Original Article

Combining Anchor and Distribution-Based Methods to Derive Minimal Clinically Important Differences on the Functional Assessment of Cancer Therapy (FACT) Anemia and Fatigue Scales

David Cella, PhD, David T. Eton, PhD, Jin-Shei Lai, PhD,
Amy H. Peterman, PhD, and Douglas E. Merkel, MD
Evanston Northwestern Healthcare and Northwestern University, Evanston, Illinois, USA

Abstract
Magnitude differences in scores on a measure of quality of life that correspond to differences in function or clinical course are called clinically important differences (CIDs). Anchor-based and distribution-based methods were used to provide ranges of CIDs for five targeted scale scores of the Functional Assessment of Cancer Therapy—Anemia (FACT-An) questionnaire. Three samples of cancer patients were used: Sample 1 included 50 patients participating in a validation study of the FACT-An; Sample 2 included 131 patients participating in a longitudinal study of chemotherapy-induced fatigue; sample 3 included 2,402 patients enrolled in a community-based clinical trial evaluating the effectiveness and safety of a treatment for anemia. Three clinical indicators (hemoglobin level; performance status; response to treatment) were used to determine anchor-based differences. One-half of the standard deviation and 1 standard error of measurement were used as distribution-based criteria. Analyses supported the following whole number estimates of a minimal CID for these five targeted scores: Fatigue Scale = 3.0; FACT-G total score = 4.0; FACT-An total score = 7.0; Trial Outcome Index—Fatigue = 5.0; and Trial Outcome Index—Anemia = 6.0. These estimates provide a basis for sample size estimation when planning for a clinical trial or other longitudinal study, when the purpose is to ensure detection of meaningful change over time. They can also be used in conjunction with more traditional clinical markers to assist investigators in determining treatment efficacy. J Pain Symptom Manage 2002;24:547–561. © *U.S. Cancer Pain Relief Committee, 2002.*

Key Words
Anemia, fatigue, quality of life, clinical significance

Address reprint requests to: David Cella, PhD, 1001 University Place, Suite 100, Evanston, IL 60201, USA.
Accepted for publication: January 19, 2002.

0885-3924/02/$–see front matter
PII S0885-3924(02)00529-8

Introduction

Fatigue and anemia are among the most common unrelieved problems of cancer.[1–6] Fatigue is a subjective sensation of weakness, lack of energy, or tiredness.[7] Anemia is a multisymptom condition that includes fatigue as its cardinal symptom, dizziness, headache, chest pain, shortness of breath, and decreased motivation.[8] Anemia is quite common in cancer patients and may result from either the disease or the myelosuppressive effects of treatment. Chemotherapy, in particular, is often associated with declines in blood hemoglobin. In many cases, these declines are so severe as to warrant adjuvant therapy via blood transfusion or recombinant human erythropoietin (rHuEPO).[9]

Fatigue and anemia are especially important from the standpoint of quality of life (QOL) as they both can profoundly impact a patient's physical and mental capacity.[1,2,10–12] The debilitating nature of fatigue and other symptoms of anemia makes their assessment a critical component of outcome evaluation in clinical trials, particularly those where QOL can be adversely affected by these symptoms.

Fatigue and anemia are measured in two ways: 1) physiological/medical indicators and 2) self-report questionnaires. Physiological approaches to the measurement of fatigue typically include hemoglobin/hematocrit level and occasionally a test of physical capacity (i.e., muscle strength; exercise tolerance). Several self-report questionnaires that rely on patient perceptions of fatigue and symptoms of anemia have recently been developed. These include fatigue and anemia subscales of the Functional Assessment of Cancer Therapy (FACT, also known as FACIT: Functional Assessment of Chronic Illness Therapy),[1,13] the Brief Fatigue Inventory (BFI),[14] the Piper Fatigue Scale (PFS),[15] the Multi-dimensional Fatigue Inventory (MFI),[12] the Fatigue Symptom Inventory,[16] and various linear analog scales.[17] The FACT-Fatigue and -Anemia scales, in particular, have shown strong associations with hemoglobin level, functional status, and global QOL.[9,13,18,19]

As interest in the prevalence and treatment of fatigue and anemia grows, clinical investigators are becoming increasingly interested in the direct measurement of these symptoms, as well as other, related, patient-reported outcomes, such as QOL. This has coincided with initiatives by regulatory agencies such as the U.S. Food and Drug Administration to include QOL as a primary endpoint in cancer clinical trials.[20] Several cancer drug submissions have included QOL data in the package, and in some circumstances these data have been influential in approval.[21,22]

While the assessment of QOL in many clinical trials is supported, aggregated scores on QOL self-reports are often less resonant to the clinician than functional parameters such as performance status and response to treatment. This can result in under-appreciation of the information learned from the QOL evaluation. One way of enhancing the clinical utility of scores on multi-item questionnaires is by investigating the importance (to patients and clinicians) of cross-sectional differences and longitudinal change scores by anchoring those differences and changes to clinically familiar events that are related to patient well-being. This area of investigation has come to be organized under the general heading of "clinical significance," as opposed to the more traditional statistical significance. The emerging importance of the area is demonstrated by the convening of over 30 experts on the topic over a two-year period, resulting in the production of six special topic papers and a summary.[23,24]

Jaeschke, Singer, and Guyatt have defined the minimal clinically important difference in a QOL score as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management."[25] A clinically important difference on a questionnaire should ideally be meaningful to both patients and providers. Some approaches have emphasized subjective importance to the patient only.[25,26]

Clinically important differences can be determined via anchor-based and distribution-based methods.[27] *Anchor-based* methods 'anchor' or map a difference score onto differences in clinical results. Anchor-based differences can be determined either cross-sectionally (differences between clinically-defined groups at one time point) or longitudinally (change in score of one group over time). *Distribution-based* methods rely on the statistical distributions of QOL scores in a given study.[27] These may include reliance on the standard deviation and/or standard error of measurement.[28] These approaches

have been used successfully to determine clinically important differences in QOL scores.[25–29]

The objective of this analysis was to establish clinically important score differences on the FACT-Fatigue and FACT-Anemia scales. In this article, the term "clinically important" is intended to be synonymous with "clinically meaningful" and "clinically significant," with the understanding that the more relevant issue is whether anchors are known to be important to patients, to clinicians, or to both. We did this by anchoring cross-sectional differences and longitudinal change scores to clinical indicators such as hemoglobin level, performance status and response to treatment. To augment the anchor-based evidence, we also used selected distribution-based measures to further specify clinically important differences.

Methods

Description of Samples

To provide for more accurate estimates of CIDs on the targeted scale scores by triangulating the results of several trials, we used three samples in the present study. Sample 1 consisted of 50 mixed diagnosis cancer patients currently receiving treatment. Quality of life was assessed twice: at baseline and 3 to 7 days later to evaluate test stability (test–retest reliability). Sample 2 consisted of 131 mixed-diagnosis cancer patients participating in a longitudinal observational study of fatigue and quality of life during chemotherapy. Quality of life was assessed at baseline (before starting treatment), 3 months, 6 months, and 1 year. Sample 3 consisted of 2,402 mixed-diagnosis cancer patients enrolled in an open-label, non-randomized, community-based clinical trial evaluating the effectiveness, safety, and clinical outcome of a treatment for anemia in cancer patients.[9] Quality of life was assessed at baseline and multiple follow-up points. Because maximal clinical data are available at baseline and study completion, we present data on only these two time points. All patients were receiving some form of chemotherapy at baseline.

