful differences in patients with nonsmall cell lung cancer were also demonstrated with the Quality of Life Questionnaire 30 core items (QLQ-C30) [37] and the Lung Cancer Symptom Scale (LCSS) [29].

Many QOL questionnaires for lung cancer patients do not have published estimates for clinically meaningful differences. If one of these questionnaires is used, a method for defining a clinically meaningful difference needs to be incorporated into the study. An anchor-based approach would tie the QOL scores to, say, the Karnofsky performance status [31, 38], the incidence of Grade 3 toxicity under the common toxicity criteria, or a patient's ability to perform or not perform a simple physical test. The patients can also be asked directly if they felt a noticeable change (small, moderate, or large) in their QOL since their last visit. Then, an estimate of how much QOL scores need to change for patients to say that they perceived a difference can be computed by correlating these answers with the observed changes in QOL scores. One needs to be cautious about the generalizability of these data. For instance, estimates need to be applied with care if the

proposed study is not going to employ the same patients, assessment, and clinical setting.

A distribution-based approach would simply declare a difference of a certain effect size (often a 1/2 SD of assessment scores in the same patient population) [12, 41, 51] as clinically significant and then the trial is powered to detect that difference. For example, if the LCSS is used, the literature provides a typical SD of LCSS scores in advanced cancer patients as 14 points [29]. Thus, any difference of one half, such as  $1/2 \cdot 14 = 7$  points, is considered clinically significant. If a reasonable estimate of the SD from the literature does not exist, an approach recommended by Sloan et al. [51] indicates that a ballpark estimate of the SD in the absence of further data is 16% of the theoretical range of the QOL assessment.

### Group vs individual approaches to understanding the clinical significance of differences or changes in quality of life [7]

Many conclusions in clinical research are based on grouped or averaged results. The average effect of a treatment does not adequately reflect that treatment's effect for an individual [20]. However, the clinician has the daily task of applying this type of information to the individual patient. He or she also has the task of interpreting the significance of some particular findings for the individual patient. How can this best be done?

First, data should be generated by assessments that meet guidelines of acceptable measurement [2, 10, 50] similar to accepted standards of practice for medical care. Clinicians may not be familiar with these guidelines, but it is important to be aware that such guidelines exist and that they need to be applied when choosing a QOL assessment.

Second, there exist differences between a deductive and an inductive approach to establishing clinical significance. The deductive approach transitions from group to individual-level data while the inductive approach transitions from individual- to group-level data. One approach does not appear to be consistently superior to the other. In fact, using both approaches can provide greater confidence in the results (i.e., the measures in use ask about the important and relevant constructs). Turning to the example from the previous section once again, if the estimate of a 1/2 SD for the FACT-L TOI is 5 points, the study can be powered to detect a five-point difference between the two lung cancer treatment groups. Hence, a significant  $p$  value under these circumstances is clinically meaningful, as well.

Without further information, the five-point difference can also be applied in the monitoring of an individual patient over time. Some argue that the variability for an individual over time may differ from the variability of the group. Optimally, data in the literature showing individual variability over time can be employed. Unfortunately, such data are not often available. One way to transform group

comparison results to longitudinal comparisons is to divide the group detectable effect by  $\sqrt{2}$  (the mathematical details of this approach can be found in Cohen [12], pp. 46–47, 71–73). For the example, a clinically significant difference for an individual over time is  $5/\sqrt{2}$  or 3.5 points. It is questionable whether a 5- or 3.5-point difference is worth arguing about. If one wishes to be conservative, the five-point cutoff is usable. As with all of the estimates, this cutoff is intended as a guideline, not a rule. Any clinical indicator is taken in context with other patient information. If further information is available that would indicate a modification of the five-point criteria, it can be readily applied provided that it is scientifically justifiable. For example, if the example is limited to advanced lung cancer patients, it might be reasonable to say that smaller improvements are clinically important because improving the QOL of advanced lung cancer patients, even a small amount, might be clinically important.

This is just one way to approach the application of group data to individual patients. There is still work to be done to find a method that is generally acceptable. The variability of a particular individual may or may not be different from the variability of an aggregated sample of individuals, much in the same way that survival data for groups needs to be applied to individual patients with caution.

---

### **Assessing the clinical significance of single items relative to summated scores [48]**

A single-item index is a single question intended to measure overall QOL or a characteristic of QOL. A multiple-item index is an index with many questions focusing on aspects of QOL. Each type of index has advantages and disadvantages depending on the clinical setting. A well-defined purpose of assessment within a specified clinical setting is required to properly weigh the advantages and disadvantages in deciding whether to use a single- or multiple-item index to assess QOL. However, the two types of indices are not mutually exclusive and may be complementary and/or confirmatory when used together in the clinical setting.

