UC Irvine
UC Irvine Previously Published Works

Title
Challenges posed by non-random missing quality of life data in an advanced-stage colorectal cancer clinical trial.

Permalink
https://escholarship.org/uc/item/3314v2s8

Journal
Psycho-oncology, 9(4)

ISSN
1057-9249

Authors
Moinpour, CM
Sawyers Triplett, J
McKnight, B
et al.

Publication Date
2000-07-01

DOI
10.1002/1099-1611(200007/08)9:4<340::aid-pon466>3.0.co;2-f

License
https://creativecommons.org/licenses/by/4.0/ 4.0

Peer reviewed

CHALLENGES POSED BY NON-RANDOM MISSING QUALITY OF LIFE DATA IN AN ADVANCED-STAGE COLORECTAL CANCER CLINICAL TRIAL

CAROL M. MOINPOUR[a,*], JULIA SAWYERS TRIPLETT[b], BARBARA MCKNIGHT[c], LAURA C. LOVATO[a], CHRISTINE UPCHURCH[d], CYNTHIA G. LEICHMAN[e], FRANCO M. MUGGIA[f], LEORA TANAKA[g], WENDY A. JAMES[h], MARTHA LENNARD[i] and FRANK L. MEYSKENS JR[j]

[a] *Southwest Oncology Group Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA*
[b] *Nashville, TN, USA*
[c] *University of Washington, Seattle, WA, USA*
[d] *Bothell, WA, USA*
[e] *Roswell Park Cancer Institute, Buffalo, NY, USA*
[f] *Kaplan Medical Center, New York, NY, USA*
[g] *University of California at Davis, Sacramento, CA, USA*
[h] *Kings County Minority-Based Community Clinical Oncology Program, Brooklyn, NY, USA*
[i] *Schumpert Medical Center, Shreveport, LA, USA*
[j] *University of California at Irvine Cancer Center, Irvine, CA, USA*

SUMMARY

Effects of variations in agent, dose, and route of treatment administration on patient reported quality of life (QOL) were examined for 279 patients enrolled on a seven-arm randomized clinical trial (S8905) of 5-FU and its modulation for advanced colorectal cancer. Patients completed QOL questionnaires at randomization and weeks 6, 11, and 21 post-randomization with five QOL endpoints considered primary: three treatment-specific symptoms (stomatitis, diarrhea, and hand/foot sensitivity); physical functioning; and emotional functioning. Patient compliance with the QOL assessment schedule was good, supporting the feasibility of including QOL measures in cooperative group trials. However, death and deteriorating health produced substantial missing data. Cross-sectional analyses indicated that the seven therapeutic arms did not differ in their impact on QOL. Unfortunately, longitudinal analyses of the QOL data were inappropriate given non-random missing data. Graphical presentation of non-random missing data identified the seriousness of this problem and its effect on potential conclusions about QOL during treatment. This problem appears to be particularly challenging in the context of advanced-stage disease. Failure to recognize the presence of non-random missing data can lead to serious overestimates of patient QOL over time. Copyright © 2000 John Wiley & Sons, Ltd.

INTRODUCTION

Clinical and quality of life issues in advanced colorectal cancer

Cancer of the colon and rectum is a common solid tumor, with approximately 140000 new cases diagnosed annually (Wingo *et al.*, 1998).

Treatment for patients with advanced or recurrent colorectal cancer is usually palliative (Bengt *et al.*, 1994). Systematic measures of patient quality of life (QOL) complement information about optimal therapeutic effects by addressing the degree of palliation achieved as a result of treatment.

The QOL of patients with advanced colorectal cancer treated with 5-FU has received little attention in clinical research. A six-arm trial reported by Poon *et al.* (1989) examined the effect of various biochemical modulations of 5-FU on the QOL of patients with advanced colorectal cancer. A regimen containing 5-FU and low-dose leucovorin (IV injection) showed improvements in

---

* Correspondence to: Southwest Oncology Group Statistical Center (MP-557), Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Box 19024, Seattle, WA 98109-1024, USA. e-mail: carolm@swog.fhcrc.org

three physician-rated measures of patient QOL (improvement in performance status, weight gain of 5% or more, improvement in symptoms). Hill et al. (1995) examined the effect on QOL of another agent, interferon alfa-2b (IFN), added to continuous infusion 5-FU for patients with advanced colorectal cancer. However, in this study, a comprehensive patient assessment of QOL was included, the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 (Aaronson et al., 1993). Patients treated with protracted venous infusion 5-FU and IFN delivered subcutaneously did not differ from patients treated with protracted venous infusion 5-FU alone with respect to most clinical variables or domains of the EORTC QLQ-C30 (e.g. global QOL and psychosocial functioning). Certain physician-rated treatment-related toxicities were greater on the combined therapy arm (e.g. leukopenia and mucositis), but on the patient-rated symptom scales, only appetite loss differentiated treatment arms with more appetite loss in the interferon modulation arm. Global QOL and emotional functioning did not differ by arm.

Describing missing QOL data in studies of advanced-stage disease

Recently, investigators have reported results of randomized trials in advanced-stage disease cancer sites in which missing data have been problematic for conducting longitudinal analyses (Troxel et al., 1998). A conference was held in Switzerland in July, 1996 with two objectives: improving submission rates for QOL assessments in cancer clinical trials (Bernhard et al., 1998; Moinpour and Lovato, 1998) and exploring appropriate options for conducting longitudinal analysis when substantial efforts to increase questionnaire submission rates still result in missing data (Troxel et al., 1998). Hopwood et al. (1994) commented on this problem and suggested the use of graphic techniques, which have been extended in the analyses presented below. Donaldson has also recommended use of graphic techniques and described the bias associated with missing data (Bertsch and Donaldson, 1995); his suggestions guided methods to address this bias in several projects (Moinpour, 1994). It is necessary to inform readers about the extent of missing data in the analyses, as we have done, and note its presumed effect on conclusions about the effects of

treatment on QOL. Staquet et al. (1996) noted the importance of describing missing data patterns in the report of QOL results; various analysis methods have been suggested (Staquet et al., 1998).

In this paper, we report the results of a QOL study conducted by the Southwest Oncology Group (SWOG) that failed to find QOL differences by treatment arm for patients with advanced colorectal cancer. The therapeutic trial, reported elsewhere (Leichman et al., 1995), will be briefly summarized. In particular, we will present an evaluation of the extent and nature of missing data found in this trial. As a result, we are able to highlight the challenges presented by non-random missing data for the analysis of change in QOL during treatment and the potential for serious overestimates of patient QOL over time when such missing data are not acknowledged. We also address the potential usefulness of baseline QOL data when they are prognostic for survival.

