



## Review article

# Assessment and data analysis of health-related quality of life in clinical trials for gastric cancer treatments

SATOSHI MORITA<sup>1</sup>, ADRIAN A. KAPTEIN<sup>2</sup>, AKIRA TSUBURAYA<sup>3</sup>, YASUHIRO KODERA<sup>4</sup>, TAKANORI MATSUI<sup>5</sup>,  
and JUNICHI SAKAMOTO<sup>6</sup>

<sup>1</sup>Department of Epidemiology and Health Care Research, School of Public Health, Kyoto University, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

<sup>2</sup>Psychology Unit, Leiden University Medical Center LUMC, Leiden, The Netherlands

<sup>3</sup>Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan

<sup>4</sup>Department of Surgery, Nagoya University, Nagoya, Japan

<sup>5</sup>Department of Clinical Research, Aichi Cancer Center Aichi Hospital, Aichi, Japan

<sup>6</sup>Medical Administration Course of Master's Degree Program, Nagoya University, Nagoya, Japan

### Abstract

**This review contains a practical guide as well as a number of design and analytical methods for conducting health-related quality of life assessments in clinical trials for gastric cancer treatment. Although we address issues specific to the assessment of gastric cancer, many are applicable to clinical trials for different cancer types.**

**Key words** Health-related quality of life · Gastric cancer · Clinical trial

### Introduction

Unlike evaluation of treatments for other common solid tumors, such as lung, colon, or breast cancers, evaluating the therapeutic efficacy of gastric cancer treatments cannot be achieved by measuring the response rate of measurable disease. More than a third of gastric cancer patients have nonmeasurable disease, such as ascites, lymphangitis carcinomatosa, and miliary liver metastasis. These diseases cannot be evaluated by the ordinary response criteria used for solid tumors. For gastric cancer patients with no measurable disease, evaluation criteria, such as health-related quality of life (HRQOL), could be a critical modality for the assessment of gastric cancer. To this end, the evaluation of HRQOL should be of principal importance when estimating disease progression and evaluating treatment options for gastric cancer.

In the late 1980s and early 1990s, the definition of HRQOL and the methodology for assessing HRQOL

in cancer patients was actively discussed [1–7]. HRQOL is generally recognized as a subjective, multidimensional construct [1,3,6,8–10] that should at minimum include physical, functional, mental (or emotional), and social well-being [11]. These variables have been measured using validated cancer-specific instruments for the assessment of HRQOL, such as the European Organization for the Research and Treatment of Cancer (EORTC) core questionnaire QLQ-C30 [2] and the Functional Assessment of Cancer Therapy Scale General (FACT-G) [12]. Subsequently, HRQOL has been investigated in a longitudinal fashion in clinical trials of cancer therapies using such questionnaires as important supplementary endpoints to traditional endpoints such as survival, tumor response, and toxicity. For patients with advanced-stage disease, such as advanced or recurrent stomach cancer, it is important to evaluate the influence of aggressive chemotherapy on HRQOL because of concerns about patients' tolerance to therapy [13]. It may also be important to assess the HRQOL of patients receiving postoperative adjuvant therapy because such treatment often continues for years. When several treatments produce only minimal differences in disease control and survival, the impact of the toxicity and treatment schedule on the individual patient's perceived well-being is a criterion that should influence the selected treatment for a particular patient. As stated above, HRQOL is an important endpoint in clinical trials of cancer treatments.

An editorial in the *Journal of Clinical Oncology* in 2002 [14] raised the issue that “There are a number of studies that have been published in the *Journal* in recent years that show significant changes in specific symptoms or toxicity but no significant changes in quality-of-life measurements.” In addition, the editorial strongly en-

Offprint requests to: S. Morita

Received: May 30, 2006 / Accepted: September 13, 2006

couraged research of what they term the “next” step: the translation of QOL measurement into clinical practice to improve patient care. What does “next” mean? To understand the trends in HRQOL evaluation, one must know the history of HRQOL research. Thus, the current review discusses the history and current status of HRQOL assessment in clinical trials of cancer treatments by providing readers with several examples, as well as the basic knowledge and methods of HRQOL assessment. This should allow readers to comprehend the problems particular to HRQOL assessment and may lead to a breakthrough in the issues addressed in the editorial.

This article is organized as follows. The next section describes the definitions of HRQOL, followed by a summary of the instruments that measure HRQOL and new concepts in HRQOL assessment in gastric cancer. We then summarize the data analysis of HRQOL and provide an example of HRQOL assessment in a randomized clinical trial in gastric cancer patients. We conclude with a general discussion of the subject.

