Assessing health-related quality of life in cancer trials


In cancer clinical trials, the evaluation of treatment options is often dominated by patient survival. However, treatment improvements have permitted survival to be measured in years rather than months for many cancer patients. This trend is likely to persist and the quality of survival as reported by patients will become an increasingly important end point. Additionally, when new cancer therapies show small survival gains, the quality of this survival period from the patient perspective should be assessed in order to adequately evaluate the new agent and to inform clinicians and patients about tradeoffs in the form of treatment-related side effects. This article describes the rationale for measuring health-related quality of life outcomes in cancer trials, and key design and methodological considerations in clinical trial settings.

Keywords: cancer • clinical trials • health-related quality of life • patient-reported outcomes

Health-related quality of life lineage

The purpose of this article is to summarize the main issues associated with health-related quality of life (HRQOL) in cancer clinical trials. The following HRQOL research issues have been examined and discussed over the last 25 years: definition; rationale for inclusion of HRQOL outcomes in a clinical trial; clinical trial design issues; assessment methods, including selection of HRQOL measures, timing of the assessments, formats for administration and the importance of quality control procedures; data analysis (including how to address missing data problems); and interpretation of HRQOL results [1–27]. We hope that this brief summary of this information will help clinicians and their research collaborators implement the inclusion of HRQOL outcomes in cancer clinical trials.

HRQOL

The HRQOL concept encompasses both specific symptoms and generic aspects of day-to-day functioning (e.g., emotional, physical, social and role function) in the specific context of having a disease or being treated for a medical condition [12]. Using HRQOL measures to evaluate how a disease and its treatment affect a patient’s life allows for the inclusion of a systematic set of treatment-specific outcomes from the patient perspective. While clinicians often ask patients how they are doing in clinical practice settings, a clinical trial requires information that is standardized and easily interpretable. Standardization is accomplished by administration of reliable and validated HRQOL questionnaires to patients at regularly scheduled times during a trial [9,26–28].

In cancer populations, HRQOL assessments add to the broad array of clinical outcomes traditionally collected in cancer clinical trials, such as survival, disease-free survival, tumor response and physician-reported treatment-related side effects [12]. A strong consensus has developed among cancer researchers that HRQOL is a reasonable trial end point [29] and its use has increased in Phase III...
cancer clinical trials [7,12]. HRQOL instruments are also commonly evaluated and used to support drug labeling claims [20]). Potential benefits of HRQOL outcomes have been recognized for comparative effectiveness research, as noted recently by Ahmed et al. [30]. Comparative effectiveness research involves a comprehensive examination of benefits and harms associated with medical care; HRQOL measures provide a systematic assessment of the patient’s perspective [30]. The widespread use of HRQOL assessments in these broad settings highlights the role of patient-reported outcomes (PROs) in adding useful, non-duplicative and treatment-specific data that can even extend into clinical practice. However, some caution about overuse without clear justification is necessary (see the ‘Value of HRQOL outcomes in determining clinical care’ section).

- Benefits of assessing HRQOL

Gotay et al. discussed situations when HRQOL adds value to trial results [31]. If survival is not the primary objective of the study, or if survival outcomes are similar, then HRQOL findings are of more interest to clinicians. HRQOL outcomes are given more attention when they are consistent with expectations or contradict expectations if a biological rationale or mechanism can be posited. The authors also noted that HRQOL outcomes are trusted more when they contrast patients receiving different treatment approaches or when a placebo control is involved. Finally, they mention the issue of outcomes that require a patient report. Depression outcomes reflect such a domain. Passik et al. found good patient/physician agreement for depressive symptoms, only when contrasting no depressive symptoms versus any depressive symptoms [32]. Since survival (or some version of the survival outcome) is usually of primary interest to the clinician, these possibilities need to be considered during the design phase of the trial when HRQOL hypotheses can be framed as either primary or secondary end points.

Systematic assessment of patient function can identify treatment-specific issues that would remain undetected in comparisons of survival or even treatment-specific symptoms, and can even challenge expectations presented by biological end points alone. A seminal study by Sugarbaker et al. examined HRQOL in soft-tissue sarcoma patients randomized to two conditions: limb preservation and radiation versus amputation [33]. Contrary to expectations, patients who received limb preservation reported poorer sexual and physical function. As a result, the clinicians revised the radiation regimen to minimize these side effects [34]. This study shows the utility of HRQOL outcomes for identifying treatment issues that can have a meaningful impact on patients, and highlights the ability of these measures to identify previously unidentified and unanticipated treatment side effects.