One example of a single-item index is the Spitzer Uniscale [52], a single item intended to assess overall QOL using the statement, “Please rate your overall quality of life,” with a response from 0 to 100 derived from a number circled or an X placed on a line or horizontal bar. Single-item indices are simple, easy to administer, and minimize patient burden. They have a high rate of completed responses and are operationally efficient in terms of data entry and analysis. Single-item indices measuring overall QOL show validity by high correlations with related multiple-item or other single-item measures of overall QOL and exhibit responsiveness to change by showing changes that are expected and discernible over time. For example, the

Spitzer Uniscale was shown to be strongly correlated with and more responsive than the multiple-item Functional Living Index–Cancer [45] in measuring QOL of advanced cancer patients [47]. A single-item index is suitable for specific, unambiguous symptoms such as, “Have you been constipated in the past week?” from the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 [1, 27, 33]. A disadvantage of single-item indices is that such indices do not provide a detailed profile of a patient’s QOL so that the clinician can better understand what factors might be driving the change in QOL.

A multiple-item index is a set of questions on a QOL questionnaire intended to measure a specific characteristic of QOL. Responses on the multiple items are usually summed in an index to arrive at a single score. For example, a multiple-item index measuring emotional functioning may ask questions about anxiety, irritability, nervousness, and depression, as opposed to the single question, “How would you rate your emotional health?” As complexity of the characteristic of QOL increases, the number of questions to adequately assess the characteristic increases. However, when the characteristic becomes too complex, such as overall QOL, a single-item index may be more appropriate. Another advantage of multiple-item indices is that individual errors on multiple items when summated tend to average out to zero. This theoretical feature of multiple-item indices increases score reliability, validity, and precision (for definition of these terms, see Nunnally and Bernstein [42]). Disadvantages of multiple-item indices include increased patient burden, potentially irrelevant questions, added difficulty in analysis, and incompleteness in patient responses. Using items on the same index for missing scores can mitigate this last disadvantage.

The clinical objective should drive the selection of single-item vs multiple-item indices. If the clinical focus is assessment of one aspect of QOL or general overall QOL, a single question might suffice. If a detailed profile of QOL is required, a multiple-item index may be necessary. For example, if the clinical focus is to identify fatigue in patients, a single question may be all that is necessary. However, if a detailed profile of fatigue is needed to recommend treatment, a multiple-item index may be more appropriate.

Regardless of the choice of single-item vs multiple-item indices, the selected index should possess several properties. First, the QOL aspect measured should be consistent with the clinical objective. Second, the selected index should have been tested previously in a comparable population and shown to detect differences over time and between subgroups of patients in the same population. Third, the selected index should be responsive to change and should not exhibit floor and ceiling effects (all patients being scored uniformly low or high, respectively). For example, an index developed for measuring QOL in a healthy population may exhibit a floor effect in terminally ill cancer patients—low scores for all such patients. Lastly, there should be an a priori definition associated with the

selected index of what constitutes a clinically significant change.

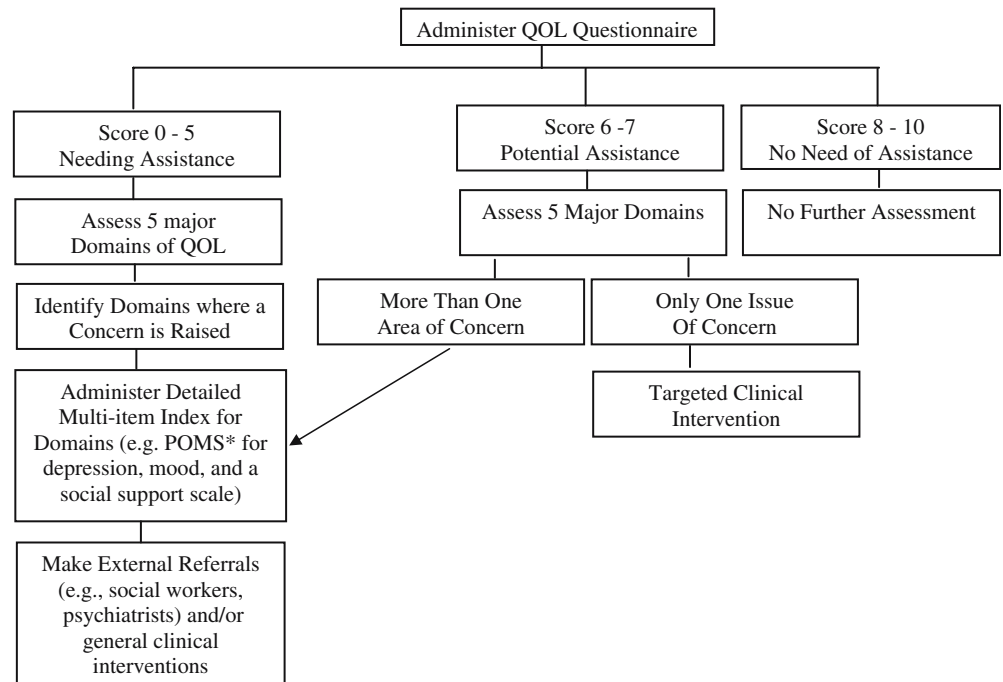
In the clinical setting, single-item and multiple-item indices can be incorporated to screen and treat patients. A clinical triage appears in Fig. 3 involving both single-item and multiple-item indices. The use of single-item or multiple-item indices is not a substitute for a complete psychological consultation, but can help ensure that limited consultative resources are applied in an effective manner. An index could optimally identify risk factors that need more specific diagnostic assessment and intervention. In general, indices can be efficient initiators of dialog between clinicians and patients by highlighting QOL issues that need to be dealt with and eliminating QOL issues not of immediate concern.