Quality of Life Measure

The 47-item Functional Assessment of Cancer Therapy—Anemia (FACT-An)[1] was used in samples 1 and 3 (see Appendix). The FACT-An consists of 5 subscales; physical well being (PWB; 7 items), social/family well being (SWB; 7 items), emotional well being (EWB; 6 items), functional well being (FWB; 7 items), and anemia symptoms (AS; 20 items). The PWB, SWB, EWB, and FWB subscales can be summed to form the FACT—General (FACT-G) score. The PWB, FWB, and AS subscales can be summed to form the Trial Outcome Index—Anemia (TOI-An).

The 40-item Functional Assessment of Cancer Therapy—Fatigue (FACT-F)[13] was used in sample 2. The FACT-F consists of the PWB, SWB, EWB, and FWB subscales and the 13-item Fatigue scale (FS). The FS consists of those items from the AS that measure fatigue alone. The PWB, FWB, and FS subscales can be summed to form the Trial Outcome Index—Fatigue (TOI-F).

Anticipating probable endpoints in clinical trials to treat anemia and fatigue, we were interested in the following 5 scores: the FACT-An (47 items), FACT-G (27 items), Fatigue subscale (13 items), TOI-F (27 items), and TOI-An (34 items). All FACT questions are rated on a 5-point Likert-type scale ranging from 0 = "not at all" to 4 = "very much." Thus, possible score ranges were the following: FACT-An (0–188), FACT-G (0–108), Fatigue Scale (0–52), TOI-F (0–108), and TOI-An (0–136). Higher scores always represent better QOL or less severe symptoms.[30]

Medical and Demographic Data

Baseline demographic (e.g., age, sex, ethnicity) and medical information (e.g., disease site, performance status rating) were obtained from patient self-reports and physician charts. A patient-reported performance status rating (PSR) was obtained in samples 1 and 2 using Eastern Cooperative Oncology Group (ECOG) criteria.[31] ECOG PSR ratings range from 0 indicating that a patient is completely asymptomatic and fully ambulatory to 4 indicating that a patient is bedridden. In sample 2, treating oncologists also provided ratings of ECOG PSR. In sample 3, physicians rated patient performance status using the Karnofsky Performance Status Rating (KPSR).[32] Comparable to ECOG PSR; the KPSR is an 11-level decile rating ranging from 0 = "dead" to 100 = "normal, no complaints". Performance status was assessed

Table 1
Sample Descriptives

	Sample 1 (n = 50)		Sample 2 (n = 131)		Sample 3 (n = 2,402)	
Sex	Female	27 (54%)	Female	92 (70%)	Female	1426 (59%)
	Male	23 (46%)	Male	39 (30%)	Male	976 (41%)
Age	Median	52.5 years	Median	56.0 years	Median	65.2 years
	Range	19–83	Range	20–82	Range	17–94
Ethnicity	Caucasian	43 (86%)	Caucasian	114 (87%)	Caucasian	1835 (76%)
	African-Am.	4 (8%)	African-Am.	12 (9%)	African-Am.	274 (11%)
	Hispanic	3 (6%)	Asian	3 (2%)	Hispanic	124 (5%)
			Hispanic	2 (2%)	Asian	95 (4%)
					Other	17 (1%)
					Missing	57 (2%)
Disease site	Breast	12 (24%)	Breast	58 (44%)	Non-hematologic[b]	
	Colorectal	11 (22%)	Colorectal	14 (11%)	Lung	559 (31%)
	Lung	8 (16%)	Lung		Breast	385 (21%)
	Lymphoma	5 (10%)	(non-small cell)	12 (9%)	Gynecologic	303 (17%)
	Leukemia	3 (6%)	Ovarian	7 (5%)	Gastro-intestinal	209 (11%)
	Ovarian	3 (6%)	Non-Hodgkins	6 (5%)	Other	365 (20%)
	Prostate	2 (4%)	Head & neck	5 (4%)		
	Myeloma	2 (4%)	Hodgkins	5 (4%)	Hematologic[c]	
	Others	2 (4%)	Lung		Non-Hodgkins	239 (46%)
	Stomach	1 (2%)	(small cell)	2 (2%)	Hodgkins	132 (25%)
	Bladder	1 (2%)	Pancreatic	3 (2%)	Leukemia	90 (17%)
			Others	19 (15%)	Mult. Myeloma	64 (12%)
Extent of disease	not available		Local	51 (39%)	Distant met.	1004 (55%)
			Regional	39 (28%)	Local or regional	757 (42%)
			Distant met.	27 (21%)	Not ascertained	60 (3%)
			N/A	13 (10%)	(Note: Non-hematologic only)	
Hemoglobin level	Median	12.2 g/dl	Median	12.6 g/dl	Median	9.4 g/dl
	Range	7–15.9 g/dl	Range	4.5–19.3 g/dl	Range	4.6–11.0 g/dl
	<10 g/dl	9 (18%)	<10 g/dl	14 (11%)	<8 g/dl	224 (9%)
	10–11.9 g/dl	13 (27%)	10–11.9 g/dl	31 (24%)	8–9.9 g/dl	1391 (58%)
	≥12 g/dl	27 (55%)	≥12 g/dl	86 (66%)	10–11 g/dl	674 (28%)
					Missing	113 (5%)
Performance status	Patient-rated ECOG PSR[a]		Patient-rated ECOG PSR		Physician-rated KPSR[d]	
					90–100	736 (31%)
	0	18 (36%)	0	79 (60%)	80	667 (28%)
	1	22 (44%)	1	36 (28%)	70	459 (19%)
	2,3	10 (20%)	2,3	16 (12%)	60	230 (10%)
					≤50	196 (8%)
			Physician-rated ECOG PSR		missing	114 (5%)
			0	77 (59%)		
			1	51 (39%)		
			2–4	3 (2%)		

[a]Eastern Cooperative Oncology Group performance status rating: 0 = 'fully ambulatory without physical symptoms', 1 = 'ambulatory with some symptoms', 2–4 = 'requiring bedrest during waking day'.
[b]Percentages are based on the total number of non-hematologic malignancies.
[c]Percentages are based on the total number of hematologic malignancies.
[d]Karnofsky Performance Status Rating (physician-rated). 90–100 = 'normal activity', 80 = 'normal activity with effort', 70 = 'unable to carry on normal activity', 60 = 'occasional assistance', ≤50 = 'disabled, requires considerable assistance'.

at baseline in sample 1 and at baseline and follow-up time points in samples 2 and 3. Among other key clinical indicators, baseline and follow-up hemoglobin data were available across the samples and best overall response to cancer treatment was available for sample 3. Across all samples, we chose 10g/dl as a cut-off point because it is the clinically recognized division between mild anemia (10–12 g/dl) and moderate anemia (8–10 g/dl).

Table 2
Baseline Scores and Internal Consistency of Targeted Endpoints

FACT Scale or Aggregate	Sample 1 ($n = 50$)[a]		Sample 2 ($n = 131$)[a]		Sample 3 ($n = 50$)[a]	
	Mean (SD)	Alpha	Mean (SD)	Alpha	Mean (SD)	Alpha
Fatigue Scale	36.8 (10.5)	0.93	38.7 (10.9)	0.95	23.9 (12.6)	0.94
FACT-G	79.4 (13.9)	0.85	83.1 (14.7)	0.86	71.9 (15.9)	0.88
FACT-Anemia	136.1 (26.0)	0.95	not administered	not administered	113.2 (29.7)	0.94
TOI-Fatigue	76.9 (20.0)	0.96	80.4 (20.3)	0.95	56.1 (22.1)	0.95
TOI-Anemia	96.7 (23.7)	0.96	not administered	not administered	73.6 (25.5)	0.95

[a]Sample sizes fluctuate slightly due to missing data within each scale.

Data Analysis

Reliability. Cronbach's alpha coefficients were used to determine internal consistency of all scales and scale aggregates. In sample 1, test–retest reliability was established using intraclass correlations of baseline and 3–7 day retest scores.