Evaluating the connection between symptoms and HRQOL can lead to the identification of unexpected issues and concerns. For example, HRQOL measurement played an important role in a Phase III trial of advanced prostate cancer patients randomized to two conditions: orchiectomy plus placebo, or orchiectomy plus the antiandrogen flutamide [35]. A side-effect of flutamide is diarrhea, which was a primary symptom outcome examined in this study. The study also included a measurement of emotional well-being, with the hypothesis that the severity of diarrhea would be associated with poorer emotional well-being. However, the placebo-arm patients had significantly better emotional well-being than the patients receiving flutamide, regardless of the presence and severity of diarrhea. The data provided a consistent picture that flutamide did not have a net palliative effect; instead, the drug compromised emotional well-being, independent of treatment-specific side effects. Had the investigators only measured treatment-specific symptoms, they would have missed the important impact of this antiandrogen agent on emotional well-being.

There are additional examples of the usefulness of HRQOL outcomes in a clinical trial. Gotay and others have shown that baseline HRQOL levels have prognostic value for survival [36,37]. Both baseline HRQOL data and those obtained during follow-up can enhance patient/physician communication [38–41]. Detmar et al., in a study of 240 patients and ten medical oncologists treating them, found that although patients wanted to discuss emotional and social issues with their physicians while receiving cancer care, they would not discuss them unless the physicians initiated discussion of these topics [38]. Importantly, although six out of ten physicians believed that such issues were part of their responsibilities, none of the ten physicians said they initiated discussions about social and emotional issues. Studies by Detmar et al. [38] and Passik et al. [32], as well as a report by the National Cancer Consensus Conference on Pain, Depression and Fatigue [29] suggest that these measurements be more routine in cancer clinical trials.

An additional benefit of HRQOL assessment is that it can provide information extending beyond the patient perspective, allowing the exploration of the biological and genetic underpinnings of patient-reported symptoms and side effects. For instance, research into the relationship between biomarkers...
and fatigue has considered the relationship between HRQOL and biological markers, such as cortisol, cytokines (IL-1, IL-6 and TNF-α) and serotonin, which could potentially regulate fatigue [42]. Studies have also evaluated biological markers with respect to other HRQOL domains in cancer populations, including depression and sleep disturbance [43], anxiety [44], stress [45] and sexual function [46]. Collection of both HRQOL and biological data allows the clinical researcher to posit more comprehensive models of patient status. For instance, Wilson and Cleary’s model associates biological/physiological factors with symptoms, functional status, general health perception and, ultimately, overall quality of life [47]. To encourage the pairing of biomarkers and HRQOL outcomes in cancer clinical trials, the NCI has established the Biomarker, Imaging and Quality of Life Studies Funding Program. A newer initiative involves the examination of genetic effects on HRQOL [48].

Clinical trial design issues for HRQOL outcomes

Design & reporting issues
In order to successfully incorporate HRQOL measurement and end points into cancer clinical trials, it is important to address their inclusion as part of the initial conceptual design. This will ensure that the HRQOL data are relevant to the research topic being addressed and are being collected for scientifically valid reasons; that appropriate measures are selected; that the HRQOL measures are administered properly and with the necessary follow-up periods; and that suitable quality-control procedures are established. This includes the use of test in all relevant sections of the protocol providing the rationale for the HRQOL outcome (e.g., literature, background and end points sections) in order to communicate the importance of the HRQOL outcomes [49–51]. It is particularly important to have specific HRQOL objectives for the trial; to include baseline HRQOL measures as eligibility criteria for the study; and to specify how the HRQOL outcomes are configured as end points, what the clinically important differences of interest are in the trial, and how these differences will be analyzed. A consistent focus on the use of HRQOL domains throughout the design and implementation phases of a study can increase the likelihood that findings will be useful, amenable to meaningful interpretation, and targeted to the outcomes being assessed by the study, while minimizing the burden on patients.

Previous authors have developed recommendations for incorporating HRQOL outcomes within the research design, implementation and analysis [25,52]. Specific a priori hypotheses examining issues such as differences by patient subgroup, domains affected and the timing of changes in HRQOL scores should be considered at the start of a trial, alongside other commonly collected patient outcomes, such as survival [7,51,53]. Including HRQOL as a primary or secondary end point from the outset will ensure that appropriate questionnaires are selected, comprehensive data collection procedures are developed and disseminated to study staff, and the study is sufficiently powered to identify true differences in the HRQOL domains studied.

The first step in informing the clinical research community about HRQOL results is to report the results adequately. Staquet et al. provided general guidelines for doing so and a checklist for authors to use in preparing an article for publication [54]. Since this time, other guidelines have been proposed. In 1997, Guyatt et al. [55] published a paper in the Journal of the American Medical Association providing guidelines for using HRQOL results from the medical literature. A full review of various guideline efforts for reporting HRQOL results is beyond the scope of this paper; readers are directed to the provided citations.