Using the example again, many questions should be asked in selecting a QOL assessment. Is it important to know just the overall QOL of patients with advanced lung cancer or do we need more detail? A typical phase III study is interested in more than just global QOL assessments, although they can serve a purpose, as well. In most settings, it would be reasonable to use one of the standard instruments such as the FACT [5], EORTC assessments, or a lung cancer-specific instrument such as the LCSS and then supplement these with any items that were specific to the conditions or treatments involved in the clinical trial proposed. For example, in a recent phase III colorectal cancer trial, the standard Symptom distress scale [36] was supplemented by a number of items to capture information on the idiosyncratic toxicities that were known to be related to one of the treatments. Because it was impossible to find any

assessment that would precisely target the toxicity this drug produced, a new single item was created to assess the degree of overall symptom distress associated with the particular toxicity produced. The new item used the format of the Brief Pain Inventory [11] to assess toxicity on a scale from 0 to 10. Note that adequate investigation into the properties of a new unvalidated assessment tool is necessary.

Returning to the lung cancer clinical trial example, a useful exercise is to list the toxicities expected to appear during the trial and then survey established assessments to see if the majority of items in the assessment are relevant. Using the 36-item short form in this patient population is likely a poor choice because it does not include enough items to distinguish the poor performance and deteriorating health of advanced lung cancer patients over the course of the study [61]. A short list of ten or so global items may be the best choice to minimize patient burden supplemented with a subscale from an existing assessment that focused on fatigue, as this patient population is expected to have specific problems in this area. The key is to fit the assessment to the research/clinical question and patient population, not the other way around. Asking potentially irrelevant questions increases patient burden and the likelihood of an insignificant finding or type II error because the assessment chosen is not sufficiently specific to be sensitive to the important changes happening during the study or treatment regimen. In addition, new questionnaires lack documentation of measurement properties. A new, untested measure may not be sufficiently sensitive to specific treatment arm differences.

**Fig. 3** Example of a clinical triage based on response to the question, “On a scale of 0 to 10, what is your overall quality of life?” (\*POMS=Profile of Mood States [16])



### **Patient, clinician, and population perspectives on determining the clinical significance of quality of life scores [22]**

There are three major perspectives that can be used to determine the clinical significance of QOL scores: the patient, clinician, and population perspectives. Each perspective can produce a different outcome in terms of what is considered clinically significant. Different stakeholders judge the type and magnitude of change using their own standards and values.

The first perspective is that of the patient. There is general agreement that patients can best judge their own QOL, and may be the best able to judge whether a change in QOL is clinically significant. This determination is dependent on the patient's perception that a change is beneficial or detrimental enough to potentially prompt a request for a change in treatment. This may be easily assessed by asking patients to rate their QOL on a 0–10 scale with zero being poor and 10 being optimal QOL. Once a clinician has obtained the patient's score, it can then be used in the clinical setting as a basis for further discussion. The patient can further be asked what has helped him/her maintain and what has negatively affected QOL. Some have argued that the patient's perspective is the only one that matters and that if a patient declares an improvement or lack of improvement, who should argue? [21]

The second perspective is that of the clinician. From the perspective of many clinicians, a clinically significant change in QOL is one that indicates the need for a change in patient treatment. Most clinicians address QOL issues informally in daily practice, focusing on the most observable aspects of QOL, such as symptoms and physical functioning. These aspects of QOL are most congruent with the traditional focus of medical care. Because clinicians are familiar with traditional biological and physical measures, these are often used informally to derive an impression of QOL. In fact, standard clinical parameters, such as hematocrit and hemoglobin, were found to be associated with important differences in overall QOL scores [13]. A counter example is the lack of association between a clinical measure of anemia (hemoglobin) and patient-reported fatigue [3, 25]. In addition, standard clinical parameters may not provide enough information about a patient's QOL. It is generally accepted that the concept of QOL includes psychological and social aspects of life, as well as physical aspects. In fact, this broader notion of QOL was found to be predictive of survival in cancer patients and to have prognostic value in terminal illness [24, 30, 55]. These findings indicate that a broader notion of QOL can be useful for determination of clinically significant changes in clinical practice.