## Definitions of HRQOL

What is the definition of HRQOL? As already stated, HRQOL can be represented as a multidimensional construct of domains. This is based on the broad definition of health described by the World Health Organization (WHO): “Health is the state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity” [15]. Although the specific terminology and conceptual formulations differ among investigators, the focus is on the patient’s perception. In day-to-day research, the term HRQOL indicates a patient-centered approach but ranges from one extreme of self-administered questionnaires completed by patients to the other extreme in which symptoms such as pain or daily activities are assessed by physicians or nurses. The EORTC QLQ-C30 and the FACT-G are well-known generic questionnaires based on multidimensional constructs. The EORTC QLQ-C30 consists of five functional subscales (physical, role, cognitive, emotional, social), nine symptom subscales (fatigue, pain, nausea and vomiting, constipation, diarrhea, sleep, dyspnea, appetite, financial), and a global QOL scale. The FACT-G is comprised of four subscales: physical, functional, social, and emotional well-being. Some clinical investigators may be uncomfortable dealing with the mental (or emotional) and social domains. One example of the symptom-based assessment performed by physicians is the Clinical Benefit Response (CBR), a value created to evaluate advanced pancreatic cancer [16–18] that is composed of pain assessment and Karnofsky performance status [19]. The intermediate

course of action between the extremes is a disease-specific module focusing on symptoms or the problems and discomfort patients experience (or both). Gastric cancer-specific modules have been developed and are available for assessment.

## Measurement instruments for HRQOL

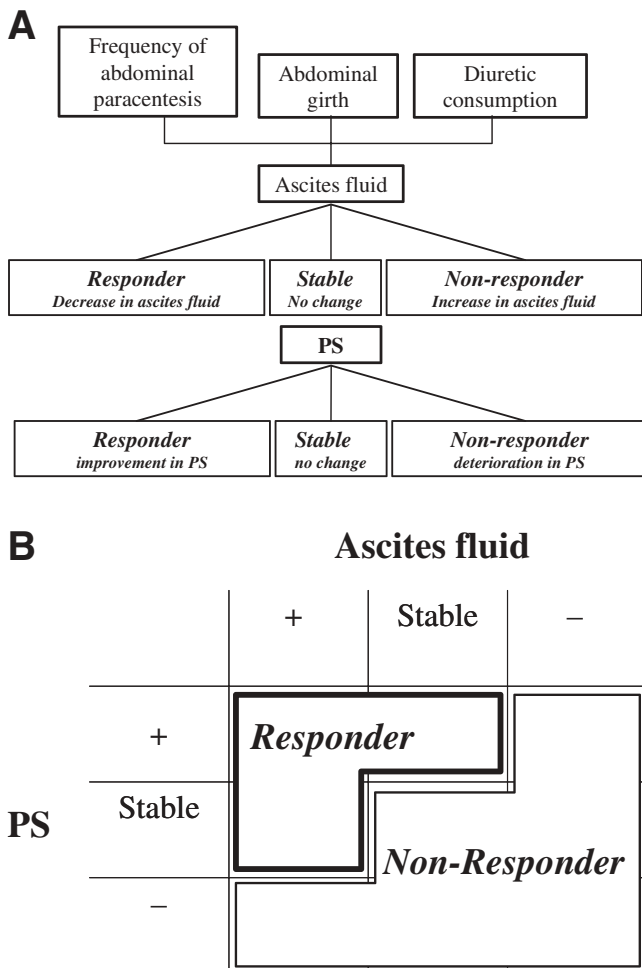
### Questionnaires

Kaptein et al. [20] conducted literature searches of MEDLINE from 1966 to 2004 using the Mesh headings “gastric cancer,” “quality of life,” “gastric cancer,” and “psychology” to find empirical studies and measure(s) used to assess HRQOL and other issues. In the 26 studies returned, a large variety of QOL instruments were reported. Homemade questionnaires and questionnaires that merely assess performance status were used frequently. Questionnaires designed specifically for gastric cancer were seldom applied. An additional literature search of Ichushi-WEB, one of the largest literature search engines in Japan, produced 119 original hits. This search revealed that 8 of the 119 studies assessed the HRQOL of patients using a QOL questionnaire and that almost all of the questionnaires were homemade.