METHODS

Design and objectives

SWOG-8905 was a seven-arm Phase II/III randomized screening trial that investigated the role of biochemical modulators, route of administration differences, and dose intensity on survival of patients treated with 5-FU (Leichman et al., 1995) (see Figure 1). Treatment arms were designed to allow multiple, parallel comparisons. The single agent 5-FU bolus regimen (arm 1) was compared to 5-FU paired with a biochemical modulator: leucovorin (arm 2, low dose; arm 3, high dose); PALA (arm 7). In addition, route of administration and schedule differences were examined: bolus injection (arms 1–3); protracted or continuous infusion 5-FU for 28 days via a battery-powered pump (arms 4 and 5); infusion of 5 FU over a
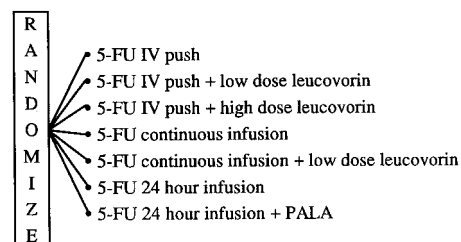
Figure 1. SWOG 8905 study schema.

24-h/day for 28 days using a pump outside the hospital or an infusion bag in the hospital (arms 6 and 7). Single agent continuous infusion 5-FU was compared to continuous infusion plus leucovorin (arms 4 and 5) and high dose, intermittent infusion of 5-FU with the same regimen plus PALA (arms 6 and 7).

The following eligibility criteria were employed for the therapeutic trial: recurrent or disseminated metastatic colorectal cancer; no previous chemotherapy for disseminated disease; at most, one prior immunotherapy regimen for metastases; at most, one prior adjuvant therapy and only if more than 1 year since discontinuation; prior surgery or radiotherapy only if more than 2 weeks since treatment; SWOG performance status of 0–2; adequate hematologic, renal, hepatic, and cardiac function. Patients enrolled on the therapeutic trial who could complete a questionnaire in English were eligible for S9045, the companion QOL study. The planned sample for the QOL study was 40 patients per arm or 280 patients.

Endpoints and treatment comparisons. The overall objective of the QOL study was to examine the effect of different treatment combinations on patient-reported QOL: three treatment-specific symptoms (mouth pain, diarrhea, and hand/foot sensitivity) (Leichman *et al*., 1990); emotional functioning; and physical functioning. Table 1 describes the *a priori* hypotheses associated with each primary endpoint and assessment period. Although SWOG-9045 addressed other components of QOL (Moinpour *et al*., 1989), the focus of this paper is the set of QOL components designated primary endpoints. We also examined

the ability of a baseline measure of general symptom status to predict survival.

QOL assessment schedule. The SWOG QOL questionnaire was administered at randomization and at weeks 6, 11, and 21 post-randomization. To minimize missing data due to deterioration of health status and death, the length of QOL follow-up was based on an estimated median survival of 24–46 weeks for patients treated with single-agent bolus 5-FU in previous studies (Nordic Gastrointestinal Tumor Adjuvant Therapy Group, 1989; Petrelli *et al*., 1989; Leichman *et al*., 1990). In this trial, our report of QOL is with respect to the past week because we wanted patients to be rating QOL at the same calendar times from randomization. Although we are missing acute toxicities that may occur just after treatment on the IV push and 24-h infusion arms, we are most interested in the overall effect of treatment at the stated time points. Two of the five primary QOL endpoints addressed the extension of treatment arm-related acute toxicities to broader areas of functioning (emotional and physical functioning).

QOL assessment methods

When QOL is assessed in SWOG studies, the goal is to compare the effect of different treatment arms on the day-to-day functioning of patients receiving these treatments. For this reason, a comprehensive, patient-reported measure of QOL is included addressing physical, role, emotional, and social functioning as well as general symptom status and treatment-related side effects

Table 1. *A priori* hypotheses associated with primary endpoints

Endpoints	Week tested[a]	Hypotheses
Stomatitis	6[b], 11, 21	More stomatitis in arms with leucovorin (arms 2, 3, 5) versus no leucovorin (arms 1, 4, 6, 7)
Diarrhea	6, 11, 21	More diarrhea in arms with IV push administration (arms 1–3) versus infusion (arms 4–7)
Hand/Foot sensitivity	6, 11, 21[b]	More sensitivity in arms with protracted (>24 h) infusion (arms 4, 5) versus bolus dosing (arms 1–3)
Emotional functioning	6, 11, 21	Better emotional functioning in arms with protracted (>24 h) infusion (arms 4, 5) versus IV push or 24 h infusion (arms 1–3, 6, 7)
Physical functioning	6, 11, 21	Worse physical functioning because of more fatigue and diarrhea in arms with IV push (arms 1–3) versus infusion (arms 4–7)

[a] The *p*-value was adjusted for testing five comparisons at each of three follow-up assessments.
[b] Hypothesized difference for stomatitis was expected to be greater at 6 weeks; for hand/foot sensitivity, at 21 weeks.

(Moin pour *et al.*, 1989). This approach is consistent with that adopted in other QOL research in cancer clinical trials (Aaronson *et al.*, 1988; Gotay *et al.*, 1992; Nayfield *et al.*, 1992; Osoba, 1994). The QOL questionnaire was administered prior to the administration of treatment and QOL was rated with respect to the previous week; questionnaires were most frequently completed in the clinic. The components of the questionnaire are described below.

Treatment-specific symptoms. Stomatitis, diarrhea, and hand/foot sensitivity were designated as primary endpoints (see Table 1). Physician and nurse investigators who had treated patients with colorectal cancer developed all three measures. The QOL Study Coordinator determined the appropriateness of the symptom items as well as the entire questionnaire for this patient population by conducting a pilot test with ten patients.

The treatment-specific questions address symptoms usually included in the set of physician-rated toxicities; these items are presented individually and not summed for a total score. Treatment-related symptom items have five- or six-level response choices ranging from no problem with the symptom to severe problems. For analysis purposes, the three primary symptom endpoints were described as dichotomous variables: no mouth pain or some but not bothersome versus bothersome to unbearable mouth pain; none or occasional diarrhea versus frequent diarrhea (several times a week to several times a day); none or some hand/foot sensitivity versus bothersome to unbearable hand/foot sensitivity. Patients who had a colostomy were omitted from analyses examining the effect of treatment arm on occurrence of diarrhea.