From the third perspective, clinical significance from the population perspective is determined largely in terms of "opportunity costs" whereby the provision of resources is measured with respect to benefits to society. This determi-

nation is based on the values of the society, such as the utilitarian value of returning people to full time employment. Society has an economical and ethical significance judged in terms of whether the outcomes are deemed worthy of resource allocation. The justification for this perspective is that society pays for much of health care, thus, society's values should determine the economic priorities. Government policymakers commonly use the population perspective, though values other than economic ones are also applied.

Returning to the lung cancer example, the three perspectives are those of the lung cancer patient, the oncologist, and the insurer. Asking the lung cancer patient what important changes he feels have occurred in his QOL might, for example, raise issues about breathing, pain, fatigue, depression, and overall QOL. Hence, these symptoms need to be included in our hypothesized phase III trial comparing two agents. The oncologist may focus more on symptoms that reflect pathologic processes. The insurer might be interested in the relative importance of these QOL domains for functional outcomes and, hence, request that some form of assessment regarding ability to perform usual activities be included in the trial so cost-benefit analyses could be undertaken. All three perspectives indicate that QOL assessments can contribute to a better understanding of exactly what is going on with the patient and what action needs to be taken. The good news is that a single assessment can incorporate all perspectives and provide data to support ameliorative interventions.

---

### **Assessing meaningful change in quality of life over time: a user's guide for clinicians [53]**

A crucial task for clinicians is to recommend the most optimal treatment for their patients. They need to consider how alternative treatments might affect their patients' QOL during treatment and beyond. Because this information needs to be retrieved primarily from the literature, this summary is intended to help clinicians critically assess and interpret longitudinal, patient-derived QOL results presented in the literature. This in turn can help clinicians use these data in the treatment decision process.

Detecting meaningful change depends on the adequacy of the research design, measurement quality, and data analysis. These study characteristics may be summarized by six key points that influence the clinical significance of score change (see Table 1). First, it is important that the QOL measure used is relevant, reliable, feasible, and responsive to change. Without these characteristics, the score has little meaning. The content areas that one would think about in making this determination include, among others, the following. Is the questionnaire appropriate given the research objective, the rationale for QOL assessment, and practical considerations (e.g., regarding respondent burden and the availability of different language versions)? Is the

questionnaire appropriate given the domains included and in light of the disease and population characteristics? Is the questionnaire reliable, feasible, and responsive to change? Finally, is the information reported in the article?

Second, timing and frequency of assessments are key issues. Timing can greatly influence the scores received. There are times during the natural course of a diagnosis that one would not expect QOL to be significantly affected. While at other times, QOL may be greatly affected. In terms of frequency, if assessments are too frequent, it is likely that the person may become sensitized to the instrument, completing it in a similar manner. If assessments are too few, one may fail to capture important changes. The following questions are important to ask. Is a baseline assessment included? Is QOL assessed at appropriate times for determining minimally important change given the natural course of the disease? Is QOL assessed long enough to determine a clinical effect, taking disease stage into account? Is QOL assessed at appropriate times to document treatment course, clinical events, and posttreatment effects? Are standard research design procedures followed (e.g., avoidance of respondent burden, collection of data before treatment or consultation)? Is the timing of the QOL assessments similar across treatment arms?

Third, it is important to know if the clinical study was adequately powered to answer the QOL question. To determine this, the following questions are helpful. Is the sample size appropriate for the research questions (e.g., by providing a power calculation)? Is the rationale and/or source for the anticipated magnitude of effect (i.e., clinically meaningful difference over time) specified?

Most commonly, a wealth of QOL data is collected in a clinical trial. QOL questionnaires measure multiple domains and are administered at several points over time. The fourth key point is to determine whether multiple QOL outcomes and multiple time points are handled adequately. The following questions may help address these issues. Is the adopted approach of handling multiple outcomes described explicitly? Did the interpretation of the results take the problem of multiple outcomes into account? Are the longitudinal data presented in a meaningful and suitable way, enabling an overview of QOL changes over time? Are the data appropriately analyzed, including all time points?