Cross-Sectional Analyses of Baseline Data. Across all samples, estimates for clinically important differences were derived from comparisons of baseline FACT scores across two clinical indicators: hemoglobin level and PSR. These boundaries set higher end limits on what constitutes a CID. Independent samples *t*-tests and one-way analyses of variance (ANOVAs) were used for group comparisons. Tukey HSD tests ($P < 0.05$) were used to specify any significant ANOVA effects. Group mean differences and effect sizes (group mean difference divided by pooled within-group standard deviation) were computed in an effort to further clarify these conservatively estimated clinically important differences.

Longitudinal Analyses of Change. Using longitudinal data from samples 2 and 3 we assessed the association of changing clinical status (hemoglobin level, PSR [samples 2 and 3], and best overall response to chemotherapy [sample 3]) with changing FACT scores to further specify clinically important differences. FACT change scores were calculated by subtracting baseline scores from follow-up scores; thus a positive change score indicated improvement in QOL and a negative change score indicated declining QOL. Independent-samples *t*-tests and ANOVAs (with Tukey HSD tests, $P < 0.05$) were used to determine differences in FACT change scores over time based on changes in clinical status. Group mean differences and ef-

Table 3
Mean FACT Scores by Hemoglobin Level at Baseline (Sample 1)

FACT Scale or Aggregate	Hemoglobin Level (1) <10 g/dl ($n = 9$)	(2) 10–11.99 g/dl ($n = 13$)	(3) ≥12 g/dl ($n = 27$)	*F* (post-hoc)	Adjacent Category Mean Differences	Adjacent Category Effect Sizes
Fatigue Scale	30.8 SD = 14.9	33.8 SD = 13	40.2 SD = 8.4	$F(2,46) = 3.8^a$ 3 > 1	3.0, 6.4	0.29, 0.62
FACT-G	76.1 SD = 16.8	79.2 SD = 15.5	90.7 SD = 11.5	$F(2,46) = 5.4^b$ 3 > 1,2	3.1, 11.5	0.22, 0.82
FACT-Anemia	124.8 SD = 33.4	131.3 SD = 24.5	152.1 SD = 20.9	$F(2,46) = 5.8^b$ 3 > 1,2	6.5, 20.8	0.25, 0.80
TOI-Fatigue	64.7 SD = 23.5	68.5 SD = 19.0	85.1 SD = 15.7	$F(2,46) = 6.2^b$ 3 > 1,2	3.8, 16.6	0.19, 0.83
TOI-Anemia	82.6 SD = 28.6	86.8 SD = 22.3	106.4 SD = 18.5	$F(2,46) = 6.0^b$ 3 > 1,2	4.2, 19.6	0.18, 0.83

[a]$P < 0.05$; [b]$P < 0.01$.

Table 4
Mean FACT Scores by Hemoglobin Level at Baseline (Sample 2)

FACT Scale or Aggregate	Hemoglobin Level (1) <10 g/dl $n = 14$	(2) 10–11.99 g/dl $n = 31$	(3) ≥12 g/dl $n = 86$	F (post-hoc)	Adjacent Category Mean Differences	Adjacent Category Effect Sizes
Fatigue Scale	32.9 SD = 14.2	37.0 SD = 10.1	40.3 SD = 10.2	$F(2,128) = 3.4^a$ 3 > 1	3.3, 4.1	0.30, 0.38
FACT-G	78.1 SD = 19.8	84.2 SD = 13.9	83.5 SD = 14.1	$F(2,128) = 0.9$	−0.7, 6.1	−0.05, 0.41
TOI-Fatigue	70.1 SD = 28.4	78.9 SD = 19.1	82.7 SD = 18.9	$F(2,128) = 2.5$	3.8, 8.8	0.19, 0.43

[a] $P < 0.05$.

fect sizes were also calculated to determine the relative magnitude of significant differences.

Distribution-Based Measures of Clinical Significance. Two distribution-based methods were used to identify the magnitudes of difference in FACT scale and scale aggregates that can be considered clinically important. First, standard deviations of FACT scores were divided by 2 to establish a 1/2 standard deviation change standard. (Cohen[33] suggested 0.50 as a medium-sized effect. Effect size of a change score is usually determined by dividing the change score by the standard deviation of patients at baseline: $M2 - M1/\sigma_1$. Because in this formula, a change of 0.50 units would result in an effect size of 0.50 if the $SD_{baseline} = 1.0$, we selected 0.50 standard deviation units to approximate Cohen's medium effect.) Next, we computed the standard error of measurement (SEM) for all FACT scores. The SEM was computed using the following formula: $SEM = \sigma_x (1 - r_{xx})^{1/2}$ where σ_x = the standard deviation of the scale/aggregate score and r_{xx} = the reliability (internal consistency) of the scale/aggregate score. Wyrwich and colleagues[28] have suggested that a difference of 1 SEM frequently corresponds to a minimally important difference.

Results

Sample Descriptives

Demographic and clinical characteristics of the samples are shown in Table 1. Most patients were Caucasian. In comparison to samples 1 and 2, patients in sample 3 were slightly older, were more likely to have been diagnosed with a hematologic malignancy, and had worse clinical characteristics (lower median hemo-

Table 5
Mean FACT Scores by Hemoglobin Level at Baseline (Sample 3)

FACT Scale or Aggregate	Hemoglobin Level (1) <8 g/dl	(2) 8–9.99 g/dl	(3) 10–11 g/dl	F (post-hoc)	Adjacent Category Mean Differences	Adjacent Category Effect Sizes
Fatigue Scale	20.6 (12.1) $n = 215$	23.3 (12.4) $n = 1349$	26.2 (12.9) $n = 653$	F(2,2214) = 20.1 $P < 0.0001$ 3 > 2 > 1	2.7, 2.9	0.21, 0.23
FACT-G	67.7 (16.6) $n = 214$	71.3 (15.6) $n = 1328$	74.3 (15.8) $n = 646$	F(2,2185) = 16.1 $P < 0.0001$ 3 > 2 > 1	3.0, 3.6	0.19, 0.23
FACT-Anemia	104.6 (30.1) $n = 213$	112.0 (29.1) $n = 1328$	118.5 (29.9) $n = 643$	F(2,2181) = 20.7 $P < 0.0001$ 3 > 2 > 1	6.5, 7.4	0.22, 0.25
TOI-Fatigue	50.2 (21.9) $n = 214$	55.0 (21.5) $n = 1338$	60.3 (22.4) $n = 648$	F(2,2197) = 21.4 $P < 0.0001$ 3 > 2 > 1	4.8, 5.3	0.22, 0.24
TOI-Anemia	66.6 (25.4) $n = 214$	72.4 (25.0) $n = 1338$	78.3 (25.8) $n = 648$	F(2,2197) = 21.2 $P < 0.0001$ 3 > 2 > 1	5.8, 5.9	0.23, 0.23

Table 6
Mean FACT Scores By Patient-rated Performance Status[a] at Baseline (Sample 1)

FACT Scale or Aggregate	Performance Status Rating 0 $n = 17$	1 $n = 22$	2–3 $n = 27$	F (post-hoc)	Adjacent Category Mean Differences	Adjacent Category Effect Sizes
Fatigue Scale	41.6 SD = 10.5	38.2 SD = 5.3	25.5 SD = 11.6	$F(2,46) = 11.0^b$ 0,1 > (2–3)	3.4, 12.7	0.32, 1.22
FACT-G	88.4 SD = 12.1	78.3 SD = 12.4	66.7 SD = 8.9	$F(2,46) = 11.2^b$ 0 > 1 > (2–3)	10.1, 11.6	0.73, 0.83
FACT-Anemia	152.1 SD = 23.9	136.4 SD = 18.1	108.1 SD = 22.5	$F(2,46) = 13.9^b$ 0,1 > (2–3)	15.7, 28.3	0.60, 1.09
TOI-Fatigue	89.0 SD = 18.5	78.7 SD = 12.8	52.6 SD = 14.3	$F(2,46) = 18.1^b$ 0,1 > (2–3)	10.3, 26.1	0.52, 1.31
TOI-Anemia	111.1 SD = 21.6	98.6 SD = 14.7	68.5 SD = 19.4	$F(2,46) = 17.3^b$ 0,1 > (2–3)	12.5, 30.1	0.53, 1.28

[a]Eastern Cooperative Oncology Group performance status: 0 = fully ambulatory with no symptoms; 1 = ambulatory with some symptoms; 2–3 = requiring bedres during the day.
[b]$P < 0.001$.

globin counts and lower PSR). Common solid tumor sites in all three samples were breast, lung, and colon.