General symptom status. The Symptom Distress Scale was used to assess patient symptom status (McCorkle and Young, 1978; McCorkle and Benoliel, 1983; Young and Longman, 1983; McCorkle, 1987McCorkle *et al.*, 1989, 1998). The scale's items address disease- and treatment-related symptoms commonly experienced by patients with cancer. McCorkle and colleagues have documented the psychometric properties of the Symptom Distress Scale (McCorkle and Young, 1978; McCorkle and Benoliel, 1983; Young and Longman, 1983; McCorkle, 1987; McCorkle *et al.*, 1989; McCorkle *et al.*, 1998). With permission

of the questionnaire developers, two of the 13 items (cough and outlook) were omitted from the measure, leaving 11 items. Internal consistency reliability was 0.84 for the 11-item version of the Symptom Distress Scale administered to this patient sample ($n = 287$). Higher scores reflect worse symptom status.

Physical and emotional functioning. The Physical Functioning Scale and the Mental Health Index address two basic QOL domains from the General Health Survey Short Form 20 (SF-20) and the 36 (SF-36) developed as part of the Medical Outcomes Study (Stewart *et al.*, 1988, 1989; Ware and Sherbourne, 1992; McHorney *et al.*, 1993, 1994) and the Rand Health Insurance Experiment (Ware *et al.*, 1980). The SWOG QOL studies were initiated during the transition between the SF-20 and the SF-36 in consultation with the developers of these questionnaires; this included a decision to change the time frame of the Short Form scales from 1 month to 1 week (there is now a 1-week time frame version of the SF-36) (John E. Ware Jr., 1988–1990, personal communications). The SWOG QOL questionnaire also includes the SF-20 Role Functioning Scale, the SF-36 Social Functioning Scale, and the SF-20/SF-36 single item rating of general health.

The ten-item Physical Functioning Scale and the five-item Mental Health Index come from the more recent and final version of these scales, the SF-36 Health Survey (Stewart *et al.*, 1989; Ware and Sherbourne, 1992; McHorney *et al.*, 1993). Psychometric properties of these scales have been established (Stewart *et al.*, 1988, 1989; Ware and Sherbourne, 1992; McHorney *et al.*, 1993). Adequate internal consistency reliability was achieved for the patient sample reported in this paper ($n = 287$): 0.92 for Physical Functioning; 0.80 for the Mental Health Index. Higher scores reflect better functioning.

Statistical methods

Compliance and missing data. A number of quality control procedures followed in SWOG QOL studies have been described elsewhere (Moinpour *et al.*, 1989; Moinpour and Lovato, 1998). A 70% questionnaire submission rate was considered the lowest acceptable for data analysis. To evaluate compliance, the questionnaire

submission rate at each time point was computed as the number of questionnaires submitted at that time point divided by the number of patients alive at that point. The decrease in sample size over time and the reasons for patient dropout were assessed. Average QOL scores at each time were compared for three groups of patients: those patients who completed all assessments, those who failed to complete all assessments due to death or deteriorating health, and those who failed to complete all assessments for other reasons.

Missing data within a QOL scale were handled as follows for Symptom Distress Scale, Physical Functioning Scale, and Mental Health Index total scores. If a patient failed to answer 20% or fewer items for a scale, his/her score for that scale was calculated with the mean of the remaining items (for that patient) substituted for missing items; if the scale had more than 20% of the items missing, the entire scale was considered missing.

Statistical comparisons

QOL treatment arm comparisons were made between combinations of treatment arms varying the route of administration or biochemical modulation at 6, 11, and 21 weeks after randomization (see Table 1). Since five primary endpoints were tested at each of three time points, the probability level for rejecting the null hypothesis at the traditional 5% (two-sided) alpha level was adjusted to 0.003. Power estimated prior to the activation of the QOL study was based on combinations of arms to address the hypotheses described in Table 1 and was deemed to be adequate for the five primary comparisons. The symptom outcome comparisons utilized one-tailed tests, while the physical and emotional functioning comparisons involved two-tailed tests. Wilcoxon rank-sum tests were used to examine group differences at each of the three time points for the Physical Functioning Scale and the Mental Health Index. For the three symptom variables, chi-square tests were employed on binary data at each of the time points. To address the effect of actual sample size on ability to detect treatment arm differences, we present 95% confidence intervals (CI) for each proportion or mean estimate. The width of the CI represents the precision of the estimates.

RESULTS

Therapeutic trial results

The therapeutic trial was activated in August 1989 and closed in January 1993 with a total accrual of 620 patients; 599 patients were determined eligible for clinical evaluation.

No significant survival differences were detected when arms 2–7 were compared to standard single-agent bolus 5-FU (24). Patients treated with PALA (arm 7) showed the shortest survival duration. Although survival for the remaining arms did not differ statistically, it was better in the 5-FU infusion regimens. A follow-up trial with continuous low-dose 5-FU versus intermittent high-dose 5-FU is underway.

QOL accrual

The QOL study opened in October 1990 and closed in January 1993. Two hundred and eighty-seven patients were registered to the QOL companion study for the therapeutic trial. Eight of these patients were later deemed ineligible for the therapeutic trial and so were deleted from the QOL analyses. Of the remaining 279 patients, four did not complete a baseline QOL questionnaire, leaving 275 patients for the QOL analyses. Patient characteristics are summarized in Table 2.

Table 2. QOL study patient characteristics ($n = 279$ eligible patients)

Age	
Minimum	21
25th percentile	55
Median	63
75th percentile	69
Maximum	84
Number of patients with baseline QOL assessment (%)	
Sex	
Male	176 (63)
Female	103 (37)
Race	
White (Non-Hispanic)	208 (75)
Black (Non-Hispanic)	40 (14)
Hispanic	18 (6)
Asian or Pacific Islander	10 (4)
American Indian	3 (1)
Other	0 (0)

QOL questionnaire submission rates:
advanced-stage disease-related attrition

Table 3(A) presents questionnaire submission rates for the four time periods. However, this table obscures the impact of the missing data problem in this study. To see the effect of missing data for longitudinal analyses, we provide the sample size available for the Symptom Distress Scale: 159 of the 279 (57%) eligible patients completed QOL questionnaires for all four scheduled

assessments; 24 patients completed only the baseline and week 6 assessments; 35 patients completed only the first three assessments; other patterns describe the remaining 67 patients (e.g. 33 patients with only a baseline questionnaire).