During the 10 years since the development of the generic QOL questionnaires, including the EORTC QLQ-C30 and the FACT-G, gastric cancer-specific modules have been developed, including the EORTC QLQ-STO22 [21–23] and the FACT-Ga [24]. The EORTC QLQ-STO22 module contains 22 items regarding dysphagia, eating restrictions, pain, reflux, anxiety, dry mouth, body image, and hair loss. The FACT-Ga consists of 16 items including weight loss, appetite loss, fatigue, dysphagia, pain, vomiting, eating restrictions, and body image. We expected that many clinical trials of gastric cancer treatments that started after the development of the modules would employ one of the two instruments to assess HRQOL specific to gastric cancer patients. The EORTC and the FACT questionnaires are available at [www.eortc.be/home/qol/](http://www.eortc.be/home/qol/) and [www.facit.org/](http://www.facit.org/), respectively. Those questionnaires have been translated into many languages; for example, the Japanese versions of the EORTC QLQ-C30 and the FACT-G were generated using a standardized, forward-backward translation protocol [25,26]. Although translation of the EORTC QLQ-STO22 and the FACT-Ga into Japanese has been completed [27], their final validation studies are currently ongoing.

### Composite endpoint

The measurement of clinical benefits was introduced to examine treatments for pancreatic cancer [16–18]. The



**Fig. 1. A** Assessment of “Clinical Benefit Response — Gastric Cancer.” Ascites fluid is assessed by the frequency of abdominal paracentesis, abdominal girth, and diuretic consumption. Functional impairment is assessed by the ECOG performance status (PS). **B** Composite endpoint comprised of ascites fluid and functional impairment. Patients are considered responders if they are rated positive for at least one of the two measures and not rated negative for the other

clinical benefit is a composite assessment of the typical debilitating symptoms of pancreatic cancer. The parameters included were pain intensity, Karnofsky performance status, and loss in body weight. In reference to this, Sakamoto et al. [28] proposed a novel design that focused on the typical features seen in patients with ascites-forming advanced gastric cancer by using a composite endpoint termed Clinical Benefit Response — Gastric Cancer (Fig. 1). Ascites fluid (assessed by the frequency of abdominal paracentesis, abdominal girth, and diuretic consumption) and functional impairment (assessed by ECOG performance status, or PS) comprise the measures of this endpoint. Each patient is classified as positive, stable, or negative for each of the two measures. With regard to ascites fluid, “positive”

indicates a decrease in the frequency of abdominal paracentesis, abdominal girth, or diuretic consumption. If an increase in either of the parameters is observed, the ascites fluid measure is considered negative. The functional impairment measure is determined on the basis of three categories — improvement, no change, and deterioration — describing the ECOG PS. To achieve a response in terms of the overall clinical benefit rating, patients must be rated “positive” for at least one of the two measures (ascites fluid or functional impairment) without being rated “negative” for the other. Otherwise, patients are classified as “nonresponsive” for the overall rating.

#### Assessment schedule

Whether using a self-administered questionnaire or a composite endpoint, HRQOL should be investigated in a longitudinal fashion in a clinical trial. This raises the question of when measurements are undertaken. Before launching a trial, the timing and frequency of HRQOL administrations must be determined. These assessment schedules are not unique to various types of cancer or anticancer treatments, whereas baseline administration, before randomization or the initiation of treatment, is mandatory in all trials. For example, in a clinical trial for chemotherapy in advanced-stage cancer, measurements are needed at least before, during, and after treatment. Acute therapy toxicity that occurs within 1–2 weeks following drug administration significantly deteriorates a patient’s HRQOL [29]. Thus, measurements are usually taken just before the next course of treatment to avoid the temporal influence of acute toxicity on a patient’s HRQOL and to evaluate the entire effect of a treatment and its effect on disease progression [30–32]. Taking into account the substantial burden on patients that filling in questionnaires entails, it may be sufficient to administer HRQOL assessments every two treatment courses.

The next question is the duration that measurements are taken. Although it is common to conduct HRQOL assessments until 1 month after the completion of chemotherapy, one must account for any possible influence of second-line therapies, such as chemotherapy or radiotherapy, when analyzing the data. If the effects of cumulative neurotoxicity on a patient’s HRQOL are of interest (e.g., that of taxanes), HRQOL assessment may require continuation over at least four courses of treatment. In contrast, HRQOL assessment of patients with postoperative breast cancer who have relatively good prognoses should continue much longer. In a clinical trial of postoperative adjuvant therapy, the follow-up duration is usually at least a year, so it is practical to administer HRQOL assessments every 3–6 months until the completion of treatment.