The fifth key point is to determine whether alternative explanations can account for the observed change or lack of observed change. It should include the following questions.

Are dissimilar baseline characteristics adequately accounted for? Are missing data handled adequately? Is observed survival difference combined with QOL in evaluating change? Have possible changes in patients' QOL perspective over time been taken into account?

If a clinical trial shows statistically significant changes in QOL outcomes over time, the final key question is the extent to which these results are clinically meaningful. Does the article provide some guidance regarding the clinical importance of the observed change in QOL? To what extent is a statement of clinical importance appropriate and empirically warranted?

These ideas can be applied to the example of a lung cancer clinical trial. Through a literature review, QOL assessments are available that are appropriate for advanced lung cancer patients and were seen to be reliable, feasible, and responsive over a number of conditions (e.g., the LCSS, EORTC QLQ Lung Cancer Module (30) [1], and FACT-L).

Next, one must decide how often to measure patient QOL in the example study. Timing of assessments for lung cancer patients is limited to the disease process. The average survival time is roughly 8 months for patients with advanced lung cancer [39]. Linking the QOL assessments to the treatment cycle and characteristics is important. If the course of treatment is months, then assessing monthly is appropriate. More frequent assessment is only needed if there is reason to believe that important changes will occur more frequently and were not previously documented. Less frequent assessment is possible if we are only interested in, for example, the first month of treatment. Tying the evaluation to clinical visits will reduce the likelihood of missing data. Otherwise, follow-up telephone or mail contact with patients to facilitate compliance is advisable. For this study, assessment at baseline and monthly for the first 3 months is appropriate because it is reasonable to believe that if the treatment has not worked in 3 months then it is not likely to be successful. If the trajectory of the disease is of interest, an evaluation at 6 and 9 months can be added, knowing that there will be many deaths by these time points. One must also be aware that patients who stay on the study are generally the best performers in terms of QOL [46].

The power considerations for the lung cancer trial were discussed earlier in that the study would be powered 80% to detect a 1/2 SD (10 points on a 100-point scale) via a two-sample *t* test for comparing QOL across the two-sided

**Table 1** Key points that influence the clinical significance of score change

1. The QOL measure used is relevant, reliable, feasible, and responsive to change.
2. The timing and frequency of measurements are adequate.
3. The study from which information is taken to use in the clinical setting is powered adequately.
4. Multiple outcomes and multiple time points are handled adequately in analysis.
5. Alternative explanations that may account for the observed change or lack thereof are adequately addressed. These may include dissimilar baseline characteristics, missing data, observed survival difference, and change of patients' quality of life perspective over time.
6. Statistical significance is translated into clinically important change.

alternative and a 5% type I error rate. This would require a sample size of 64 evaluable patients per treatment arm. Inflation of the sample size by 10% is often used to account for anticipated cancellations or ineligibilities.

### **The clinical significance of quality of life results: practical considerations for specific audiences [54]**

Unique audiences—whether they be patients, research or practicing clinicians, policymakers, or others—do not always interpret QOL outcomes similarly. All the same, these outcomes can be used by clinicians to explain the effects of treatment alternatives to patients. The importance or clinical significance of these outcomes rests in the additional information they provide to both clinicians and patients, and the subsequent influence they have in the clinical decision-making process. The relevance of these data to a clinician and a patient is likely to depend on the relationship between the two, the seriousness of the illness, the needs of the particular patient, and the familiarity (and comfort level) of the treating clinician with these outcomes.

QOL information is not always of value in clinical decision-making. For example, if treatment options for noncurative conditions are similar in their efficacy but differ in their adverse effects, then QOL data may have a significant impact on the selection of the therapeutic treatment. If, on the other hand, a treatment option results in a cure for the underlying disease and has tolerable side effects, then QOL data are likely to carry less weight in the section of a therapy.

Any number of techniques used to convey relevant health-related information to patients can be used by clinical researchers and practitioners to communicate QOL outcomes. Statistical data, such as means, medians, or raw scores, afford the patient limited, if any, useful information. Conversely, providing information about the likelihood of QOL improvements, the meaningfulness of the improvements, and a comparison with the patient's current status provide valuable intelligence. This type of information can be delineated as the percentage of persons

experiencing benefits or the number of persons who experience adverse side effects. Table 2 provides an outline of various approaches by which the clinical significance of QOL outcomes can be communicated by a clinician to a patient. Some approaches in Table 2 are more reasonable with respect to clinical practice than other approaches. Further, the approaches in Table 2 do not represent all possible approaches.