Table 3(B) displays the patterns of missing data. Most patients completed all questionnaires up to the one missed. However, there were examples of patients with intermittent patterns. Table 3(C) describes the reasons provided on the Quality of Life Questionnaire Cover Sheet (required

Table 3. Missing data

(A) QOL Questionnaire submission rates

Assessment time	Number alive	n submitted (%)
Baseline	279	275 (99)
6 weeks	272	230 (85)
11 weeks	262	207 (79)
21 weeks	234	182 (78)

(B) Patterns of missing data

Pattern of missing data	n	%
Baseline only	33	11.8
Baseline and 6 weeks	24	8.6
Baseline, 6 and 11 weeks	35	12.5
All assessments	159	56.9
Baseline and 11 weeks	4	1.4
Baseline, 21 weeks	6	2.1
Baseline, 6 and 21 weeks	8	2.8
Baseline, 11 and 21 weeks	6	1.0
No baseline, 6 weeks only	1	0.3
No baseline; 6, 11, and 21 weeks	3	1.0

(C) Reasons for missing forms by assessment time

Reason for missing data	6 weeks n (%)	11 weeks n (%)	21 weeks n (%)
Death	7 (14)	17 (24)	45 (46)
Illness-related	2 (4)	5 (7)	5 (5)
Patient refused, not ill	3 (6)	3 (4)	6 (6)
Patient refused phone interview	0 (0)	1 (1)	3 (3)
Not contacted	2 (4)	3 (4)	2 (2)
Institution error	12 (24)	13 (18)	10 (10)
Off treatment	1 (2)	2 (3)	1 (1)
Other reason[a]	6 (12)	7 (10)	11 (11)
No reason provided	16 (33)	21 (29)	14 (14)
Total	49	72	97

[a] Other reasons at 6 weeks ($n = 6$): death or deteriorating health ($n = 2$); wife took home but patient did not complete ($n = 1$); other ($n = 3$). Other reasons at 11 weeks ($n = 7$): deteriorating health ($n = 1$); off treatment or study ($n = 3$); out of the country ($n = 1$); patient refusal ($n = 1$); unknown ($n = 1$). Other reasons at 21 weeks ($n = 11$): death or deteriorating health ($n = 3$); off study ($n = 2$); out of the country ($n = 1$); patient refusal ($n = 1$); patient told QOL form not needed ($n = 1$); unknown ($n = 3$).

with submitted questionnaires and if a question-naire was not completed by the patient) for missing QOL questionnaires. Institution staff did not always use the categories for missing question-naires as indicated by the examples of reasons in the 'other-specify' category. If 'other reasons' having to do with deteriorating health are added to the two specific categories of death and deteriorating health in Table 3(C), we have the following percentages for non-ignorable missing data at each time point: 6 weeks (11/49 = 22%); 11 weeks

(23/72 = 32%); 21 weeks (19/47 = 55%). The percentage of missing forms due to death or deteriorating health would be even higher if we added the 'off-treatment/off study' reasons to the deteriorating category since patients are often taken off treatment if they are not responding or are doing poorly. However, one still notes approximately 24%, 18%, and 10% of the questionnaires missing due to institution error at 6, 11, and 21 weeks, respectively. Figure 2 represents the missing data problem graphically; the 67 patients with intermit-
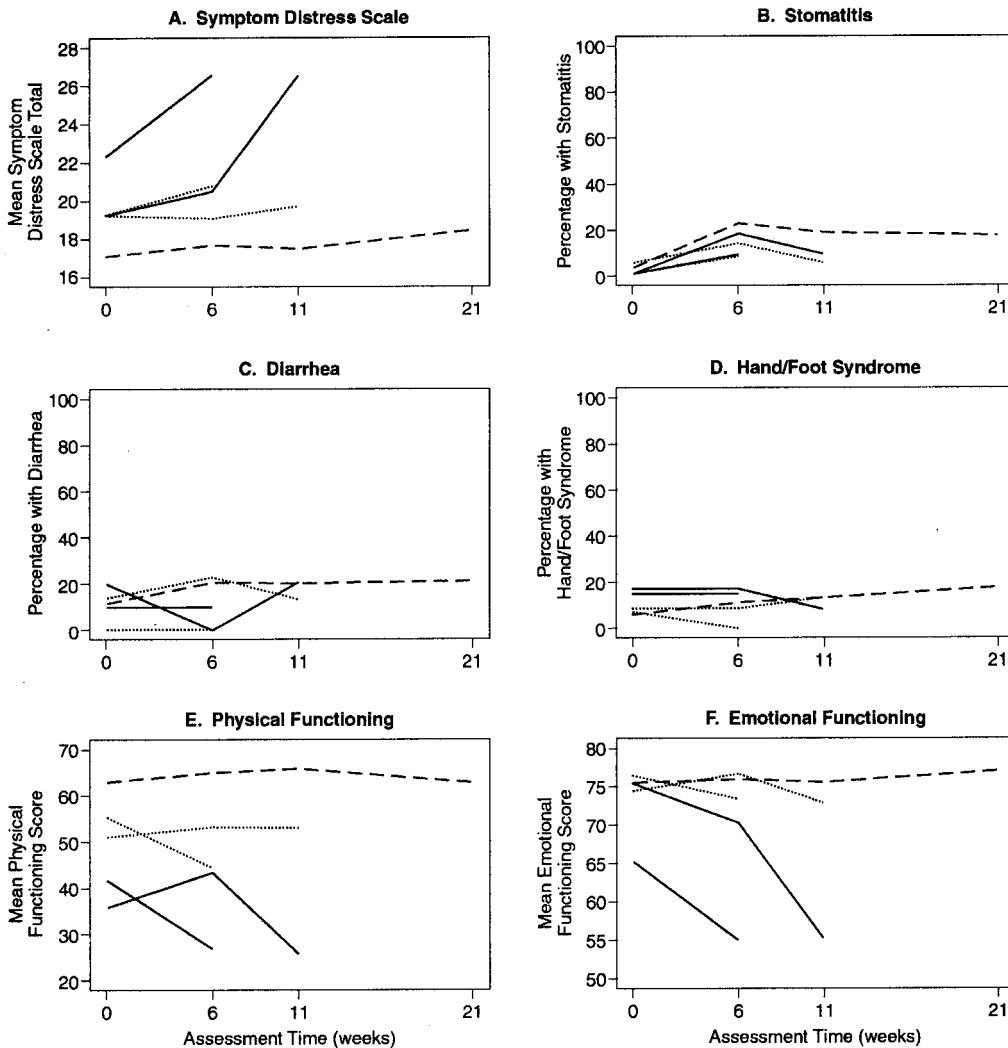

Figure 2. QOL by type and length of follow-up. Six endpoints: mean symptom status total score (2A); % with bothersome stomatitis (2B); % with bothersome diarrhea (2C); % with bothersome hand/foot syndrome (2D); mean physical functioning score (2E); mean emotional functioning score (2). Patients with incomplete data are grouped by last questionnaire submitted: – – – (complete QOL data at four time points); — (incomplete data because of death or deteriorating health); - - - (incomplete data unrelated to health status).