FACT Descriptives and Reliability

Baseline descriptive and internal consistency (coefficient alpha) statistics on targeted FACT scales and scale aggregates are displayed in Table 2. FACT scores were highly internally consistent across all samples (all alpha coefficients above 0.85). Because the questionnaire was administered 3–7 days after baseline testing in sample 1, test stability (test–retest reliability) could also be assessed. Intra-class correlation coefficients for the five targeted scales showed very good stability: $r_{FS} = .89$; $r_{FACT\text{-}An} = .86$; $r_{FACT\text{-}G} = .82$; $r_{TOI\text{-}F} = .88$; $r_{TOI\text{-}An} = .87$.

Cross-Sectional Analyses

FACT scores were first compared across baseline hemoglobin level. Hemoglobin level was trichotomized in each of the three samples: <10 g/dl, 10–11.9 g/dl, and ≥12 g/dl in samples 1 and 2; and <8 g/dl, 8–9.9 g/dl, and 10–11 g/dl in sample 3. As patients in sample 3 were entered into a clinical trial of the effects of epoetin alfa on anemia, they had relatively low baseline hemoglobin levels (mean = 9.3, SD = 1.0; range 4.5–11.0). Results of these analyses appear in Tables 3–5. We were particularly interested in mean differences and effect sizes (ES) of adjacent clinical categories (represented in the final two columns of each table). Adjacent categories can be construed as representing clinically distinguishable groups.

Next, we compared FACT scores across baseline PSR (patient-rated ECOG PSR for samples 1 and 2 and physician-rated KPSR for sample 3). For the purpose of analysis, ECOG PSR was trichotomized (0, 1, 2–3), while the KPSR was collapsed into 5 categories (90–100, 80, 70, 60, 50 and below). Results of these analyses appear in Tables

Table 7
Mean FACT Scores By Patient-rated Performance Status at Baseline (Sample 2)

FACT Scale or Aggregate	Performance Status Rating 0 $n = 79$	1 $n = 36$	2–3 $n = 16$	F (post-hoc)	Adjacent Category Mean Differences	Adjacent Category Effect Sizes
Fatigue Scale	42.2 SD = 9.4	38.1 SD = 8.3	23.1 SD = 9.1	$F(2,128) = 29.6$*** 0,1 > (2–3)	4.1, 15.0	0.38, 1.38
FACT-G	86.7 SD = 13.1	83.2 SD = 12.9	65.1 SD = 13.4	$F(2,128) = 18.0$*** 0,1 > (2–3)	3.5, 18.1	0.24, 1.23
TOI-fatigue	86.7 SD = 17.7	79.1 SD = 16.1	52.5 SD = 17.6	$F(2,128) = 26.2$*** 0,1 > (2–3)	7.6, 26.6	0.37, 1.31

[a]Eastern Cooperative Oncology Group performance status: 0 = fully ambulatory with no symptoms; 1 = ambulatory with some symptoms; 2–3 = requring bedrest during the day.
***$P < 0.001$.

Table 8
Mean FACT Scores by Karnofsky Performance Status at Baseline (Sample 3)

FACT Scale or Aggregate	Performance Status Rating (1) 90–100	(2) 80	(3) 70	(4) 60	(5) 50 and below	*F* (post-hoc)	Adjacent Category Mean Differences	Adjacent Category Effect Sizes
Fatigue Scale	29.4 (12.4) $n = 722$	24.0 (12.3) $n = 651$	20.5 (10.8) $n = 438$	19.4 (11.0) $n = 226$	15.6 (10.8) $n = 182$	$F(4,2214) = 78.6$ $P < 0.0001$ $1 > 2 > 3,4 > 5$	1.1, 3.5 3.8, 5.4	0.09, 0.28 0.30, 0.43
FACT-G	77.9 (15.0) $n = 697$	72.8 (14.9) $n = 627$	68.9 (14.5) $n = 425$	66.5 (14.9) $n = 219$	59.5 (15.5) $n = 173$	$F(4,2136) = 70.1$ $P < 0.0001$ $1 > 2 > 3,4 > 5$	2.4, 3.9 5.1, 7.0	0.15, 0.25 0.32, 0.44
FACT-Anemia	126.5 (28.6) $n = 712$	114.3 (28.0) $n = 642$	105.7 (25.4) $n = 434$	102.0 (26.2) $n = 223$	89.6 (27.2) $n = 176$	$F(4,2182) = 91.9$ $P < 0.0001$ $1 > 2 > 3,4 > 5$	3.7, 8.6 12.2, 12.4	0.12, 0.29 0.41, 0.42
TOI-Fatigue	66.3 (21.4) $n = 717$	56.8 (21.1) $n = 648$	50.1 (18.4) $n = 435$	47.3 (19.5) $n = 224$	38.9 (19.0) $n = 19.0$	$F(4,2197) = 96.7$ $P < 0.0001$ $1 > 2 > 3,4 > 5$	2.8, 6.7 8.4, 9.5	0.13, 0.30 0.38, 0.43
TOI-Anemia	85.6 (24.7) $n = 717$	74.2 (24.3) $n = 648$	66.5 (21.4) $n = 435$	63.3 (22.3) $n = 224$	53.6 (22.1) $n = 178$	$F(4,2197) = 99.3$ $P < 0.0001$ $1 > 2 >> 3,4 > 5$	3.2, 7.7 9.7, 11.4	0.13, 0.30 0.38, 0.45

6–8. Mean differences and effect sizes of adjacent clinical categories were again of primary interest.

Longitudinal Analyses of Change

To further specify clinically important differences on the FACT, we compared change scores across changing clinical status. For sample 2, we calculated baseline to 6-month and baseline to 12-month changes in Fatigue Scale, FACT-G, and TOI-F scores across change in hemoglobin (Hgb) level and changes in ECOG PSR. We did not analyze baseline to 3-month changes in this sample as patients were either still receiving chemotherapy or they had very recently finished. Chemotherapy on its own caused patients' scores to decline regardless of Hgb level or PSR, and this introduced some variability into the CID estimates. For sample 3, we calculated baseline to completion changes in Fatigue Scale, FACT-G, FACT-An, TOI-F, and TOI-An scores across changes in Hgb, change in KPSR, and best overall response to chemotherapy. The mean time to completion for patients in sample 3 was 89 days (SD = 37).