Research and practicing clinicians are becoming familiar with approaches for communicating the clinical significance of QOL outcomes to their patients and better appreciating how this information can benefit them. Through shared review of QOL assessments and outcomes, clinicians are better able to determine which factors are most important to their patients to achieve both optimal treatment and optimal treatment adherence. The end result is a triple-win situation: On the part of the patient, greater satisfaction with the health care provided; on the part of the clinical care provider, greater appreciation of QOL outcomes and their clinical meaningfulness; and between the two, a closer partnership.

Applying these ideas to the lung cancer clinical trial example, clinical colleagues can be asked to review each item in each QOL assessment for relevance, clarity, and focus on the specific needs of the particular patient population. The assessments can then be selected based on this review. A pilot study with a small number of patients can test the chosen assessments to ensure that the assessments are indeed appropriate and that the proposed definitions for clinically meaningful changes are relevant. Having at least one individual serving as a resource for all QOL aspects often facilitates the logistics of patient assessment and maintains communication among the three perspectives. However, this may not be feasible in all situations.

### **Concluding remarks**

The basic message clinicians should take from this paper is that assessing the clinical significance of QOL assessments is vital and achievable. The tenets of good science apply to

**Table 2** Selected approaches for communicating the clinical significance of QOL outcomes to patients

| What patients may want to know about QOL information  | How QOL information can best be communicated to patients  |
|---|---|
| Whether a single aspect of patient QOL (e.g., physical function) is likely to improve with a specific treatment | With use of related studies, present a patient's baseline score relative to treatment intervals at which time QOL was assessed  |
| Whether QOL improvement with a specific treatment is meaningful   | Discuss aggregated outcomes for relevant profiles from disease-specific instruments used under related (treatment) circumstances, and compare these to "standard" clinical outcomes from related studies                        |
| An individual patient wants to know his or her level of QOL relative to where it was or could be                | Describe how research showed that scores of this nature sometimes indicate a need for help and that interventions are useful to reverse this situation  |
| A study indicates a significant impact on patient QOL. How can this be interpreted?                             | With the individual patient, review the proportion of patients that are helped by these interventions by using traditional clinical endpoints and compare these results to aggregated QOL outcomes and individual domain scores |

QOL assessment as they would in any other aspect of clinical investigation. Following the scientific methods described in this manuscript will ensure that the research has built-in clinical relevance. A significant  $p$  value, in and of itself, may or may not be evidence of a clinically meaningful result. However, incorporating clinical significance into the methodology of the trial at the design phase is the most important step for statistically significant results having clinical meaning. Too often QOL was an add-on to a trial after other elements were solidified. No scientific method performs at its best when it is included as

an afterthought or artificially molded to fit an unreasonable situation. QOL is no different.

It is hoped that the lung cancer clinical trial example provided throughout this article substantiates the conceptual content of the six manuscripts produced by the group of 30 researchers. The importance of providing concrete examples has spurred the group of 30 to commence a second set of forthcoming papers that will be example-driven in much more detail than what can be given here. Collectively, these works are intended to make QOL assessment feasible, understandable, and scientifically viable in oncology research and practice.