Despite the existence of guidelines for trials with an HRQOL outcome, however, reviews of randomized controlled trials continue to show that while HRQOL data are often reported, a priori hypotheses are used in as few as 15% of studies [56,57]. This highlights the growing use of HRQOL in clinical trials, and the necessity of incorporating HRQOL early in the study design process. Cocks et al. reviewed the degree to which EORTC QLQ-C30 results are reported in randomized trials of cancer treatment according to a checklist of quality criteria [58]. The authors noted that generally these studies met reporting guidelines but that statistical significance was more commonly used versus some method for addressing clinical or meaningful differences. Other similar studies have been conducted to evaluate the extent to which studies with HRQOL outcomes use rigorous design and reporting methods [51,52,59,60]. In 1981, Najman and Levine reported that few studies of medical interventions reviewed in the paper were conducted such that HRQOL findings could be defended [61]. The authors noted that the bulk of the findings supported the positive impact of the intervention on HRQOL. However, most could be misinterpreted due to poor study designs (particularly sampling plans) or inappropriate indicators (e.g., studies commonly used ‘objective’ vs quality of life indicators, such as social and family networks). They also proposed that the assessment of the gap between HRQOL expectations and their achievement needs to be incorporated into study designs. More recently, Brundage and colleagues [62,63] have found great variability in how thoroughly HRQOL results are reported...
in clinical trial manuscripts. They note the challenges such variability presents to clinicians who are trying to interpret the results and use the information in their clinical practice. Variables of interest in the Brundage evaluations included: specification of hypotheses, rationale for instrument selection, description of psychometric properties of the HRQOL measures, report of compliance/missing data, and a distinction between statistical and clinical significance.

- **Appropriate HRQOL measure selection**

Measure selection is an important consideration when collecting HRQOL data in cancer clinical trials. Many widely validated options are available to measure a range of symptoms, HRQOL attributes and functional domains [25]. However, it is critical to consider specific issues relevant to the cancer clinical and research populations when integrating HRQOL data into cancer clinical research. Cancer patient populations have unique HRQOL, symptomatic and functional concerns that vary dramatically by treatment type [64,65]. For example, as a recent review pointed out, it is essential that ovarian cancer clinical trials include a measure of sexual function [66]. Although HRQOL measures for sexual function in cancer populations exist, they are not part of measure sets commonly used in cancer studies, so researchers must pay special attention to this issue in order to collect this crucial aspect of HRQOL data. As another example, studies that focus on advanced cancer patients will need to account for different symptom-specific end points and a higher risk of patient burden due to questionnaire length. Reviews of measures used in cancer clinical trials have shown that disease-specific measures, such as the EORTC QLQ-C30 [25] and FACT-G [68] are commonly selected measures in clinical trials. The model for both measures and a common approach is to have a core set of general domains of HRQOL. For example, the Physical, Functional, Social and Emotional Function subscales comprise the FACT-G; the functional status, disease-/treatment-related physical symptoms, psychological distress, social interaction, financial/economic impact, perceived health status and overall quality of life comprise the EORTC QLQ-C30. Each of these two questionnaires is supplemented by an appropriate module assessing disease- (e.g., the EORTC QLQ-C30 and CR29 for colon cancer) [67] or treatment-related symptoms (e.g., FACT-Taxane) [68]. There is currently no consensus as to which cancer-specific HRQOL measure is preferable. Often, researchers have selected the EORTC or FACT as a matter of custom or historical practice, without specifying a substantive reason for preferring one measure to the other [69].

General HRQOL measures, such as the Short Form Health Survey-36 (SF-36) [14] are also used, and can be sensitive to changes within cancer populations [35]. Concerns about measure sensitivity to distinguish general declines due to illness from trial-specific effects have limited the use of generic measures in cancer trials [70]; when they are used it is often in addition to disease-specific measures [56].

While the proliferation of many different HRQOL assessment measures hinders broader population comparisons [64], current work developing new measures based in psychometric theory has the potential to address this issue. For example, the Patient Reported Outcome Information System (PROMIS) was developed using psychometric methodologies that allow item linking, opening up future opportunities for backward compatibility with scores from legacy measures [71]. This work will allow for additional assessment options and more streamlined interpretation of instrument scores.