**Table 1.** Methods of analyses used for longitudinal QOL data

Method	Advantages	Disadvantages
Multiple univariate analysis <i>t</i> -test; Wilcoxon rank sum test, etc. at each time point	Simple implementation Useful for exploratory analyses	Inflation of alpha error due to multiple testing Difficulty in interpretation of results
Longitudinal data analysis General linear model Generalized estimation equation	Greater statistical power compared to multiple univariate analysis Secular trends can be examined	More advanced statistical skills required
Summary measures AUC, slope, etc.	Simple implementation Easy to interpret results	Arbitrary selection of measures

QOL, quality of life; AUC, area under the curve

## Data analysis

As mentioned above, HRQOL assessments are usually conducted at multiple time points before, during, and after treatment with the goal of characterizing a patient's HRQOL in a longitudinal fashion. Thus, statistical techniques designed for longitudinal data should be used to analyze HRQOL data. In the following sections, we summarize examples of such methods (Table 1) and discuss the issue of missing data, which is a fundamental difficulty in HRQOL methodology [33–36].

### Descriptive statistics

As a first step, it is important to summarize fluctuations in observed HRQOL data graphically. The appropriate use of graphics can quickly and clearly summarize complex data and simultaneously emphasize the high degree of variability in HRQOL data. Graphical techniques can highlight changes in HRQOL that would be considered clinically significant and make it apparent to readers which changes are unimportant even if statistically significant. For multidomain HRQOL data, it is useful to show mean HRQOL scores with a 95% confidence interval (approximately two times the standard error) at each time point according to treatment group. It may be suitable to compute changes in the HRQOL score from the baseline and express the mean scores in the same way. Binary data, such as the clinical benefit response (responder/nonresponder), require tabulation. In addition to summarizing the proportion of responders, presenting the data of each component — such as changes in PS and the consumption of analgetics or diuretics — may provide a deeper interpretation of treatment effects. In any case, compliance with HRQOL assessment at each time point should include the number of patients who completed the questionnaires or the number of patients from whom the required data were observed.

### Multiple univariate analyses

The most commonly used statistical technique is probably univariate analysis at each time point using statistical tests, including Student's *t*-test and Wilcoxon rank sum tests. This approach is easy to implement and a useful way to explore data, displaying *P* values for descriptive purposes. However, if a large number of comparisons are required, this approach may lead to difficulty in the interpretation of results and result in confusion. From a statistical perspective, such multiple comparisons can also inflate the type I error. Although one strategy for dealing with multiple testing is to use a procedure of multiple comparisons (e.g., the Bonferroni correction), the statistical power to detect differences between the effects of treatments on HRQOL unavoidably decreases. A more practical strategy may be to specify a limited number of time points for comparison during the design stage of the study. This requires a deeper discussion about which points are clinically appropriate for comparison.

### Analysis of longitudinal data

An analytical approach to analyze repeat measures all at one time and not analyze data at each time point, is available for HRQOL data. These procedures include general linear modeling [37–39] and the generalized estimation equation (GEE) [39,40]. Although these methods require greater knowledge of and experience with statistics, they have several potential advantages, including (1) larger statistical power and (2) the ability to examine the differences between long-term trends in multiple treatment groups. Readers should refer to textbooks where specific statistical methods are explained using real examples [38,39]. For continuous variables, such as the domain scores of a QOL questionnaire, a linear general model has recently gained popularity owing to developments in statistical methods and software. Analysts must choose what type of correlation matrix can best handle serial correlations between data

measured repeatedly over multiple time points. Next, a linear model that includes explanatory variables such as treatment groups, time points, and their interaction effects is fitted to the data. If a response variable is longitudinal data but not continuous — for example, clinical benefit response data with a binary response of 0 or 1 — the GEE method is commonly used. A serial correlation matrix should also be chosen when using GEE. When using either method, missing data is a critical issue, and both methods require assumptions about missing data to guarantee unbiased comparisons. The issue of missing data is discussed below.

### *Summary measures*

An alternative approach to multiple univariate analyses and longitudinal data analysis is the use of summary measures. Examples include slope over time, the change from baseline to the last measurement, the maximum or minimum value, the area under the curve (AUC), and the time to a peak or to a prespecified value. The choice of measure depends on the disease and available treatments. Nevertheless, one may be able to select a summary measure by considering several possible patterns of HRQOL changes over time. For example, if HRQOL scores change steadily over time, showing constant improvement or deterioration, a slope may be the best choice. If the HRQOL of a patient temporarily improves or deteriorates but then returns to its original status, AUC or maximum or minimum values may be suitable to examine the effects of treatments. A third pattern may be that all patients' HRQOL values improve to complete health, which depends on curing the disease in each patient. In such a case, the time to a peak or a prespecified value may be an appropriate endpoint for the clinical trial. Clinicians may be familiar with all of the above summary measures, which are easy to interpret. Moreover, summary measures do not increase the multiplicity problem due to multiple statistical testing. Even when using these summary measures, however, missing data can lead to biased estimates of comparison. For example, when using an AUC measure, it is obvious that comparisons will be invalid if the time periods between HRQOL assessments differ among patients owing to missing data or dropouts.