tent patterns of missing data are not included in Figure 2. Figure 2(A) portrays the effect of missing data on potential conclusions about symptom status during treatment. The following can only be descriptive in nature, as the informative dropout problem reflected in Figure 2 makes statistical comparisons inappropriate. Figure 2(A) indicates that patients who submitted all assessments had lower (better) Symptom Distress Scale scores at baseline and throughout the follow-up period. The average Symptom Distress Scale score for patients who dropped out of the QOL database due to death or deteriorating health status was worse at baseline than that of patients with complete assessments. This finding is consistent with the ability of the baseline Symptom Distress Scale scores to predict survival as presented below. Figure 2(A) further indicates that on average the Symptom Distress Scale score was worst (highest) at the last assessments taken before these patients dropped out. Average values for patients who had missing data for other reasons (e.g. institution error) looked more like those for patients who submitted complete QOL data. Conclusions about the effect of treatment on patient QOL over time are subject to bias by the tendency of patients with the poorest QOL to drop out. The effect of data missing because of death or deteriorating health was similar for physical and emotional functioning (Figure 2(E) and (F)) but did not appear to be as problematic for the three treatment-related symptom endpoints (Figure 2(B), (C), and (D)).

Treatment-specific symptoms

Of the hypothesized differences for treatment-specific symptoms, chi-square tests were significant only for hand/foot sensitivity at 11 ($p = 0.003$) and 21 ($p < 0.001$) weeks.

Stomatitis. Significant differences in stomatitis (percent with bothersome or worse symptoms) were not detected for patients with a regimen including leucovorin versus no leucovorin at any of the follow-up assessments, although differences were in the hypothesized direction (i.e. more stomatitis with leucovorin) (see Tables 1 and 4(A)). Figure 2(B) shows no evidence of bias due to incomplete follow-up.

Diarrhea. Significant differences in amount of diarrhea reported by patients receiving treatment by IV push were not detected at any of the follow-up assessments, although differences were again in the hypothesized direction (more diarrhea with IV push) (see Tables 1 and 4(B)). As with stomatitis, Figure 2(C) shows no evidence of bias due to incomplete follow-up.

Hand/Foot sensitivity. As hypothesized, patients whose treatment was administered via protracted continuous infusion reported significantly more hand and/or foot sensitivity at 21 weeks than did patients whose treatment was administered as a bolus dose (chi-square tests, $p < 0.001$); the difference was also significant at 11 weeks ($p = 0.003$) (see Tables 1 and 4(C)). Figure 2(D) shows a slightly increasing percentage of patients reporting troublesome levels of hand/foot syndrome over time for those patients with complete QOL data, and shows little evidence of bias due to incomplete data.

Physical and emotional functioning

Hypothesized differences for physical and emotional functioning were not observed as tested by Wilcoxon rank sum tests. Figure 2(E) and (F) indicate poorer physical and emotional functioning for patients who drop out of the QOL study for health-related reasons relative to those with complete data. This bias prevented the conduct of longitudinal analyses for the emotional and physical functioning variables.

Physical functioning. Patients who received treatment via IV push or bolus dose did not report poorer physical functioning (lower SF-36 Physical Functioning Scale scores) than did patients receiving treatment via infusion (see Tables 1 and 5(A)). Figure 2(E) indicates poorer physical functioning for patients who drop out of the QOL study for health-related reasons relative to those with complete data.

Emotional functioning. Patients administered treatment via protracted continuous infusion reported similar emotional distress (Mental Health Index scores) to patients without a pump (Table 5(B)). It was hypothesized that after an initial adjustment period, patients with a continuous infusion pump would demonstrate better emotional functioning than would those receiving treatment by bolus dosing or 24-h infusion. Although the Wilcoxon rank sum tests for Mental Health Index

Table 4. Treatment-specific symptoms

(A) Incidence of 'bothersome' or worse stomatitis

Assessment	Leucovorin Arms 2, 3, 5			No leucovorin Arms 1, 4, 6, 7		
	n	%	95% CI	*n*	%	95% CI
Baseline	114	2	(0–4)	155	3	(0–5)
6 weeks	95	21	(13–29)	134	17	(11–24)
11 weeks	85	19	(11–27)	122	12	(6–18)
21 weeks	76	18	(10–27)	104	12	(5–18)

(B) Incidence of frequent diarrhea

	IV push or bolus Arms 1–3			Infusion (24 h and protracted) Arms 4–7		
Baseline	93	15	(8–22)	126	12	(6–18)
6 weeks	75	24	(14–34)	105	13	(7–20)
11 weeks	71	20	(10–29)	94	20	(12–28)
21 weeks	70	24	(14–34)	79	15	(7–23)

(C) Incidence of 'bothersome' or worse hand/foot sensitivity

	Protracted continuous infusion (>24 h) Arms 4–5			No continuous infusion Arms 1–3		
Baseline	75	7	(1–12)	114	10	(4–15)
6 weeks	65	11	(3–18)	97	12	(6–19)
11 weeks	59	22	(11–33)	89	6	(1–10)
21 weeks	48	31	(18–44)	83	7	(2–13)

CI, 95% confidence interval for the proportion.

scores did not detect combined treatment arm differences at any of the time points, the difference was in the hypothesized direction at 21 weeks. Figure 2(F) suggests that patients with data missing due to death or deteriorating health show poorer emotional functioning.

Overall symptom status

Although there were no specific hypotheses with respect to the total Symptom Distress Scale score, descriptive data for this questionnaire are presented in Table 6 broken down by one agent and two route of administration treatment arm combinations (see Table 1). In general, there were minimal differences. The data suggest better symptom status at the last assessment for patients receiving drug via continuous infusion versus 24-h infusion or IV push route of administration. However, this suggestion must be tempered, given

the potential for missing data bias just demonstrated for symptom status data; in addition, *a priori* hypotheses regarding overall symptom status had not been identified.

DISCUSSION

Missing data problems in advanced-stage disease

This study showed that good quality control procedures can result in good compliance with QOL assessment schedules in cooperative group trials. Submission rates for live patients ranged from 99% at baseline to 78% at 21 weeks. These submission rates were achieved in the cooperative group context where many institutions contribute patients to trials, complicating quality control procedures. In general, the QOL questionnaires were administered on schedule, in the face of

Table 5. QOL Scales

(A) Physical functioning total score[a]: median, mean, and 95% CI for the mean

Assessment	IV push or bolus Arms 1–3				Infusion (24 h and protracted) Arms 4–7			
	n	Median	Mean	95% CI	*n*	Med	Mean	95% CI
Baseline	119	70	61	(55–66)	153	55	55	(50–60)
6 weeks	95	60	60	(54–67)	130	60	59	(54–65)
11 weeks	89	75	75	(56–70)	117	65	61	(55–67)
21 weeks	84	73	73	(56–71)	98	70	60	(53–67)

(B) Emotional functioning total score[a]: median, mean, and 95% CI for the mean

	IV push or 24 h infusion Arms 1–3, 6, 7				Protracted continuous infusion (>24 h) Arms 4–5			
Baseline	195	80	75	(72–78)	77	76	72	(66–77)
6 weeks	162	76	75	(72–78)	64	80	73	(67–80)
11 weeks	148	80	74	(70–78)	58	80	72	(66–78)
21 weeks	134	80	76	(72–80)	47	88	77	(69–84)

[a] Higher scores = better functioning.

demanding data management requirements associated with clinical trials research. Patients also demonstrated their willingness to complete questionnaires. However, missing data due to death and deteriorating health can make it impossible to conduct of longitudinal analyses due to reductions in sample size and non-ignorable missing data biases. With advanced-stage disease, even the best quality control procedures cannot avoid missing data that are associated with deteriorating patient health and patient death.