- **HRQOL assessment timing considerations**

General timing issues

There are three points during the trajectory of cancer treatment and survival when incorporating HRQOL measurement into clinical trials is particularly useful in monitoring the impact of cancer and its treatment on patients. The first is during adjuvant therapy, given when the patient is free of detectable disease but has a risk of recurrence. In this setting, a key objective is the identification of mild or brief toxicities. Cancer researchers are continually testing adjuvant interventions to prevent or ameliorate treatment-related symptoms. For example, skin toxicities from EGFR inhibitors and peripheral neurotoxicity symptoms are particularly bothersome, and depending upon their severity, can affect compliance with trial/treatment dosing [72,73]. HRQOL measures can document the extent to which these side effects compromise the patient’s daily activities and ability to function.

The second setting is in the treatment of advanced metastatic cancer. In this setting, the current regimens often only improve survival by a few months and are rarely curative, thus quality of patient survival may be more important than the duration of survival. HRQOL measures provide an intuitively appropriate way of detecting successful palliation [18]. In this population in particular, the relationship between patient symptoms and broader concepts of HRQOL are key issues in the evaluation and interpretation of clinical trial findings. While conceptually, a link between symptom severity and function is anticipated, this is not always supported in the research. A meta-analysis of the cachexia-related treatments for advanced cancer patients showed no differences in HRQOL [74],

---

**Clinical Trial Perspective**

Jensen, Moinpour & Fairclough

---

566

www.future-science.com

future science group
Assessing health-related quality of life in cancer trials

Protocol-specific timing issues
When adding HRQOL measures to a cancer clinical trial, it is important to think carefully about the time points when they are administered. Table 1 summarizes some of the factors that should be considered in order to select clinically meaningful time points, as well as time points that are ‘fair’ for all treatment arms under study. Most factors require discussion with clinicians involved in the trial. For example, one may want to document HRQOL status at known points of remission or deterioration for the particular cancer site, or assess patient HRQOL at the earliest point when an agent could be expected to have a positive effect on the disease to determine whether there would be a comparable effect on HRQOL. The clinicians involved in the trial have previous experience with the agent and can be good sources of suggestions for meaningful time points. Tang et al. discuss the need for careful attention to assessment timing issues when end-of-life or palliative care interventions are being evaluated (Table 1) [80]. The following references also address timing issues [46,27].

Ensuring compliance with these assessment time points is another consideration. Factors affecting compliance include patient response burden, staff administration burden and quality control issues (Table 1). Strategies that balance the value of patient reports with burden should be considered in the initial study design and assessment selection. The total number of assessment questions, length of time to complete, and specific population characteristics (e.g., age and illness severity) should each be considered and adjusted if necessary to lower patient burden. The more frequent the HRQOL assessments, the more work is involved. Clinical staff will spend more time administering questionnaires, the data coordinating center will face higher volumes of data and will require greater quality control, and statisticians will have more complex datasets to analyze. Quality control problems, such as issues with data collection and management, and variability in completion date, can be reduced through the establishment of specific protocols and procedures. For example, the assessment time window [81] should ideally be controlled so that it occurs at the beginning of each treatment cycle, before the patient receives that cycle’s agent. However, treatment delays are common and tracking those delays at the data center presents a serious challenge, since data centers may not learn of the delays in a timely fashion, and some systems may not be able to accommodate ongoing demands for due date revisions. Given that delays occur for many reasons, not just because a patient experiences serious side effects from the agent (e.g., inability to get to the clinic due to non-health-related issues), two common options are to:

■ Specify an assessment time prior to the treatment date (e.g., 1–3 days prior) or;

■ Count follow-up assessments from the date of the first assessment without connecting them to the delivery of treatment.

The first option is preferred because it allows some flexibility while still following the measurement intent. The second strategy is problematic with respect to evaluation of the treatment agents because it does not permit the consistent examination of the effect of
a treatment on HRQOL. Even small quality control decisions such as this one can have important implications for systems used to monitor timely submission of questionnaires.

In trials with high levels of patient mortality or treatment discontinuation, a critical design question is how long to attempt follow-up assessments of HRQOL. Two issues should be considered. The first is whether to continue HRQOL follow-up assessments past the discontinuation of treatment. Traditionally, assessments of side effects have stopped with the discontinuation of treatment for obvious reasons (e.g., patients are too sick to continue treatment). This model has often been extended to HRQOL domains without careful thought about the differences between symptoms and measures of function. Clinical studies examine links between a therapy and acute side effects, whereas broader measurements of patient well-being and functioning examine the reach of a treatment’s impact on day-to-day functioning. Treatment discontinuation is often based on pre-symptomatic evidence (e.g., the radiologic