## References

- Bergman B, Aaronson NK, Ahmedzai S et al (1994) The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. *Eur J Cancer* 30A: 635–642
- Berzon RA (1994) Quality of life determinations in chronic conditions: perspectives of the pharmaceutical industry. In: Trimble MR, Dodson WE (eds) *Epilepsy and quality of life*. Raven, New York
- Brown DJ, McMillan DC, Milroy R (2005) The correlation between fatigue, physical function, the systemic inflammatory response, and psychological distress in patients with advanced lung cancer. *Cancer* 103:213–215
- Cella DF (1996) Quality of life outcomes: measurement and validation. *Oncology* 10:233–246
- Cella DF, Tulsky DS, Gray G et al (1993) The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol* 11:570–579
- Cella DF, Bonomi AE, Lloyd SR et al (1995) Reliability and validity of the Functional Assessment of Cancer Therapy–Lung (FACT-L) quality of life instrument. *Lung Cancer* 12:199–220
- Cella D, Bullinger M, Scott C et al (2002) Group versus individual approaches to understanding the clinical significance of differences or changes in quality of life. *Mayo Clin Proc* 77:384–392
- Cella D, Eton DT, Fairclough DL et al (2002) What is a clinically meaningful change on the Functional Assessment of Cancer Therapy–Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) study 5592. *J Clin Epidemiol* 55: 265–295
- Cella D, Eton DT, Lai JS, Peterman AH, Merkel DE (2002) Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. *J Pain Symptom Manage* 24:547–561
- Chassany O, Sagnier P, Marquis P et al (2002) Patient-reported outcomes: the example of health-related quality of life—a European guidance document for the improved integration of health-related quality of life assessment in the drug regulatory process. *Drug Inf J* 36:209–238
- Cleeland CS, Ryan KM (1994) Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singapore* 23:129–138
- Cohen J (1998) *Statistical power analysis for the behavioral sciences*, 2nd edn. Hillsdale, New Jersey
- Cortesi E, Gascon P, Henry D, Littlewood T, Milroy R, Pronzato P, Reinhardt U, Shasha D, Thatcher N, Wilkinson P (2005) Standard of care for cancer-related anemia: improving hemoglobin levels and quality of life. *Oncology* 68:22–32
- Cox DR, Fitzpatrick R (1992) Quality-of-life assessment: can we keep it simple? *J R Stat Soc* 155:353–393
- Cull A, Gould A, House A, Smith A, Strong V, Velikova G, Wright P, Selby P (2001) Validating automated screening for psychological distress by means of computer touchscreens for use in routine oncology practice. *Br J Cancer* 85:1842–1849
- Curran SL, Andrykowski MA, Studts JL (1985) Short form of the profile of mood states (POMS-SF): psychometric information. *Psychol Assess* 7:80–83
- Curt GA, Breitbart W, Cella D, Groopman JE, Horning SJ, Itri LM, Johnson DH, Miaskowski C, Scherr SL, Portenoy RK, Vogelzang NJ (2000) Impact of cancer-related fatigue on the lives of patients: new findings from the Fatigue Coalition. *Oncologist* 5: 353–360
- Davis K, Cella D (2002) Assessing quality of life in oncology clinical practice: a review of barriers and critical success factors. *J Clin Outcomes Manag* 9:327–332
- Detmar SB, Muller MJ, Schornagel JH, Wever LDV, Aaronson NK (2002) Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *JAMA* 288:3027–3034
- Donaldson GW, Moinpour CM (2002) Individual differences in quality of life treatment response. *Med Care* 40:III-39–III-53
- Frost MH, Sloan JA (2002) Quality of life measurements: a soft outcome—or is it? *Am J Manag Care* 8:S574–S579
- Frost MH, Bonomi AE, Ferrans CE et al (2002) Patient, clinician, and population perspectives on determining the clinical significance of quality-of-life scores. *Mayo Clin Proc* 77:488–494
- Funch DP (1988) Predictors and consequences of symptom reporting behaviors in colorectal cancer patients. *Med Care* 26:1000–1008
- Ganz P, Lee J, Siau J (1991) Quality of life assessment: an independent prognostic variable for survival in lung cancer. *Cancer* 67:3131–3135