Our experience in this trial suggests that it is risky to draw firm conclusions about the effect of treatment on patient functioning over time, given the substantial number of missing follow-up assessments. Patients who died or who had deteriorating health status were increasingly not available for analyses over the 21-week QOL assessment period. Using only QOL data for patients who survive and are well enough to complete all questionnaires clearly leads to an overestimate of QOL status and is only informative for patients who survive. When QOL data are missing related to an aspect of the QOL endpoint (i.e. the health status of the patient), the resulting bias is problematic: it compromises our ability to describe patterns of QOL change, to draw conclusions about the maintenance of QOL during the

course of treatment, and to document the presence or absence of palliation.

QOL and patient benefit

In this study, we were unable to determine the extent to which the treatment arms differed in their ability to achieve palliation or confer patient benefit (Moinpour *et al.*, 1999). Substantial problems with missing data due to death and deteriorating health complicated our ability to evaluate patient benefit by treatment arm. In addition, confidence intervals indicated a wide range of potential values for the reported measures. For example, the confidence interval for the continuous infusion arms at 21 weeks indicates that the proportion of patients reporting bothersome or worse hand/foot sensitivity could be as high as 44%.

Hill *et al.* (1995), using the comprehensive EORTC QC-30 measure, did not find QOL differences by treatment arm in the advanced-stage disease setting. The Hill *et al.* trial compared protracted infusion of 5-FU alone or with a biochemical modulator (IFN). Although the biochemical modulator in the Hill study differed from that used in this study, the continuous infusion 5-FU alone arm was the same. It may be that

Table 6. Symptom Distress Scale total scores (median, mean, 95% CI for the mean[a])

(A) Arms containing leucovorin versus no leucovorin

Assessment	Leucovorin Arms 2, 3, 5				No leucovorin Arms 1, 4, 6, 7			
	n	Median	Mean	95% CI	*n*	Median	Mean	95% CI
Baseline	114	17	18	(16–20)	158	17	19	(17–20)
6 weeks	94	17	18	(16–20)	131	18	19	(17–20)
11 weeks	84	15	17	(16–19)	122	17	19	(17–21)
21 weeks	76	16	18	(16–20)	101	18	19	(17–21)

(B) IV push versus infusion routes of administration

	IV push or bolus Arms 1–3				Infusion (24 h and protracted) Arms 4–7			
Baseline	116	17	18	(16–19)	156	17	19	(17–20)
6 weeks	95	17	18	(17–20)	130	18	19	(17–20)
11 weeks	90	16	18	(16–20)	116	16	19	(17–20)
21 weeks	83	18	18	(17–20)	94	17	19	(17–20)

(C) Protracted infusion versus other forms of administration (24 h infusion or IV push)

	Protracted continuous infusion (>24 h) Arms 4–5				IV push or 24 h infusion Arms 1–3, 6, 7			
Baseline	78	17	19	(17–21)	194	17	18	(17–19)
6 weeks	64	17	19	(17–21)	161	17	18	(17–20)
11 weeks	58	16	18	(16–20)	148	17	19	(17–20)
21 weeks	46	15	18	(15–20)	131	18	19	(17–20)

[a] Higher scores reflect worse symptom status.

patients with advanced-stage disease, even those with good performance status, who respond modestly to treatment, are not likely to show significant improvements in broader aspects of day to day functioning such as physical and emotional functioning. That is, the modest clinical outcomes may not translate to significant improvement in ability to carry out normal activities and in emotional outlook. In addition, modest clinical improvements may also be associated with substantial levels of toxicity. Saltz *et al.* (1999) reported confirmed response rates of 33% for a combined regimen of 5-FU, leucovorin, and CPT-11 compared to confirmed response rates of <20% for 5-FU/leucovorin and for CPT-11 alone. Although QOL was not measured in this study, the authors reported non-trivial problems with grade 3 or 4 diarrhea in the arms containing CPT-11 and grade 4 neutropenia in the arms with 5-FU and leucovorin. It may be that in the ad-

vanced-stage disease setting, stability in the patient's report of both symptom status and general QOL dimensions is the best we can expect. However, QOL data have the potential to help us identify patient benefit (or harm) attributable to treatment regimens when response rates and survival are minimally affected.

Attempts to compare the toxicity profile reported for the therapeutic trial (Leichman *et al.*, 1995) with that available for the companion QOL study are problematic given that our analyses were based on combinations of arms rather than individual arms as reported in the therapeutic paper. For example, we combined all arms with the modulator, leucovorin, and compared symptom reports to those provided by patients not treated with leucovorin. The direction of the effect was as hypothesized but not significantly different from zero. The report of the therapeutic trial showed significantly more severe stomatitis in

arms 2 and 5, two of the three arms in our combined comparison. It also reported the highest proportion of severe diarrhea in arm 3 (IV bolus administration of 5-FU plus high dose leucovorin). We reported more diarrhea (not significantly so) in arms 1–3 versus arms 4–7. Arms 1 and 2 differed from arm 3 in that arms 1 and 2 administered 5-FU over 5 days versus a 1-day administration in arm 3; the single day administration would be expected to result in more diarrhea.

The Meta-analysis Group in Cancer evaluated administration of 5-FU by bolus versus continuous infusion (Meta-analysis Group in Cancer, 1998); arm 7 and the two arms with high dose leucovorin included in the Southwest Oncology Group trial were omitted. Although the authors of the meta-analysis reported, as we did, significantly more hand/foot sensitivity for continuous infusion compared to other treatment administration routes, its conclusion was that continuous infusion administration of 5-FU was superior to bolus administration for a number of clinical outcomes. A follow-up SWOG trial (SWOG-9420) used continuous infusion of two doses of 5-FU (no QOL data were collected). In a *post hoc* examination of QOL data for SWOG 8905/9045, we did not observe superiority in QOL when 5-FU alone was administered by continuous infusion compared to the arms where 5-FU was modulated by leucovorin (bolus or infusion).