<table>
<thead>
<tr>
<th>Table 1. Assessment schedules: important issues to consider.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>Baseline assessment is mandatory</td>
</tr>
<tr>
<td>Data collection prior to administration of treatment and/or discussions with clinical staff</td>
</tr>
<tr>
<td>Timing of HRQOL assessments should be similar for all treatment arms</td>
</tr>
<tr>
<td>Natural course of the disease</td>
</tr>
<tr>
<td>Disease stage</td>
</tr>
<tr>
<td>Timing of important clinical events or monitoring</td>
</tr>
<tr>
<td>Effects associated with the treatment course or administration</td>
</tr>
<tr>
<td>Completion of treatment and/or a short time after completion of treatment</td>
</tr>
<tr>
<td>Scheduling issues for special populations</td>
</tr>
<tr>
<td>Compliance with assessment schedule</td>
</tr>
</tbody>
</table>

HRQOL: Health-related quality of life; XRT: Radiotherapy.
Adapted with permission from [79].
assessment of tumor size); based on this type of criterion, the impact of treatment failure is likely to occur after discontinuation. Therefore, continuing HRQOL assessments is necessary to assess the real impact of a treatment choice. Investigators should try to maintain this schedule even if a patient discontinues therapy. Follow-up data from such patients can be more difficult to collect but analyses that do not include these patients can bias conclusions about HRQOL in the positive direction [27].

The second issue is how long to continue assessment in populations with high levels of mortality, because after some point the sparse data on a very select group of survivors will not generate useful information. Exactly where this cutoff should occur has not been carefully studied. However, it is recommended that investigators consider carefully the value of follow-up past the expected median survival for that patient population.

**Quality control issues & strategies**

As noted above, it is important to describe the HRQOL outcomes in all relevant sections of the protocol [50]. This emphasizes the role of HRQOL data as important outcomes in the trial and makes certain that those collecting the data understand why the data are being collected as well as the correct procedures for collecting them (quality control at work). It is important to establish a centralized quality control monitoring procedure for tracking the submission of required HRQOL assessments. When HRQOL assessments are included in the study calendar, research staff and clinicians will be reminded about the specific times at which these measures must be administered and the importance of the HRQOL measures to the study will be reinforced. Staff at participating institutions can assist in this effort, particularly if one or more are involved in helping to coordinate the HRQOL component of the trial. Ongoing training in the administration of HRQOL questionnaires needs to be available; the most feasible method for doing so is to provide such information online and to state in the protocol how this training information will be accessed.

**Administration methods**

HRQOL questionnaires have historically been administered solely on paper. However, recent technological advances have included the development of new electronic data capture platforms that are able to collect HRQOL data. These electronic methods present improvements over paper-based administrations by reducing patient, staff and financial burden [81] and many systems provide the option of at-home follow-up patient assessments over the internet [82]. Data are automatically stored and scored in real-time when entered by the patient. Missing and incomplete data are quickly identified and relevant items can be re-administered to patients at the end of the questionnaire; with this approach, Buxton *et al.* reported that no missing items remained unanswered [83]. Patient benefits include a quicker completion time, lowering burden and increasing satisfaction [84,85]. Electronic administration also allows for computerized adaptive testing [86,87], which uses a real-time selection of items based on a patient’s previous responses. Although there are barriers to clinical researchers’ adoption of these methods [88], they offer the advantages of reduced patient response burden and the ability to use smaller sample sizes while still identifying clinically meaningful differences in HRQOL.

**Trial conduct**

There are important considerations for the administration of HRQOL surveys in a trial. How patients are approached is a key consideration when a HRQOL survey is administered. During the administration of surveys, patients should be encouraged to complete a survey as independently as possible to ensure unbiased responses. For an interviewer, this entails being present to answer patient questions and provide clarification for questions and words that are unclear, while not providing interpretation or response options to the patient. If a caregiver or family member is present during the questionnaire administration, care must be taken to ensure that the questionnaire is completed by the patient without influence from others. Research has shown that family members and caregivers can have very different views of a patient’s HRQOL than the patient themself [89]. Prompting or involvement by the caregiver may cause a patient’s responses to incorporate the caregiver’s perspective, adding measurement bias to the results. If the survey is administered as an interview, additional attention should be paid to any possible social desirability bias, where patients may respond to questions as they believe the researcher would like them to – in ways they would not have if their responses had been anonymous.

**Key challenges in analysis & interpretation of HRQOL**

Three major challenges exist when analyzing and interpreting HRQOL measures in populations of cancer patients. First, due to high levels of morbidity and mortality, missing data is common in cancer populations. Second, the existence of multiple end points due to the multivariate nature of the patient-reported data collected in cancer clinical trials presents both analytic and interpretation challenges. Third, the
interpretation and clinical significance of HRQOL scores, although ignored in the early days of HRQOL research, now present a prominent trial design issue.