25. Ganz PA, Moinpour CM, McCoy S (2004) Predictors of vitality (energy/fatigue) in early stage Hodgkin's disease (HD): results from Southwest Oncology Group (SWOG) study 9133. *J Clin Oncol* 22:14S (abstract 6546)
26. Greenhalgh J, Meadows K (1999) The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literary review. *J Eval Clin Pract* 5:401-416
27. Groenvold M, Klee MC, Sprangers MAG et al (1997) Validation of the EORTC QLQ-C30 Quality of Life Questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. *J Clin Epidemiol* 50:441-450
28. Guyatt GH, Osoba D, Wu AW et al (2002) Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 77:371-383
29. Hollen PJ, Gralla RJ, Kris MG et al (1999) Normative data and trends in quality of life from the Lung Cancer Symptom Scale (LCSS). *Support Care Cancer* 7:140-148
30. Justice AC, Aiken LH, Smith HL et al (1996) The role of functional status in predicting inpatient mortality with AIDS: a comparison with current predictors. *J Clin Epidemiol* 49:193-201
31. Karnofsky DA (1967) Clinical evaluation of anticancer drugs. *Gann Monogr Cancer Res* 2:223-234
32. Kazis LE, Anderson JJ, Meenan RF (1989) Effect sizes for interpreting changes in health status. *Med Care* 27:S178-S189
33. King MT (1996) The interpretation of scores from the EORTC Quality of Life Questionnaire QLQ-C30. *Qual Life Res* 5:555-567
34. Levinson W, Roter D (1995) Physicians' psychosocial beliefs correlate with their patient communication skills. *J Gen Intern Med* 10:375-379
35. Lydick E, Epstein RS (1993) Interpretation of quality of life changes. *Qual Life Res* 2:221-226
36. McCorkle R, Cooley ME, Shea JA (1998) A user's manual for the Symptom Distress Scale. University of Pennsylvania, Philadelphia
37. Moinpour CM, Lyons B, Grevstad PK et al (2002) Quality of life in advanced non-small lung cancer: results of a Southwest Oncology Group randomized trial. *Qual Life Res* 11:115-126
38. Mor V, Laliberte L, Morris JN et al (1984) The Karnofsky Performance Status scale: an examination of its reliability and validity in a research setting. *Cancer* 53:2002-2007
39. Non-small Cell Lung Cancer Collaborative Group (1995) Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomized clinical trials. *BMJ* 311:899-909
40. Norman GR, Gwadry Sridhar F et al (2001) The relation of distribution- and anchor-based approaches in interpretation of changes in health related quality of life. *Med Care* 39:1039-1047
41. Norman GR, Sloan JA, Wywich KW (2003) Interpretation of changes in health-related quality of life. The remarkable universality of a half a standard deviation. *Med Care* 41:582-592
42. Nunnally CC, Bernstein IH (1994) Psychometric theory, 3rd edn. McGraw-Hill, USA
43. Osoba DA (2002) Taxonomy of the uses of health-related quality-of-life instruments in cancer care and the clinical meaningfulness of the results. *Med Care* 40:III-31-III-38
44. Osoba D, Rodrigues G, Myles J et al (1998) Interpreting the significance of changes in health-related quality of life scores. *J Clin Oncol* 16:139-144
45. Schipper H, Clinch J, McMurray A et al (1984) Measuring the quality of life of cancer patients: the Functional Living Index-Cancer: development and validation. *J Clin Oncol* 2:472-483
46. Sloan JA, Dueck A (2004) Issues for statisticians in conducting analyses and translating results for quality of life end points in clinical trials. *J Biopharm Stat* 14:73-96
47. Sloan JA, Loprinzi CL, Kross SA et al (1998) Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. *J Clin Oncol* 16:3662-3673
48. Sloan JA, Aaronson N, Cappelleri JC (2002) Assessing the clinical significance of single items relative to summated scores. *Mayo Clin Proc* 77:479-487
49. Sloan JA, Cella D, Frost M et al (2002) Assessing clinical significance in measuring oncology patient quality of life: introduction to the symposium, content overview, and definition of terms. *Mayo Clin Proc* 77:367-370
50. Sloan J, Symonds T, Vargas-Chanes D et al (2003) Practical guidelines for assessing the clinical significance of health-related quality of life changes within clinical trials. *Drug Inf J* 37:23-31
51. Sloan JA, Vargas-Chanes D, Kamath CC et al (2003) Detecting worms, ducks and elephants: a simple approach for defining clinically relevant effects in quality-of-life measures. *J Cancer Integ Med* 1:41-47
52. Spitzer WO, Dobson AJ, Hall J et al (1981) Measuring the quality of life of cancer patients: a concise QL-index for use by physicians. *J Chronic Dis* 34:585-597
53. Sprangers MAG, Moinpour CM, Moynihan TJ et al (2002) Assessing meaningful change in quality of life over time: a user's guide for clinicians. *Mayo Clin Proc* 77:561-571
54. Symonds T, Berzon R, Marquis P et al (2002) The clinical significance of quality-of-life results: practical considerations for specific audiences. *Mayo Clin Proc* 77:572-583
55. Tamburini M, Brunelli C, Rosso S et al (1996) Prognostic value of quality of life scores in terminal cancer patients. *J Pain Symptom Manage* 11:32-41
56. Taylor KM, Macdonald KG, Bezjak A, Ng P, DePetrillo AD (1996) Physicians' perspective on quality of life: an exploratory study of oncologists. *Qual Life Res* 5:5-14
57. Unruh ML, Weisbord SD, Kimmel PL (2005) Health-related quality of life in nephrology research and clinical practice. *Semin Dial* 18:82-90
58. Valderas JM, Alonso J, Prieto L, Espallargues M, Castells X (2004) Content-based interpretation aids for health-related quality of life measures in clinical practice. An example for the Visual Function Index (VF-14). *Qual Life Res* 13:35-44
59. Velikova G, Brown JM, Smith AB, Selby PJ (2002) Computer-based quality of life questionnaires may contribute to doctor-patient interactions in oncology. *Br J Cancer* 86:51-59
60. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, Selby PJ (2004) Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol* 22:714-724
61. Ware JE, Kosinski M, Keller SD (1994) SF-36 physical and mental health summary scales: a user's manual. Health Institute, New England Medical Center, Boston, Massachusetts
62. Wright EP, Selby PJ, Crawford M, Gillibrand A, Johnston C, Perren TJ, Rush R, Smith A, Velikova G, Watson K, Gould A, Cull A (2003) Feasibility and compliance of automated measurement of quality of life in oncology practice. *J Clin Oncol* 21:374-382
63. Wywich KW, Nienaber NA, Tierney WM et al (1999) Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 37:469-478