### *Missing data*

In clinical trials, the data collection schedule is fully described in detail in the study protocol. Perfect data collection is not achievable in real trials, however, as some data are always missing. In the case of CBR assessment, it is necessary to observe data for all components of CBR in order to compute the CBR of a patient.

The questionnaire-based HRQOL assessment can be faced with missing data for many reasons, such as lost questionnaires or the deteriorating health of patients [33–36]. In any case, missing data, at minimum, leads to the following two problems [34]. The first is the loss of statistical power due to the reduced number of data points. This problem can be addressed by increasing the planned sample size for the trial. The second is a nonresponse bias that potentially occurs due to systematic differences between the observed and unobserved data. Particularly in studies of patients with advanced-stage disease, the amount of missing HRQOL data increases over the course of treatment owing to deterioration of patients' health and sometimes patient death [41–47].

To understand missing data issues, it is helpful to make important distinctions between missing data processes [35,45,48]. A missing data process is said to be missing completely at random (MCAR) if data are missing for reasons unrelated to the study treatments or patients' illnesses (e.g., avoidable causes such as administrative or logistical issues). In such a case, all the statistical methods described above provide unbiased comparisons. The “missingness” is considered to be missing at random (MAR) if, conditional on the observed data, the missingness is independent of the unobserved measurements. In this case, a general linear model can be used for data analysis. If data are missing for a reason that would affect the unobserved (and potentially observable) data values, such as deteriorating health, they are missing not at random (MNAR). If the missing data are MNAR, analyzing only the observed data would generate serious bias. Thus, specific statistical approaches are required to overcome the problems of MNAR, and numerous statistical approaches have been proposed [42,44,45,49,50].

Data imputation is an attractive approach for the analysis of incomplete data due to its flexibility in handling the incomplete-data problem [30,34,47,49]. Although simple imputation methods that substitute a single reasonable value, such as the mean score of the observed data, for a missing measurement have been traditionally used, they have a major weakness: underestimation of the variance. Rubin's multiple imputation [31,34] solves this problem while retaining many of the advantages of single imputation [32,34,51–53]. This approach combines  $m$  ( $>1$ ) completed data sets that are generated by imputing unobserved data  $m$  times through an imputation model to “properly” take into account the uncertainty inherent to the imputed values [31,51]. The imputation model, which incorporates relevant variables that not only predict dropout but also correlate with outcomes (i.e., HRQOL), is an effective way to achieve approximate comparability between respondents and nonrespondents for unobserved missing

HRQOL data that are MNAR [54]. Simes et al. [47] and Morita et al. [30] proposed incorporating auxiliary HRQOL information — such as physician-assessed health status — for exactly that purpose.

As an alternative, pattern mixture models have been previously employed to address the problem of MNAR by partitioning patients into several strata (patterns) based on measures of deteriorating health, such as disease progression and duration of survival [49,54]. Estimates to compare treatment groups are obtained by calculating weighted averages of the estimates obtained within each of the subgroups. Weighting is assigned according to the proportion of patients divided into each subgroup.

As mentioned above, missing data is a common problem in QOL studies that leads to analytical problems. The best strategy for minimizing the problem of missing data is to prevent it [43,45,47]. The best efforts (better staff training and more support of clinical research coordinators) should be made to improve data collection by reducing uncollected questionnaires due to administrative errors. In addition, data collection can be improved by organizing a QOL data center and taking measures to motivate patients to fill in questionnaires. However, missing data, especially data loss associated with deteriorating health of patients, is unavoidable in the context of advanced stage disease.

Given the numerous potential methods of analysis, how do we choose between strategies? No gold standard has been established, and it is difficult in practice to choose the “best” approach. We recommend carrying out a sensitivity-type analysis to examine the effects of each analytical method; in other words, investigate the robustness of the conclusions obtained from each alternative analysis [44].