**Missing data**

While HRQOL data are a rich source of information, data analysis is often complicated by problems of missing information. Patients sometimes fail to complete HRQOL assessments because of negative events they experience, such as treatment toxicities, disease progression and death. Because not all patients are subject to these missing observations at the same rate, especially when treatment failure or survival rates differ between arms, available observations are not always representative of the total group; analyses using only complete observations are, therefore, potentially biased.

In a clinical trial of patients with a good prognosis (e.g., adjuvant therapy recipients or post-treatment survivors), most missing data are preventable. However, in trials with significant morbidity or mortality, missing assessments are inevitable and impact both the analysis and interpretation of the results. There are three types of analytical techniques that account for missing data: techniques that use part of the data, all available data, or all available data plus auxiliary information. These three types correspond to three assumptions about the missing data.

The first missing data technique assumes that assessments are missing completely at random, for reasons unrelated to the patient’s health status. These types of missing data are rare. Analysis methods that assume data are missing completely at random include those that exclude patients with any missing data, such as multivariate analysis of variance, repeated cross-sectional tests at each assessment and unadjusted generalized estimating equation methods. In the presence of missing data due to morbidity or mortality, all of these methods will overestimate HRQOL and underestimate symptom measures.

The second approach assumes that assessments are missing at random (MAR). The occurrence of missing assessments is assumed to be independent of the patient’s current HRQOL after adjusting for observed HRQOL and other covariates. Analysis methods include maximum likelihood estimation of mixed models using either a repeated measures [89] or growth curve framework [90], multiple imputation (MI) techniques [91,92] using available data and covariates, and doubly robust generalized estimating equations [93]. While these methods tend to underestimate the decline in HRQOL over time within groups, they can provide relatively unbiased estimates of differences between treatment groups when there are similar patterns and reasons for dropout across groups. Most experts recommend these methods for the primary analysis, but because the impact of the missing data is generally unknown, sensitivity analyses are recommended using one of the methods described below [27,94–97,202].

The third technique for missing data is designed to address assessments that are not MAR, where missing assessments are more likely to occur in patients with poorer morbidity or mortality outcomes. Recommended analysis methods include pattern mixture models [98,99], joint or shared parameter models [100,101], and multiple imputation with surrogate or auxiliary information [27]. In all of these approaches, the missing data are assumed to be conditionally MAR. In pattern mixture models, the assumption is that the data are MAR within each pattern. In joint models, the data are assumed to be MAR conditional on the other outcomes. MI techniques assume the data are MAR conditional on the variables included in the imputation models.

Each approach using not-missing-at-random methods has advantages and disadvantages. Pattern mixture models are attractive because it is not necessary to specify a model for the missingness mechanism, but this is balanced by the need to extrapolate curves or place restrictions on the models to estimate all the parameters in each pattern. The joint models and MI require that surrogate or auxiliary information has been prospectively gathered. Notably, there are currently no formal tests to determine which of the models yields the correct result, so sensitivity analyses involving either multiple methods or variations on a selected method should be considered in the analysis plan.

An unresolved question is how to account for HRQOL assessments scheduled to occur after a patient has died. While some argue that assigning a value to HRQOL or even symptoms after death does not make sense, all analysis methods either explicitly (e.g., MI) or implicitly (e.g., expectation-maximization algorithm) impute the assessments that occur after death. Among the explicit techniques is imputing the minimum possible score on the scale. While this approach is reasonable for some scales where a score is explicitly anchored to zero (e.g., utilities, functional well-being), it will not work for symptom scales where a zero could mean, for example, that the deceased patient is experiencing severe symptoms such as nausea, vomiting or pain. Kurland et al. provide an overview of different approaches, arguing that the choice should be made based on the research aims [102]. In fact, the research question is the most critical issue. For example, if the goal is to compare treatments from an intent-to-treat perspective, a
method that penalizes the arm with poorer survival would be appropriate. In contrast, if the goal is to describe the trajectories of survivors conditional on the duration of survival, the estimates could be displayed until the time of death. In this latter approach, it is important to be cautious about potential selection bias if making comparisons between treatment arms.

**Summarizing multiple HRQOL end points**

PROs in cancer research typically assess multiple HRQOL symptoms and domains over time. This introduces the potential for multiple end points that may inflate the number of Type I errors and adds complexity to the interpretation of the results. Data analysis strategies need to be driven primarily by well-defined research questions. Two specific considerations need to be addressed when deciding how to incorporate multiple end points: whether to summarize data from different measures at each assessment or over time, and whether to summarize HRQOL scores at the individual patient or group level.