Table 9
FACT Change Scores by Change in Hemoglobin (Hgb) Level From Baseline to Follow-up (Sample 2)

FACT Scale or Aggregate	Change in Hgb: Increased or no change[a]	Decreased[b]	*t*	Mean Differences	Effect Sizes
$\Delta_{\text{Fatigue Scale}}$					
Baseline to 6 months	−0.5 (SD = 11.6) $n = 54$	−3.7 (SD = 10.5) $n = 26$	$t(78) = 1.2$	3.2	0.29
Baseline to 12 months	3.6 (SD = 9.2) $n = 45$	−3.8 (SD = 6.6) $n = 11$	$t(54) = 2.5^c$	7.4	0.68
$\Delta_{\text{FACT-G}}$					
Baseline to 6 months	0.7 (SD = 13.9) $n = 54$	−1.7 (SD = 11.4) $n = 26$	$t(78) = 0.7$	2.3	0.16
Baseline to 12 months	1.7 (SD = 11.2) $n = 45$	−6.4 (SD = 7.9) $n = 11$	$t(54) = 2.2^c$	8.1	0.55
$\Delta_{\text{TOI-Fatigue}}$					
Baseline to 6 months	−0.2 (SD = 20.8) $n = 54$	−6.5 (SD = 19.1) $n = 26$	$t(78) = 1.3$	6.2	0.31
Baseline to 12 months	4.7 (SD = 14.9) $n = 45$	−9.8 (SD = 12.2) $n = 11$	$t(54) = 3.0^d$	14.5	0.71

[a]Follow-up Hgb equals or exceeds baseline Hgb. Mean change in $\text{Hgb}_{\text{baseline–6months}} = 1.04$.
[b]Follow-up Hgb falls below baseline Hgb. Mean change in $\text{Hgb}_{\text{baseline–6months}} = -2.28$.
Mean change in $\text{Hgb}_{\text{baseline–12months}} = -2.18$.
[c]$P < 0.05$; [d]$P < 0.01$.

Table 10
FACT Change Scores by Change in Hemoglobin (Hgb) Level From Baseline to Completion (Sample 3)

FACT Scale or Aggregate	Change in Hgb: Improved[a] (I)	Unchanged[b] (U)	Worsened[c] (D)	*F* (post-hoc)	Adjacent Category Mean Differences	Adjacent Category Effect Sizes
$\Delta_{\text{Fatigue Scale}}$	6.6 (13.7) $n = 1011$	1.7 (11.2) $n = 303$	−4.3 (12.7) $n = 64$	F(2,1375) = 32.8 $P < 0.0001$ I > U > W	4.9, 6.0	0.39, 0.48
$\Delta_{\text{FACT-G}}$	4.9 (15.9) $n = 1003$	−1.6 (14.9) $n = 299$	−5.6 (14.7) $n = 62$	F(2,1361) = 29.4 $P < 0.0001$ I > U,W	4.0, 6.5	0.25, 0.41
$\Delta_{\text{FACT-Anemia}}$	13.2 (30.9) $n = 979$	0.1 (26.4) $n = 295$	−11.6 (28.3) $n = 61$	F(2,1332) = 37.2 $P < 0.0001$ I > U > W	11.7, 13.1	0.39, 0.44
$\Delta_{\text{TOI-Fatigue}}$	10.6 (23.3) $n = 989$	1.3 (19.6) $n = 297$	−9.2 (20.6) $n = 63$	G(2,1346) = 38.7 $P < 0.0001$ I > U > W	9.3, 10.5	0.42, 0.48
$\Delta_{\text{TOI-Anemia}}$	12.3 (26.7) $n = 988$	1.2 (22.3) $n = 297$	−10.8 (23.3) $n = 63$	F(2,1345) = 40.9 $P < 0.0001$ I > U > W	11.1, 12.0	0.44, 0.47

[a]Completion Hgb exceeded baseline Hgb by ≥1 g/dl. Mean change in Hgb = 2.9.
[b]Completion Hgb within 1 g/dl of baseline Hgb. Mean change in Hgb = 0.1.
[c]Completion Hgb falls below baseline Hgb by ≥1 g/dl. Mean change in Hgb = −1.8.

Changes in Hgb Level. We derived change categories by setting 1.0 g/dl as a meaningful difference. However, in the case of sample 2 where there would not be sufficient cases across three change categories, we classified patients as 'increased/no change' if their 6-month Hgb level equaled or exceeded their baseline level and 'decreased' if their 6-month Hgb level fell below their baseline level. We then combined the increased/no change categories into a single category because relatively few patients showed improvements in Hgb over time. For sample 3, we classified patients as 'improved' if completion Hgb exceeded baseline Hgb by ≥1 g/dl; 'worsened' if completion Hgb fell below baseline Hgb by ≥1 g/dl; and 'unchanged' if completion Hgb was within 1 g/dl of baseline Hgb. Results of these analyses appear in Tables 9 and 10. As in the cross-sectional analyses we were mainly interested in mean differences and effect sizes of adjacent clinical categories as they represent clinically

Table 11
FACT Change Scores by Change in Performance Status Rating (PSR) From Baseline to Follow-up (Sample 2)

FACT Scale or Aggregate	Change in PSR: Improved PSR (I)[a]	Unchanged PSR (U)[b]	Worsened PSR (W)[c]	*F* (post-hoc)	Adjacent Category Mean Differences	Adjacent Category Effect Sizes
$\Delta_{\text{Fatigue Scale}}$						
Baseline to 6 months	7.9 (SD = 11.2) $n = 11$	0.2 (SD = 9.5) $n = 47$	−6.0 (SD = 10.8) $n = 38$	$F(2,93) = 9.0^{d}$ I > U > W	6.2, 7.7	0.57, 0.71
Baseline to 12 months	9.6 (SD = 8.2) $n = 14$	0.8 (SD = 9.9) $n = 51$	1.0 (SD = 8.1) $n = 17$	$F(2,79) = 5.2^{e}$ I > U,W	0.2, 8.8	0.02, 0.81
$\Delta_{\text{FACT-G}}$						
Baseline to 6 months	8.6 (SD = 13.9) $n = 11$	3.6 (SD = 11.7) $n = 47$	−6.3 (SD = 12.8) $n = 38$	$F(2,93) = 9.5^{d}$ I,U > W	5.0, 9.9	0.34, 0.67
Baseline to 12 months	5.4 (SD = 12.2) $n = 14$	−0.4 (SD = 10.3) $n = 51$	1.5 (SD = 11.5) $n = 17$	$F(2,79) = 1.6^{d}$ I > U,W	1.9, 5.8	0.13, 0.39
$\Delta_{\text{TOI-Fatigue}}$						
Baseline to 6 months	16.2 (SD = 18.0) $n = 11$	2.1 (SD = 16.8) $n = 47$	−11.2 (SD = 20.7) $n = 38$	$F(2,93) = 11.0^{d}$ I,U > W	13.3, 14.1	0.66, 0.69
Baseline to 12 months	14.3 (SD = 14.5) $n = 14$	0.1 (SD = 16.6) $n = 51$	0.9 (SD = 14.9) $n = 17$	$F(2,79) = 4.5^{f}$ I > U,W	0.8, 14.2	0.04, 0.70

[a]Follow-up PSR exceeded baseline PSR.
[b]Follow-up PSR equaled baseline PSR.
[c]Follow-up PSR fell below baseline PSR.
[d]$P < 0.001$; [e]$P < 0.01$; [f]$P < 0.05$.