### Example of HRQOL assessment in a clinical trial

A reference example for this section is that of Bouche et al. [55] published in 2004. They assessed HRQOL in patients with previously untreated metastatic gastric cancer in a randomized phase II clinical trial. Three chemotherapeutic treatments were compared: leucovorin (LV) 200mg/m<sup>2</sup> (2-h infusion) followed by fluorouracil (FU) 400mg/m<sup>2</sup> (bolus) and FU 600mg/m<sup>2</sup> (22-hour continuous infusion) on days 1 and 2 every 14 days (LV5FU2; arm A); LV5FU2 plus cisplatin 50mg/m<sup>2</sup> (1-h infusion) on day 1 or 2 (arm B); and LV5FU2 plus irinotecan 180mg/m<sup>2</sup> (2-h infusion) on day 1 (arm C). Treatment was continued for at least four courses or until disease progression, unacceptable toxicity, or patient refusal. HRQOL was assessed as a secondary endpoint. The primary endpoint was the overall response rate, and the other secondary endpoints were

progression-free survival, overall survival, safety, and the duration of hospitalization.

HRQOL assessment was carried out using the EORTC QLQ-C30 before randomization and every 2 months thereafter. The observed data were analyzed in the following manner. Summary statistics, including the mean, standard deviation, median, and range, were computed for each subscale of the QLQ-C30 at baseline and at each 2-month time point. The longitudinal change in scores was graphically presented along with the missing data information. In addition, the longitudinal changes during the first three follow-up assessments were compared among the treatment groups using a general linear model analysis.

A total of 134 patients were enrolled in the study between January 1999 and October 2001. Arms A, B, and C included 45, 44, and 45 patients, respectively, of whom 37, 33, and 38 patients completed the questionnaires at baseline. HRQOL data at the third evaluation were observed in 22 (41%), 21 (38%), and 29 (48%) patients in arms A, B, and C, respectively. Thereafter, the number of patients who completed the questionnaire at follow-up was fewer than 10 in each arm of the study.

Longitudinal analysis revealed higher HRQOL scores in arm C than in arms A and B for all functional subscales and nearly all symptom scales, but the differences were not statistically significant. The secular trend in global QOL was roughly equivalent among arms and increased from the baseline over time.

Although the participation of patients in HRQOL assessments was low in all arms, no specific statistical method was employed to deal with the missing data issues. The authors did discuss the missing data issue and noted the following related limitations of the study. The reduction in the availability of HRQOL data, especially after the third follow-up, could have biased the analyses of the longitudinal HRQOL. Because patients with shorter survival times or progression (or both) had poor compliance with completing the HRQOL assessments and probably poor HRQOL, the mean scores may be overestimated, especially at the later follow-ups.

### Conclusion

This article has discussed the design and analytical methods for HRQOL assessment during cancer clinical trials. Although several disease-specific modules require development, the instruments for generic HRQOL assessments are roughly satisfactory. There are at least two issues, however, that have not yet been fully resolved. The first is the statistical issue of how to handle missing HRQOL, a problem that occurs because of, for example, deterioration of patient health or pro-

gression of the disease. The second issue is the clinical issue regarding how to interpret HRQOL data generated during clinical trials and how to utilize HRQOL evaluations when making decisions about individual patient care [56]. When these issues are solved, HRQOL will be widely accepted as a truly useful endpoint for clinical trials. The future challenge is to incorporate QOL measures into decisions regarding patient care in the clinic.