The first dimension is the measured study outcome. Combining multiple symptoms or domains into a composite score will increase the likelihood of detecting small to moderate differences between treatments. However, this method has the disadvantage of only identifying changes that are in the same direction, thus obscuring changes that may occur in different directions by the treatment arm. Interpreting results based on composite measures may also be misleading unless the components are also examined individually.

Summary measurement across time, such as the overall slope or the area under the curve (AUC), can reduce multiple measurements to a single measure that is easier to interpret than either F-statistics or multiple t-tests. The choice of summary measure (slope vs AUC) will depend on the expected trajectory (linear vs nonlinear) and on whether the study focuses on early or late patient outcomes (slope over post-treatment vs AUC over early treatment).

Finally, there are two strategies for forming the summary measures: at the individual level (raw data summaries) or at the group level (parameter estimate summaries). When there are no missing data, the results of the two approaches are the same. With missing data, calculating raw data summaries becomes burdensome because explicit rules must be developed and defended. Strategies for parameter estimate summaries require addressing the missing data issues described above.

**Interpretation & clinical significance**

Given appropriate trial design and adequate presentation of HRQOL data as discussed above, it is equally important to provide context or guidelines for clinical interpretation of differences, or change scores in studies where statistically significant differences are observed. Previous publications have described approaches for determining clinically important changes, such as distribution-based measures (e.g., effect size) and anchor-based measures (e.g., a person’s perception of the extent of change in a HRQL domain) [105–109]. Recently, Wyrwich et al. reviewed methods for interpreting the clinical importance of change over time in HRQOL scores [110]. Example publications of guidelines for interpreting differences/change for the EORTC QLQL-C30 and its modules include King [111], Osoba et al. [112], and Maringwa et al. [113]. Example publications of guidelines for minimal clinically important differences for the FACT and Functional Assessment of Chronic Illness Therapy include Cella et al. [114], Yost et al. [115] and King et al. [116]. Cocks and colleagues have also proposed a new method of determining clinical significance using a totally different approach [114–119]. Clinicians were asked to predict trivial, small, medium, and large effects with respect to clinical relevance but their judgments were based only on the clinical information from the articles they were given (not the HRQOL scores). That is, the clinicians were asked to estimate which HRQOL domains would be affected and how much change they expected for the patient given the available clinical information.

**Value of HRQOL outcomes in determining clinical care**

There is obvious interest in incorporating HRQOL measures in the routine monitoring of patient status during care as well as extrapolating results of well-controlled clinical trials to individual patients seen by physicians. There are two main issues with these goals for HRQOL data. The first is individual deviation of patient trajectories relative to the group mean. Clinical trials present the average HRQOL score for patients receiving treatment A versus treatment B, but we know that within each treatment arm, some patients are responding well, some not so well, and some are not responding at all [120]. Large treatment effects facilitate extrapolation to patients but individual variation will likely play a much larger role when small effects are observed. Therefore, individual variation make it less credible to infer that these effects will hold for a new patient considering treatment [120]. While clinical studies with measurements of HRQOL provide important information about patient populations, researchers and clinicians should be aware of the measurement error present in any single score for an individual patient [121]. For example, in clinical trials...
where results are presented for large groups of patients, a change of five points has been documented as clinically important for the FACT Trial Outcome Index for lung cancer patients (which has three subscales: physical well-being, functional well-being and the lung cancer symptom module) [122]. A clinically important change for an individual would need to be larger, in this case 15–20 points [121]. This issue is relevant to both extrapolation of trial results as well as to using patient scores on HRQOL measures at specific points to monitor the effect of clinical care on patient HRQOL. Regarding the routine monitoring use of HRQOL measures for individual patients, using repeated measures helps reduce this error because it improves measure precision and allows the use of sophisticated statistical models of change [123]. Another method for reducing measurement error is to use computer adaptive testing measures, which have been adopted as one format for the PROMIS initiative [124]. Computer adaptive testing measures allow fewer items to be administered (a good feature for a busy clinician’s office) while generating increased precision in the estimation of the HRQOL area of interest [86]. The movement to incorporate patient-reported data into electronic medical or health records (EMR/EHR) in practice-based networks will encourage the use of computer-based assessment in medical care settings [125] and make electronic capture of HRQOL/PRO data in the clinic more feasible. The Critical Path Initiative (C-PATH) [203] has established an electronic PRO Consortium to develop guidelines to incorporate PROs in EMR/EHR.