Table 12
FACT Change Scores by Change in Karnofsky Performance Status Rating (KPSR) (Sample 3)

FACT Scale or Aggregate	Change in KPSR: Improved KPSR (I)[a]	Change in KPSR: Unchanged KPSR (U)[b]	Change in KPSR: Worsened KPSR (W)[c]	*F* (post-hoc)	Adjacent Category Mean Differences	Adjacent Category Effect Sizes
$\Delta_{\text{Fatigue Scale}}$	10.5 (12.5) $n = 404$	4.8 (12.1) $n = 606$	−0.1 (14.4) $n = 401$	F(2,1408) = 68.0 $P < 0.0001$ I > U > W	4.9, 5.7	0.36, 0.42
$\Delta_{\text{FACT-G}}$	8.8 (14.9) $n = 402$	3.8 (14.3) $n = 601$	−3.9 (16.9) $n = 394$	F(2,1394) = 70.5 $P < 0.0001$ I > U > W	5.0, 7.7	0.31, 0.48
$\Delta_{\text{FACT-Anemia}}$	21.7 (28.3) $n = 396$	9.3 (27.0) $n = 583$	−4.1 (32.8) $n = 389$	F(2,1365) = 77.4 $P < 0.0001$ I > U > W	12.4, 13.4	0.40, 0.43
$\Delta_{\text{TOI-Fatigue}}$	17.2 (21.2) $n = 399$	8.0 (20.4) $n = 590$	−2.5 (24.6) $n = 393$	F(2,1379) = 79.6 $P < 0.0001$ I > U > W	9.2, 10.5	0.40, 0.45
$\Delta_{\text{TOI-Anemia}}$	19.7 (24.4) $n = 398$	9.0 (23.4) $n = 590$	−2.5 (28.2) $n = 393$	F(2,1378) = 77.6 $P < 0.0001$ I > U > W	10.7, 11.5	0.40, 0.43

[a]Follow-up KPSR exceeded baseline KPSR.
[b]Follow-up KPSR equaled baseline KPSR.
[c]Follow-up KPSR fell below baseline KPSR.

distinguishable groups. These are shown in the final two columns of the tables.

Changes in PSR. Performance status changes were based on changes in PSR categories as previously defined (ECOG PSR: 0, 1, 2–3; and KPS 90–100, 80, 70, 60, 50 and below). For samples 2 and 3 patients were classified 'improved' if their follow-up PSR exceeded their baseline PSR; 'unchanged' if their follow-up PSR equaled their baseline PSR; or 'worsened' if their follow-up PSR fell below their baseline PSR. Results of these analyses appear in Tables 11 and 12.

Best Overall Response to Treatment. Finally, for sample 3 we compared FACT change scores from baseline to completion by best overall response to treatment while on study. Response to treatment was categorized as either complete or partial response (CR/PR), stable disease (SD), or progressive disease (PD). Results appear in Table 13.

Summary of Anchor-Based Measures

To synthesize the data obtained from these analyses, we summarize mean differences and effect sizes in Table 14. Only those mean differences that correspond to an effect size of at

Table 13
FACT Change Scores by Best Overall Response to Treatment (Sample 3)

FACT Scale or Aggregate	Best Response to Treatment: Complete or Partial Response (CR/PR)	Best Response to Treatment: Stable Disease (SD)	Best Response to Treatment: Progressive Disease (PD)	*F* (post-hoc)	Adjacent Mean Differences	Adjacent Effect Sizes
$\Delta_{\text{Fatigue Scale}}$	8.5 (12.9) $n = 656$	4.6 (12.4) $n = 415$	−2.0 (13.4) $n = 367$	F(2,1368) = 68.6 $P < 0.0001$ CR/PR > SD > PD	3.9, 6.6	0.31, 0.52
$\Delta_{\text{FACT-G}}$	7.0 (15.0) $n = 656$	2.9 (14.5) $n = 405$	−6.0 (16.4) $n = 296$	F(2,1354) = 75.1 $P < 0.0001$ CR/PR > SD > PD	4.1, 8.9	0.26, 0.56
$\Delta_{\text{FACT-Anemia}}$	17.3 (29.0) $n = 642$	8.6 (27.9) $n = 398$	−8.8 (30.8) $n = 289$	F(2,1326) = 82.2 $P < 0.0001$ CR/PR > SD > PD	8.7, 17.4	0.29, 0.59
$\Delta_{\text{TOI-Fatigue}}$	14.1 (21.9) $n = 645$	7.0 (21.0) $n = 404$	−5.1 (23.1) $n = 294$	F(2,1340) = 78.1 $P < 0.0001$ CR/PR > SD > PD	7.1, 12.1	0.32, 0.55
$\Delta_{\text{TOI-Anemia}}$	16.1 (25.2) $n = 645$	8.1 (24.0) $n = 404$	−5.8 (26.6) $n = 293$	F(2,1339) = 76.9 $P < 0.0001$ CR/PR > SD > PD	8.0, 13.9	0.31, 0.55

Table 14
Summary of Mean Differencs and Effect Sizes for Anchor-Based Analyses

FACT Score or Aggregate	Cross-sectional Analyses		Longitudinal Analyses	
	Mean Difference Range	Effect Size Range	Mean Difference Range	Effect Size Range
Fatigue Scale	2.7–15.0	0.21–1.38	3.2–8.8	0.29–0.81
FACT-G	3.1–18.1	0.22–1.23	4.0–9.9	0.25–0.67
FACT-Anemia	6.5–28.3	0.22–1.09	8.7–17.4	0.29–0.59
TOI-Fatigue	4.8–26.6	0.22–1.31	6.2–14.5	0.31–0.71
TOI-Anemia	5.8–30.1	0.23–1.28	8.0–13.9	0.31–0.55

least 0.20 appear in the table. This magnitude corresponds to a "small effect"[33] and is a likely approximation of a difference that is minimally clinically important.

Summary of Distribution-Based Measures

Standard deviations of FACT scores were divided by 2 to determine a clinical significance standard of 1/2 of the standard deviation (SD). Standard errors of measurement (SEM) were also calculated for each FACT score. The results appear in Table 15.

Discussion

Across three data sets of cancer patients reporting fatigue and anemia-related concerns using the FACT measurement system, we estimated clinically important differences. Analyses included cross-sectional differences in clinical indicators (Hgb level and PSR), longitudinal differences associated with changes in clinical function (e.g., Hgb change, PSR change, and response to treatment), and distribution-based methods (1/2 SD and 1 SEM). We focused our efforts on 5 self-report endpoints of the FACT—Fatigue and FACT—Anemia questionnaires; the Fatigue Scale, the FACT-G, the FACT-An, the TOI-F, and the TOI-An. All selected endpoints had high internal consistency, with coefficients at 0.85 or higher. Furthermore, test–retest data from sample 1 reflected good test stability over time (all intraclass r's > 0.80). Thus, the targeted endpoints appear to be reliably measured. They also represent the most clinically relevant outcomes of patients experiencing fatigue and/or anemia in the context of cancer treatment.

Tables 14 and 15 summarize the anchor and distribution-based analyses used in estimating CIDs. For the anchor-based criteria, we required observed differences to exceed an effect size threshold of 0.20 (a "small" effect). Mean differences corresponding to this size of effect and higher were retained to estimate CIDs on the FACT-An and FACT-F. For distribution-based criteria we used 1/2 of the SD and 1 SEM. Some have suggested that 1/2 SD corresponds to a reasonable estimate of a CID. This can be justified inasmuch as a "moderate" effect size corresponds to an average change score equal to 0.5 SD of the baseline (e.g., pretreatment) score.[34] The single SEM criterion has recently been offered as a less sample-dependent CID criterion. It has been reassuring to note in some recent studies that CIDs estimated using the single SEM criterion have agreed closely with CID estimates using anchor-based methods.[28,29] These tables suggest that based on this particular compilation of clinical anchors (performance status, hemoglobin level, response to treatment), it would be difficult to consider differences of less than 2.7 on the Fatigue Scale, 3.1 on the FACT-G, 6.5 on the FACT-An, 4.8 on the TOI-F, and 5.8 on the TOI-An to be clinically important. This offers indirect evidence for the whole number just above these lower bound values as the minimal clinically important differences on the five targeted summary scores of this questionnaire. In concrete terms, this magnitude of change can be thought of as responding one category different (better or worse) on 3 of the 13 (23%) questions on the Fatigue Subscale, with no change in the other 10 questions. On the FACT-G, this translates to a one category change on 4 of 27 (15%) questions; 7 of 47 (15%) of the FACT-An questions; 5 of 27 (19%) of the TOI-Fatigue questions; and 6 of 34 (18%) of the TOI-Anemia questions. The similarity in percentages (15–23%) is noteworthy.