## References

- Aaronson NK, Bullinger M, Ahmedzai S. A modular approach to quality-of-life assessment in cancer clinical trials. *Recent Results Cancer Res* 1988;111:231–49.
- Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, 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. *J Natl Cancer Inst* 1993;85:365–76.
- Osoba D. Lessons learned from measuring health-related quality of life in oncology. *J Clin Oncol* 1994;12:608–16.
- Moinpour CM, Feigl P, Metch B, Hayden KA, Meyskens FL, Crowley J. Quality of life end points in cancer clinical trials: reviews and recommendations. *J Natl Cancer Inst* 1989;81:485–95.
- Gotay CC, Korn EL, McCabe MS, Moore TD, Cheson BD. Quality-of-life assessment in cancer treatment protocols: research issues in protocol development. *J Natl Cancer Inst* 1992;84:575–9.
- Nayfield SG, Ganz PA, Moinpour CM, Cella DF, Hailey BJ. Report from a National Cancer Institute (USA) workshop on quality of life assessment in cancer clinical trials. *Qual Life Res* 1992;1:203–10.
- Machin D. Assessment of quality of life in clinical trials of the British Medical Research Council. *J Natl Cancer Inst Monogr* 1996;20:97–102.
- Ganz PA. Quality of life and the patient with cancer: individual and policy implications. *Cancer* 1994;74(suppl 4):1445–52.
- Lipscomb J, Donaldson MS, Hiatt RA. Cancer outcomes research and the arenas of application. *J Natl Cancer Inst Monogr* 2004; 33:1–7.
- Lipscomb J, Donaldson MS, Arora NK, Brown ML, Clauser SB, Potosky AL, et al. Cancer outcomes research. *J Natl Cancer Inst Monogr* 2004;33:178–97.
- Rowland JH. Outcome assessment: cancer-specific quality-of-life measures: beyond the research setting. Educational Book. ASCO 33rd Annual Meeting, Denver, CO, May 17–20, 1997. p. 342–9.
- Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;11: 570–9.
- Moinpour CM. Measuring quality of life: an emerging science. *Semin Oncol* 1994;21(suppl 10):48–63.
- Levine MN, Ganz PA. Beyond the development of quality-of-life instruments: where do we go from here?. *J Clin Oncol* 2002;20: 2215–6.
- World Health Organization. Constitution in basic documents. Geneva; World Health Organization: 1948.
- Burriss H, Storniolo AM. Assessing clinical benefit in the treatment of pancreas cancer: gemcitabine compared to 5-fluorouracil. *Eur J Cancer* 1997;33(suppl 1):18–22.
- Carmichael J, Fink U, Russell RCG, Spittle MF, Harris AL, Blatter J. Phase II study of gemcitabine in patients with advanced pancreatic cancer. *Br J Cancer* 1996;73:101–5.
- Burriss HA 3rd, Moore MJ, Andersen J, Green MR, Rothenberg ML, Modiano MR, et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. *J Clin Oncol* 1997;15:2403–13.
- Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod CM, editor. Evaluation of chemotherapeutic agents. New York: Columbia University Press; 1949. p. 199–205.
- Kaptein AA, Morita S, Sakamoto J. Quality of life in gastric cancer. *World J Gastroenterol* 2005;11:3189–96.
- Vickery CW, Blazeby JM, Conroy T, Johnson CD, Alderson D. Development of an EORTC module to improve quality of life assessment in patients with gastric cancer. *Br J Surg* 2000;87:362.
- Vickery CW, Blazeby JM, Conroy T, Arraras J, Sezer O, Koller M, et al. Development of an EORTC disease-specific quality of life module for use in patients with gastric cancer. *Eur J Cancer* 2001;37:966–71.
- Blazeby JM, Conroy T, Bottomley A, Vickery C, Arraras J, Sezer O, et al. Clinical and psychometric validation of a questionnaire module, the EORTC QLQ-STO 22, to assess quality of life in patients with gastric cancer. *Eur J Cancer* 2004;40:2260–8.
- Eremenco SL, Cashy J, Webster K, Ohashi Y, Locker GY, Pelletier G, et al. FACT-Gastric: a new international measure of QOL in gastric cancer. *ASCO Proc* 2004;23:755.
- Kobayashi K, Takeda F, Teramukai S, Gotoh I, Sakai H, Yoneda S, et al. A cross-validation of the European Organization for Research and Treatment of Cancer QLQ-C30 (EORTC QLQ-C30) for Japanese with lung cancer. *Eur J Cancer* 1998;34:810–5.
- Fumimoto H, Kobayashi K, Chang CH, Eremenco S, Fujiki Y, Uemura S, et al. Cross-cultural validation of an international questionnaire, the General Measure of the Functional Assessment of Cancer Therapy scale (FACT-G), for Japanese. *Qual Life Res* 2001;10:701–9.
- Morita S, Shimozuma K, Sato A, Nakamachi M, Blazeby J, West K, et al. Pilot-testing of the Japanese version of the EORTC QLQ-STO22 gastric cancer module. *Gan To Kagaku Ryoho* 2004;31:1195–9 (in Japanese).
- Sakamoto J, Morita S, Yumiba T, Narahara H, Kinoshita K, Nakane Y, et al. A phase II clinical trial to evaluate the effect of paclitaxel in patients with ascites caused by advanced or recurrent gastric carcinoma: a new concept of clinical benefit response for non-measurable type of gastric cancer. *Jpn J Clin Oncol* 2003;33: 238–40.
- Morita S, Kobayashi K, Eguchi K, Matsumoto T, Shibuya M, Yamaji Y, et al. Influence of clinical parameters on quality of life during chemotherapy in patients with advanced non-small-cell lung cancer: application of a general linear model. *Jpn J Clin Oncol* 2003;33:470–6.
- Morita S, Kobayashi K, Eguchi K, Matsumoto T, Shibuya M, Yamaji Y, et al. Analysis of incomplete quality of life data in advanced stage cancer: a practical application of multiple imputation. *Qual Life Res* 2005;14:1533–44.
- Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
- Lavori PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. *Stat Med* 1995;14: 1913–25.
- Staquet MJ, Hays RD, Fayers PM. Quality of life assessment in clinical trials. Oxford: Oxford University Press; 1998.
- Ohashi Y, Morita S. Statistical analysis of quality of life data. In: Ikegami N, et al., editors, Handbook of quality of life assessment for clinicians. Tokyo: Igaku-Shoin; 2001. p. 19–29.
- Fairclough DL. Design and analysis of quality of life studies in clinical trials. New York: Chapman & Hall/CRC; 2002.
- Fairclough DL. Patient reported outcomes as endpoints in medical research. *Stat Methods Med Res* 2004;13:115–38.
- Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–74.

38. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer; 2000.
39. Diggle PJ, Heagerty P, Liang KY, Zeger SL. Analysis of longitudinal data, 2nd edition. Oxford: Oxford University Press; 2002.
40. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988;44:1049–60.
41. Kopp I, Lorenz W, Rothmund M, Koller M. Relation between severe illness and non-completion of quality-of-life questionnaires by patients with rectal cancer. *J R Soc Med* 2003;96:442–8.
42. Moinpour CM, Sawyers Triplett J, McKnight B, Lovato LC, Upchurch C, et al. Challenges posed by non-random missing quality of life data in an advanced-stage colorectal cancer clinical trial. *Psychooncology* 2000;9:340–54.
43. Conroy T, Bleiberg H, Glimelius B. Quality of life in patients with advanced colorectal cancer: what has been learnt? *Eur J Cancer* 2003;39:287–94.
44. Fairclough DL, Gagnon DD, Zagari MJ, Marschner N, Dicato M, Epoetin Alfa Study Group. Evaluation of quality of life in a clinical trial with nonrandom dropout: the effect of epoetin alfa in anemic cancer patients. *Qual Life Res* 2003;12:1013–27.
45. Moinpour CM, Lyons B, Grevstad PK, Lovato LC, Crowley J, Czaplicki K, et al. Quality of life in advanced non-small-cell lung cancer: results of a Southwest Oncology Group randomized trial. *Qual Life Res* 2002;11:115–26.
46. Bernhard J, Cella DF, Coates AS, Fallowfield L, Ganz PA, Moinpour CM, et al. Missing quality of life data in cancer clinical trials: serious problems and challenges. *Stat Med* 1998;17:517–32.
47. Simes RJ, Grooten V, GebSKI VJ. Practical approaches to minimize problems with missing quality of life data. *Stat Med* 1998;17:725–37.
48. Little RA, Rubin DB. Statistical analysis with missing data, 2nd edition. New York: Wiley; 2002. p. 3–23.
49. Fairclough DL, Peterson HF, Cella D, Bonomi P. Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Stat Med* 1998;17:781–96.
50. Matsuyama Y. Sensitivity analysis for the estimation of rates of change with non-ignorable drop-out: an application to a randomized clinical trial of the vitamin D<sub>3</sub>. *Stat Med* 2003;22:811–27.
51. Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res* 1999;8:17–36.
52. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999;8:3–15.
53. Fairclough DL. Multiple imputation for non-random missing data in longitudinal studies of health-related quality of life. In: Mesbah M, Cole BF, Lee MLT, editors. *Statistical methods for quality of life studies*. Boston: Kluwer Academic; 2002. p. 323–37.
54. Pauler DK, McCoy S, Moinpour C. Pattern mixture models for longitudinal quality of life studies in advanced stage disease. *Stat Med* 2003;22:795–809.
55. Bouche O, Raoul JL, Bonnetain F, Giovannini M, Etienne PL, Lledo G, et al. Randomized multicenter phase II trial of a biweekly regimen of fluorouracil and leucovorin (LV5FU2), LV5FU2 plus cisplatin, or LV5FU2 plus irinotecan in patients with previously untreated metastatic gastric cancer: a Federation Francophone de Cancerologie Digestive Group Study — FFCD 9803. *J Clin Oncol* 2004;22:4319–28.
56. Morita S, Ohashi Y, Kobayashi K, Matsumoto T, Eguchi K, Shibuya M, et al. Individually different “weights” of quality of life assessment in patients with advanced non-small-cell lung cancer. *J Clin Epidemiol* 2003;56:744–51.