Patients with complete data had relatively stable QOL (Symptom Distress Scale, Physical Functioning, and Mental Health Index scores) over the QOL study period, unlike those who dropped out of the study as they deteriorated or died. However, information based only on patients providing a complete schedule of assessments is not useful for new patients at the initiation of treatment. As noted above, the inappropriate use of complete case analysis approaches would suggest a palliative effect of treatment. Patients who are not doing well at the last time they contributed data are not represented in such analyses. Their reported QOL would certainly temper conclusions regarding palliation. From a common sense perspective, we would conclude that palliation did not occur for the large number of patients whose condition deteriorated while on treatment. The physician-rated improvements reported by Poon et al. (1989) for patients receiving 5-FU and low-dose leucovorin documented palliation but this research involved a less comprehensive measure of

QOL and one that did not include patient self-administered questionnaires.

Finally, we must remember that the opposite of palliation can occur in treatment trials. It is possible that treatment can worsen patient functioning; for example severe treatment side-effects can negatively affect patient emotional or physical functioning. An example from advanced-stage disease in prostate cancer documented negative impacts on emotional functioning at months 1, 3, and 6 months post-randomization for patients receiving total androgen ablation (orchiectomy plus flutamide) versus treatment with orchiectomy alone (plus placebo) (Moinpour et al., 1998). Hill et al. (1995) noted that transient symptom problems during the trial did not affect broader areas of patient functioning, providing some evidence that the treatments did achieve palliation (i.e. overall patient QOL and functioning were maintained during the treatment period). Symptom Distress Scale scores did not go above 18 for any of the combination of arms over the 21-week assessment period in this trial. Symptom Distress Scale total scores above 24–26 points have been found by the developer of the questionnaire to reflect moderate symptom problems (McCorkle et al., 1998). This would suggest that, although we did not observe differences in symptom palliation by treatment arm, most of the patients treated with these approaches appeared to be maintaining a fairly good level of symptom status.

We believe that until there is a more complete QOL database for patients with advanced-stage disease, it is important to continue to include comprehensive measures of QOL in clinical trials addressing QOL issues. Without a comprehensive measure, we cannot properly evaluate the *extent to which* palliation objectives are achieved or the full extent of negative effects of treatment such as those found in the advanced-stage prostate cancer trial described above (Moinpour et al., 1998).

CONCLUSIONS

The primary purpose of the QOL companion study was to monitor the effects on patient QOL of agent and route of administration variations in an advanced colorectal cancer treatment trial. Although we comprehensively measured the effect of treatment on QOL, we could not complete the planned analyses due to the presence of non-

ignorable missing data. Therefore, the primary purpose of this paper became the interpretation of QOL results in the context of missing longitudinal data. Quality control procedures were in place to ensure adequate submission rates for patients who were alive and not experiencing deteriorating health status. However, institution error accounted for 10–24% of missing questionnaires over the follow-up period, indicating that quality control procedures can always be improved. Although we observed adequate submission rates for living patients, these rates did not make up for the decreasing sample size over time and the bias associated with attrition due to death. For example, by the end of the QOL assessment schedule (21 weeks), 46% of the reasons for missing questionnaires were associated with death and deteriorating health. Therefore, longitudinal analyses of the QOL data could not be justified. The challenges associated with QOL assessment in the advanced-stage disease setting have been noted (Bernhard *et al.*, 1998). For this reason, many investigators 'front-load' QOL assessments in order to capture patient functioning before deteriorating health prevents submission of questionnaires. The disadvantage to this approach is the failure to capture longer-term positive effects of treatment on patient QOL as well as the potential for later negative effects. However, in advanced-stage disease, we are trying to document support for new therapies where we do not expect to see substantial gains in survival. The ability to differentiate treatment arms with respect to symptom status and general areas of functioning can be important, although some justification of the clinical meaningfulness of differences is required (Lydick and Epstein, 1993; Osoba *et al.*, 1998; Sloan *et al.*, 1998). In this context, a focus on QOL assessment during treatment is probably defensible and may result in the identification of outcomes other than survival that demonstrate patient benefit.

ACKNOWLEDGEMENTS

REFERENCES

Aaronson NK, Bullinger M, Ahmedzai S. 1988. A modular approach to quality of life assessment in cancer clinical trials. *Recent Results Cancer Res* **111**: 231–249.

Aaronson NK, Ahmedzai S, Bergman B *et al.* 1993. 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* **85**: 365–373.

Bengt G, Hoffman K, Graf W, Påhman L, Sjödén P-O for the Nordic Gastrointentinal Tumor Adjuvant Therapy Group. 1994. Quality of life during chemotherapy in patients with symptomatic advanced colorectal cancer. *Cancer* **73**: 556–562.

Bernhard J, Cella DF, Coates AS, Fallowfield L, Ganz PA, Moinpour CM, Mosconi P, Osoba D, Simes J, Hurny C. 1998. Missing quality of life data in cancer clinical trials: Serious problems and challenges. *Stat Med* **17**: 517–532.

Bertsch L A, Donaldson G. 1995. Quality of life analyses from vinorelbine (navelbine) clinical trials of women with metastatic breast cancer. *Semin Oncol* **22**: 45–54.

Gotay CC, Korn EL, McCabe M, Moore TD, Cheson BD. 1992. Quality of life assessment in cancer treatment protocols: research in issues in protocol development. *J Natl Cancer Inst* **84**: 575–579.

Hill M, Norman A, Cunningham D, Findlay M, Watson M, Nicolson V, Webb A, Middleton G, Ahmed F, Hickish T, Nicolson M, O'Brien M, Iveson T, Evans C. 1995. Impact of protracted venous infusion fluorouracil with or without interferon Alfa-2b on tumor response, survival, and quality of life in advanced colorectal cancer. *J Clin Oncol* **13**: 2317–2323.

Hopwood P, Stephens R, Machin D for the MRC Lung Cancer Working Party. 1994. Approaches to the analysis of quality of life data: experiences gained form a Medical Research Council Lung Cancer Working Party palliative chemotherapy trial. *Qual Life Res* **3**: 339–352.

Leichman CG, Leichman L, Spears P, Rosen PJ, Muggia FJS. 1990. Biological modulation of protracted infusion of 5-fluorouracil with weekly leucovorin. A dose-seeking clinical trial for patients with disseminated gastrointestinal cancers. *Cancer Chemother Pharmacol* **26**: 57–61.