These estimates can provide a basis for sample size estimation when planning for a clinical trial or other longitudinal study where the pur-

Table 15
Summary of Distribution-Based Criteria of Clinical Significance

	Criterion	
	1/2 SD	SEM
Fatigue subscale		
Sample 1	5.2	2.8
Sample 2		
Baseline	5.5	2.4
3 months	6.2	2.8
6 months	5.7	2.3
12 months	4.8	2.3
Sample 3		
Baseline	6.3	3.1
Completion	6.7	2.7
Mean	5.8	2.6
FACT-G		
Sample 1	7.0	5.4
Sample 2		
Baseline	7.4	5.5
3 months	7.9	4.4
6 months	7.7	5.1
12 months	8.2	4.5
Sample 3		
Baseline	8.0	5.5
Completion	8.8	5.3
Mean	7.9	5.1
FACT-Anemia		
Sample 1	13.0	5.8
Sample 3		
Baseline	14.9	7.3
Completion	16.6	6.6
Mean	14.8	6.6
TOI-Fatigue		
Sample 1	10.0	4.0
Sample 2		
Baseline	10.2	4.5
3 months	11.3	3.9
6 months	10.5	3.6
12 months	8.2	3.7
Sample 3		
Baseline	11.1	4.9
Completion	12.0	4.8
Mean	10.5	4.2
TOI-Anemia		
Sample 1	11.8	4.7
Sample 3		
Baseline	12.8	5.7
Completion	14.0	5.6
Mean	12.9	5.3

pose is to ensure detection of meaningful change over time. Sample size estimation for a clinical trial is not the only potential use of CID information. CID information can also be used to classify patients according to changes in QOL over time. Some investigators[34,35] have proposed using CIDs to determine proportions of people who may benefit from treatment. In this manner patients are classified into "QOL improved," "QOL worsened," and "QOL unchanged" categories, based upon whether QOL change scores improve by the factor of the CID, worsen by the factor of the CID, or remain below the absolute values of the CID threshold. Proportionate differences in these categories across treatment arms could be a valuable piece of evidence when decisions about the efficacy of investigational drugs are made. This information can be used in conjunction with more traditional endpoints like response to treatment, change in PSR and change in Hgb to assist investigators in determining treatment efficacy. CID information offers an important supplement, but does not replace or minimize the importance of continued attention to traditional endpoints. Finally, whereas classification of individuals into improved, stable or unchanged groups based on these criteria may produce the most accurate estimate of the proportion of people in a group who change, the reliability of accurate classification of any one patient must be questioned. For many self-reported health endpoints, where internal consistency is often below .90 (e.g., FACT-G in this study, see Table 2) there may be too much error to conclude that a 'true' change has occurred. Therefore, when monitoring individual patients using scales in which the internal consistency falls below .90, these minimal CID values are best considered as suggestive adjuncts to care monitoring. The CID may help to inform or alert the practicing clinician to the possibility that a patient's status has changed in a meaningful way, but this should be confirmed in clinical interview.

One way group-based data on CIDs can be used in individual treatment decision-making involves the concept of "number needed to treat" (NNT).[36,37] The NNT refers to that number of patients one would need to treat in order to see a meaningful benefit in one. If, for example, one sets the Fatigue Scale minimum CID at 3.0, and determined in a clinical trial of 200 patients (100 intervention; 100 placebo) that 60 patients in the intervention arm and 40 patients in the placebo arm improved by scores of 3.0 or more, one can conclude that 20 patients benefited from the intervention above and beyond the benefit of placebo. To convert this number (20 of 100 treated patients) into the NNT in order to receive the measured benefit, one divides the number benefiting (20) into the number treated (100), indicating that aside from the placebo benefit five (5) patients must be treated for one to benefit. This en-

ables the clinician to better weigh the probability of benefit into each clinical decision.

This study, while comprehensive in its inclusion of clinically meaningful anchors and distributional data across three different samples, does have its limitations. First, there is no generally accepted single criterion for meaningful change of self-reported health. Some anchors are approximations of latent (i.e., unobservable) variables. The anchors selected here (hemoglobin level, performance status, and response to chemotherapy) were selected because they are familiar to practitioners and therefore offer the opportunity to provide interpretive benchmarks for observed change scores. For example, a change score of 3.9 on the Fatigue Scale can be equated to the difference a clinician sees in fatigue between the average patient with stable disease during a course of therapy compared to the average patient who shows a partial or complete tumor response (see Table 13; row 1). Each of these three anchors is clinically relevant; none of them is definitive in determining the minimal CID. Here is where the distribution-based analyses offer some further refining; for example by eliminating any anchor-based differences that do not exceed the .20 effect size criterion. This combination of anchor-based and distribution-based methods, while useful, can lead to a range of estimates of the minimal CID rather than a clear-cut single answer. Indeed, it may be that a search for a single number that represents a minimally important difference score is unrealistic. By their nature, anchor-based estimates of CIDs are fully dependent upon the choice of anchor. When one uses different anchors to estimate important differences on a scale related to that anchor, the magnitude of change will range depending upon the strength of the relation between those questions and the chosen anchor. This will naturally lead to a range of estimates of important change that should in turn be interpreted in the context of the chosen anchor(s) and the relation of the scale score to that anchor.

This combination of methods did not include direct patient-estimated global ratings of change or "subjective significance," which has been recommended by some.[25,26] It is possible that subjective global estimates of change across the time period separating measurements might yield different estimates of clinically important change. There is an important difference between selecting more objective, interpretable clinical anchors, not within the direct control of the respondent, and global estimates of change derived from direct query of the respondent. Each has its advantages over the other, and it is yet to be determined in the same samples of patients whether they would derive the same estimates of CID. In a different sample of cancer outpatients, using a global rating of change calibration approach similar to Jaeschke et al.[25] and Osoba et al.,[26] we estimated a minimal CID on the FACT-G.[38] In that study, the "lower bound" estimate for minimal CID on the FACT-G (5.5) was slightly higher than the lower bounds in this study (3.1 to 4.0). This may reflect method-dependent or sample-dependent differences. When one maps a change score of a relatively generic set of questions such as those in the FACT-G to specific, fairly circumscribed clinical anchors such as the three used here, the associated average differences or changes in the overall scores are likely to be fairly low, with modest effect sizes. This could suggest that the 3.1–4.0 minimum CID on the FACT-G be used with caution. This concern is muted when the questions in the scale bear more direct relationship to the associated clinical anchor, such as in those aggregated scores including fatigue questions.

In conclusion, through a series of analyses of FACT-F and FACT-An data, using anchor-based and distribution-based methods, we have provided estimates of CIDs for 5 commonly-aggregated summary scores. Rounding to the whole number above each CID, conservative estimates for each scale are: Fatigue Scale = 3.0; FACT-G = 4.0; FACT-An total = 7.0; TOI-Fatigue = 5.0; and TOI-Anemia = 6.0. These figures can be useful for planning future clinical trials and studies where the purpose is to determine sample size requirements, and in determining clinically important QOL changes in cancer clinical trials.

Acknowledgments

Supported in part by grants from Ortho-Biotech, Inc. and Aventis Pharmaceuticals. The authors would like to thank the patients and staff of the Kellogg Cancer Care Center at Evanston

Northwestern Healthcare for their generous participation in study 2.

References

1. Cella D. The Functional Assessment of Cancer Therapy-Anemia (FACT-An) Scale: a new tool for the assessment of outcomes in cancer anemia and fatigue. Semin Hematol 1997;34(Suppl):13–19.

2. Cella D, Peterman A, Passik S et al. Progress toward guidelines for the management of fatigue. Oncology 1998;12:369–377.