The second issue related to the extrapolation of HRQOL data from clinical trials is that clinical trial participants are a highly select group of patients who are treated in a very uniform manner. In practice, the patients will represent a broader population and the treatment delivery will be more varied. The efforts described above to collect data electronically in clinics and to merge HRQOL data with EHR will also help address this second issue because over time increasingly more HRQOL data will be collected for patients receiving a wide variety of treatments; the use of merged HRQOL and clinical data will provide better characterizations of patients on these trials. Improvements in the design and reporting of HRQOL outcomes will increase their value to clinicians [85,62,63]. Incorporating PROs in clinical settings will present new challenges, but can provide an important context in which HRQOL information from clinical trials and practice can be evaluated in tandem [116]. The third meeting of the Clinical Significance Consensus Meeting Group addressed how to translate what we have learned about HRQOL assessment into the clinical practice setting [117–129].

Two reviews of the extent to which HRQOL outcomes have made a difference in the interpretation of clinical trial results and in the interpretation of these results for clinical care of cancer patients reached similar conclusions when the context was primary management of the cancer (surgery, hormone therapy or radiation therapy) [110,111]. The Goodwin et al. evaluation was restricted to randomized trials for breast cancer treatment. The authors noted that when equivalent medical outcomes were observed in the primary management setting (i.e., treatment of the primary breast tumor/local therapy), HRQOL data was useful in recommending treatments [130]. Goodwin et al. did not see added benefit from HRQOL for patients with breast cancer in the adjuvant therapy, metastatic disease or symptom control, supportive care or longer-term follow-up settings contexts. Trials involving a psychosocial intervention showed improved HRQOL outcomes in 10 out of 11 trials in the adjuvant setting, less value in the metastatic setting, and a benefit associated with the only symptom control intervention (a comprehensive menopausal assessment). The authors did not find support for including HRQOL outcomes in every breast cancer trial, but advised inclusion when treatment equivalence for medical outcomes was expected or when there was a strong rationale for an effect on HRQOL or where the outcome of interest was one that was best supplied through a patient report. Blazeby et al. [131] evaluated the value of including HRQOL outcomes in surgical oncology trials – breast, stomach, prostate, cerebral metastases, rectal, larynx, esophageal, testes, colon, perianal, melanoma and pancreas (most with localized or operable disease but some trials for advanced stage disease). Studies could include surgery for any stage or tumor grade but more than two out of three of the reviewed studies addressed local or locally advanced disease. The authors noted that HRQOL outcomes influenced decisions about care in 22 out of 33 reviewed trials and only one study involved metastatic disease; in four of these studies, the HRQOL data were helpful in the informed consent process. HRQOL data were more consistently used in the trials for localized disease, which is consistent with the Goodwin et al. report [130]. One might think that HRQOL data would matter more in the advanced stage setting where clinical outcomes are often small, leading to increased interest in patient HRQOL (at what cost come the small clinical gains?). This is an area that requires more research and documentation but represents a salient question in a time of decreasing resources for clinical cancer trials. To date, HRQOL outcomes in trials meeting criteria
Reports. In addition, we recognize patient advocates who have contributed their health-related quality of life to cancer clinical trials and have supported informed decision making in this disease setting.

**Conclusion**

HRQOL has been established as an important, distinct outcome in cancer clinical trials, and should be considered when planning clinical trials but with a strong rationale for the HRQOL research question. When HRQOL measures are included in a trial, several considerations must be addressed before, during and after the study. Fortunately, checklists and protocols have been developed specifically for HRQOL assessments, which can provide a useful framework for researchers. With the development of new electronic-based assessment options and psychometric methods, the assessment of PROs, such as HRQOL, will allow for further opportunities to both incorporate and evaluate the patient perspective in clinical trial settings, while enabling exciting potential applications in clinical practice settings.

**Future perspective**

There is increasing pressure from multiple stakeholders to include patient-centered outcomes, such as HRQOL, in the evaluation of interventions in clinical trials and the larger observational studies used for comparative effectiveness research. The future of HRQOL measurement will involve the increasing substitution of electronic measures for existing paper-based methods. This will allow for practical benefits, such as decreased data entry burden and increased quality control, while providing new opportunities, such as web-based reporting and limiting the need for clinic-based assessments. These technological developments, coupled with computer-based item administration that shifts assessments from fixed forms to item banks, will facilitate comparisons of patient scores to population norms and lower patient burden.

**Acknowledgements**

The authors acknowledge the patients who have participated in cancer clinical trials and have contributed their health-related quality of life reports. In addition, we recognize patient advocates who have become more involved in the trial design process, which helps make the process more patient friendly and more likely to increase patient participation.

**References**

12. Nayfield SG, Ganz PA, Moinpour CM,
Assessing health-related quality of life in cancer trials

Clinical Trial Perspective


117 King MT, Stockler MR, Cella DF et al. Meta-analysis provides evidence-based


Websites


203 Critical Path Institute. www.c-path.org