Leichman CG, Fleming TR, Muggia FM, Tangen CM, Ardalan B, Doroshow JH, Meyers FJ, Holcombe RF, Weiss GR, Mangalik A, MacDonald JS. 1995. Phase II study of fluorouracil and its modulation in advanced colorectarl cancer: a Southwest Oncology Group study. *J Clin Oncol* **13**: 287–290.

Lydick E, Epstein RS. 1993. Interpretation of quality of life changes. Commentary. *Qual Life Res* **2**: 221–226.

McCorkle R. 1987. The measurement of symptom distress. *Semin Oncol Nurs* **3**: 248–256.

McCorkle R, Benoliel JQ. 1983. Symptom distress, current concerns, and mood disturbance after diagnosis of life threatening disease. *Soc Sci Med* **17**: 431–438.

McCorkle R, Young K. 1978. Development of a symptom distress scale. *Cancer Nurs* **1**: 373–378.

McCorkle R, Benoliel JQ, Donaldson G, Fortini G, Moinpour CM, Goodell B. 1989. A randomized clinical trial of home nursing care for lung cancer patients. *Cancer* **64**: 1375–1382.

McCorkle R, Cooley ME, Shea JA. 1998. *A User's Manual for the Symptom Distress Scale*. University of Pennsylvania School of Nursing: Philadelphia, PA.

McHorney C, Ware JE Jr, Raczek A. 1993. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* **31**: 247–263.

McHorney CA, Ware JE Jr, Lu RJF, Sherbourne CD. 1994. The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions and reliability across diverse patient groups. *Med Care* **32**: 40–46.

Meta-analysis Group in Cancer. 1998. Efficacy of intravenous continuous infusion of fluorouracil compared with bolus administration in advanced colorectal cancer. *J Clin Oncol* **16**: 301–308.

Moinpour CM. 1994. Measuring quality of life: an emerging science. *Semin Oncol* **21**: 48–63.

Moinpour CM, Lovato LC. 1998. Ensuring the quality of quality of life data: the Southwest Oncology Group experience. *Stat Med* **17**: 641–651.

Moinpour CM, Feigl P, Metch B, Hayden KA, Meyskens FL Jr, Crowley J. 1989. Quality of life end points in cancer clinical trials: review and recommendations. *J Natl Cancer Inst* **81**: 485–495.

Moinpour CM, Savage Marguerite J, Troxel A, Lovato LC, Eisenberger M, Veith RW, Higgins B, Skeel R, Yee M, Blumenstein BA, Crawford ED, Meyskens FL Jr. 1998. Quality of life in advanced prostate cancer: results of a randomized therapeutic trial. *J Natl Cancer Inst* **90**: 1537–1544.

Moinpour CM, Ganz PA, Rowland JH, Gritz ER, Gotay CC. 1999. Guest editorial: the value of quality of life data in judging patient benefit. *Cancer Lett* **25**: 1–5.

Nayfield SG, Ganz PA, Moinpour CM, Cella DF, Hailey BJ. 1992. Report from a National Cancer Institute (USA) workshop on quality of life assessment in cancer clinical trials. *Qual Life Res* **1**: 203–210.

Nordic Gastrointestinal Tumor Adjuvant Therapy Group. 1989. Superiority of sequential methotrexate, fluorouracil, and leucovorin to fluorouracil alone in advanced symptomatic colorectal carcinoma: a randomized trial. *J Clin Oncol* **7**: 1437–1446.

Osoba D. 1994. Lessons learned from measuring health-related quality of life in oncology. *J Clin Oncol* **12**: 608–616.

Osoba D, Rodrigues G, Myles J, Zee B, Pater J. 1998. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* **16**: 139–144.

Petrelli N, Douglass HO Jr, Herrera L, Russell D, Stablein DM, Bruckner HW, Mayer RJ, Schinella R, Green MD, Muggia FM, Megibow A, Greenwald ES, Bukowski RM, Harris J, Levin B, Gaynor E, Loutfi A, Kalser MH, Barkin JS, Benedetto P, Woolley PV, Nauta R, Weaver DW, Leichman LP for the Gastrointestinal Tumor Study Group. 1989. The modulation of fluorouracil with leucovorin in metastatic colorectal carcinoma: a prospective randomized phase III trial. *J Clin Oncol* **7**: 1419–1426.

Poon MA, O'Connell J, Moertel G, Wieand HS, Cullinan SA, Everson LK, Krook JE, Malliard JA, Laurie JA, Tschetter LK, Wieserfeld M. 1989. Biochemical modulation of fluorouracil: evidence of significant improvement of survival and quality of life in patients with advanced colorectal carcinoma. *J Clin Oncol* **7**: 1407–1418.

Saltz LB, Locker PK, Pirotta N, Elfring GL, Miller LL. 1999. Weekly irinotecan (CPT-11), leucovorin (LV), and fluorouracil (FU) is superior to daily X5 LV/FU in patients (PTS) with previously untreated metastatic colorectal cancer (CRC). *Proc ASCO* **18**: 233a.

Sloan JA, Loprinzi CL, Kuross SA, Miser AW, O'Fallon JR, Mahoney MR, Heid IM, Bretscher ME, Vaught NL. 1998. Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. *J Clin Oncol* **16**: 3662–3673.

Staquet M, Berzon R, Osoba D, Machin D. 1996. Guidelines for reporting results of quality of life assessments in clinical trials. *Qual Life Res* **5**: 496–502.

Staquet MJ, Hays RD, Fayers PM (Eds). 1998. *Quality of Life Assessment in Clinical Trials. Methods and Practice*. Oxford University Press: Oxford.

Stewart AL, Hays RD, Ware JE Jr. 1988. The MOS Short-form General Health Survey. Reliability and validity in a patient population. *Med Care* **26**: 724–735.

Stewart AL, Greenfield S, Hays RD, Wells K, Rogers WH, Berry S, McGlynn EA, Ware JE Jr. 1989. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. *J Am Med Assoc* **262**: 907–913.

Troxel AB, Fairclough DL, Curran D, Hahn EA. 1998. Statistical analysis of quality of life with missing data in cancer clinical trials. *Stat Med* **17**: 653–666.

Ware JE Jr, Sherbourne CD. 1992. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* **30**: 473–483.

Ware JE, Jr, Brook R, Davies-Avery A. 1980. Conceptualization and measurement of health for adults in the Health Insurance Study. Model of Health and Methodology, Vol. 11. R-1987/1-HEW.

Wingo PA, Ries LAG, Rosenberg HM, Miller DS, Edwards BK. 1998. Cancer incidence and mortality, 1973–1995. *Cancer* **82**: 1197–1207.

Young KJ, Longman AJ. 1983. Quality of life and persons with melanoma: a pilot study. *Cancer Nurs* **6**: 219–225.