3. Mooney K, Ferrel B, Nail L, et al. Oncology Nursing Society research priorities survey. Oncol Nurs Forum 1991;18:1381–1388.

4. Stetz K, Haberman M, Holcombe J, et al. Oncology Nursing Society research priority survey. Oncol Nurs Forum 1994;22:785–789.

5. Stone P, Richardson A, Ream E, et al. Cancer-related fatigue: inevitable, unimportant and untreatable? Results of a multi-centre patient survey. Ann Oncol 2000;11:971–975.

6. World Health Organization. Cancer pain relief and palliative care. Geneva: WHO (technical report series 804), 1990.

7. Stone P, Richards M, Hardy J. Fatigue in patients with cancer. Eur J Cancer 1998;34:1670–1676.

8. Cella D. Factors influencing quality of life in cancer patients: anemia and fatigue. Sem Oncol 1998;25:43–46.

9. Demetri GD, Kris M, Wade J, et al. Quality-of-life benefit in chemotherapy patients treated with epoetin alfa is independent of disease response or tumor type: results from a prospective community oncology study. J Clin Oncol 1998;16: 3412–3425.

10. Jamar S. Fatigue in women receiving chemotherapy for ovarian cancer. In: Funk SG, Tornquist EM, Champagne MT, Archer GL, Weiss RA, eds. Key aspects of comfort, management of pain, fatigue, and nausea. New York: Springer, 1989.

11. Portenoy RK, Itri LM. Cancer-related fatigue: guidelines for evaluation and management. Oncologist 1999;4:1–10.

12. Smets EMA, Garssen B, Bonke B, DeHaes JCJM. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychometric Research 1995;39:315–325.

13. Yellen SB, Cella DF, Webster K, et al. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Symptom Manage 1997; 13:63–74.

14. Mendoza TR, Wang XS, Cleeland CS, et al. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. Cancer 1999;85:1186–1196.

15. Piper BF, Lindsey AM, Dodd MJ, et al. The development of an instrument to measure the subjective dimensions of fatigue. In: Funk SG, Tornquist EM, Champagne MT, et al., eds. Key aspects of comfort management of pain and nausea. Philadelphia: Springer, 1989.

16. Hann DM, Jacobsen PB, Azzarello LM, et al. Measurement of fatigue in cancer patients: development and validation of the Fatigue Symptom Inventory. Qual Life Res 1998;7:301–310.

17. Glaspy J, Bukowski R, Steinberg D, et al. for the Procrit Group. The impact of therapy with epoetin alfa on clinical outcomes in patients with nonmyeloid malignancies during cancer chemotherapy in community oncology practice. J Clin Oncol 1997;15: 1218–1234.

18. Gabrilove JL, Cleeland CS, Livingston RB, et al. Clinical evaluation of once-weekly dosing of epoetin alfa in chemotherapy patients: improvements in hemoglobin and quality of life are similar to three-times-weekly dosing. J Clin Oncol 2001;19: 2875–2882.

19. Littlewood TJ, Bajetta E, Nortier JWR, et al. Effects of epoetin alfa on hematologic parameters and quality of life in cancer patients receiving nonplatinum chemotherapy: results of a randomized, double-blind, placebo-controlled trial. J Clin Oncol 2001;19:2865–2874.

20. Johnson JR, Temple R. Food and Drug Administration requirements for approval of new anticancer drugs. Cancer Treat Rep 1985;69:1155–1159.

21. Moore MJ, Osoba D, Murphy K, et al. Use of palliative end points to evaluate the effects of mitoxantrone and low-dose prednisone in patients with hormonally resistant prostate cancer. J Clin Oncol 1994;12:689–694.

22. Thatcher N, Hopwood P, Anderson H. Improving quality of life in patients with non-small cell lung cancer: research experience with gemcitabine. Eur J Cancer 1997;33(Suppl 1):S8–S13.

23. Sloan J, Symonds T. Clinical and statistical significance: differences and similarities. Workshop presented at the International Society of Quality of Life Methods Workshop, Washington, DC, February 5–6, 2001.

24. Sloan JA, Cella D, Frost MH, and the Clinical Significance Consensus Meeting Group. Assessing clinical significance in measuring oncology patient quality of life: Introduction to the symposium, content overview, and definition of terms. Mayo Clinic Proc 2002;77:367–370.

25. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989;10: 407–415.

26. Osoba D, Rodrigues G, Myles J, et al. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998;16: 139–144.

27. Lydick E, Epstein RS. Interpretation of quality of life changes. Qual Lif Res 1993;2:221–226.

28. Wyrwich K, Tierney W, Wolinsky F. Further evidence supporting a SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999;52: 861–873.

29. Wyrwich K, Nienaber N, Tierney W, Wolinsky F. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999;37:469–478.

30. Cella D. Manual of the Functional Assessment of Chronic Illness Therapy (FACIT) measurement system. Evanston, IL: Evanston Northwestern Healthcare and Northwestern University, 1997.

31. Zubrod CG, Schneiderman M, Frei E. et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and triethylene thiophosporamide. J Chron Dis 1960;11:7–33.

32. Karnofsky DA, Abelmann WH, Craver LF, et al. The use of the nitrogen mustards in the palliative treatment of carcinoma. Cancer 1948;1:634–656.

33. Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic, 1977.

34. Norman GR, Sridhar FG, Guyatt GH, Walter SD. Reconciling effect size and MID approaches to interpretation of quality of life. Manuscript submitted for publication.

35. Guyatt GH, Juniper EL, Walter SD, et al. Interpreting treatment effects in randomized trials. BMJ 1998;316:690–693.

36. Guyatt GH, Juniper ED, Walter SD, et al. Measuring health-related quality of life. Ann Int Med 1993;118:622–629.

37. Bobbio M, Demichells B, Glustetto C. Completeness of reporting trial results: effect on physicians' willingness to prescribe. Lancet 1994;343: 1209–1211.

38. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. Qual Life Res 2002;11:207–211.

Appendix

Functional Assessment of Cancer Therapy-Anemia (FACT-An)

Physical Well-Being (PWB)
- I have a lack of energy.
- I have nausea.
- Because of my physical condition, I have trouble meeting the needs of my family.
- I have pain.
- I am bothered by the side effects of treatment.
- I feel sick.
- I am forced to spend time in bed.

Social/Family Well-Being (SFWB)
- I feel close to my friends.
- I get emotional support from my family.
- I get support from my friends.
- My family has accepted my illness.
- I am satisfied with family communication about my illness.
- I feel close to my partner (or the person who is my main support).
- I am satisfied with my sex life.

Emotional Well-Being (EWB)
- I feel sad.
- I am satisfied with how I am coping with my illness.
- I am losing hope in the fight against my illness.
- I feel nervous.
- I worry about dying.
- I worry that my condition will get worse.

Functional Well-Being (FWB)
- I am able to work (include work at home).
- My work (include work at home) is fulfilling.
- I am able to enjoy life.
- I have accepted my illness.
- I am sleeping well.
- I am enjoying the things I usually do for fun.
- I am content with the quality of my life right now.

Anemia Subscale (AS)
- *I feel fatigued.
- *I feel weak all over.
- *I feel listless ("washed out").
- *I feel tired.
- *I have trouble *starting* things because I am tired.
- *I have trouble *finishing* things because I am tired.
- *I have energy.
- I have trouble walking.
- *I am able to do my usual activities.
- *I need to sleep during the day.
- I feel lightheaded (dizzy).
- I get headaches.
- I have been short of breath.
- I have pain in my chest.
- *I am too tired to eat.
- I am interested in sex.
- I am motivated to do my usual activities.
- *I need help doing my usual activities.
- *I am frustrated by being too tired to do the things I want to do.
- * I have to limit my social activities because I am tired.

Note. * Items that comprise the Fatigue Scale