Volume 12 • Number 8 • 2009
VALUE IN HEALTH

Evaluating Minimal Important Differences for the FACT-Melanoma Quality of Life Questionnaire

Robert L. Askew, BA,[1] Yan Xing, MD, PhD, MS,[1] J. Lynn Palmer, PhD,[2] David Cella, PhD,[3] Lemuel A. Moye, MD, PhD,[4] Janice N. Cormier, MD, MPH[1]

[1]Department of Surgical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA; [2]Department of Palliative Care and Rehabilitation Medicine, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA; [3]Center on Outcomes, Research, and Education, Evanston Northwestern Healthcare, Evanston, IL, USA; [4]Department of Biostatistics, The University of Texas School of Public Health, Houston, TX, USA

ABSTRACT

Objectives: Minimal Important Differences (MIDs) establish benchmarks for interpreting mean differences in clinical trials involving quality of life outcomes and inform discussions of clinically meaningful change in patient status. The purpose of this study was to assess MIDs for the Functional Assessment of Cancer Therapy-Melanoma (FACT-M).

Methods: A prospective validation study of the FACT-M was performed with 273 patients with stages I through IV melanoma. FACT-M, Karnofsky Performance Scales, and Eastern Cooperative Oncology Group Performance Status scores were obtained at baseline and 3 months following enrollment. Anchor- and distribution-based methods for assessing MIDs were compared, and pattern-mixture modeling was employed to derive multivariate adjusted estimates.

Results: This study indicates that an approximate range for MIDs of the FACT-M subscales is between 5 to 9 points for the Trial Outcome Index, 4 to 6 points for the Melanoma Combined Subscale, 2 to 4 points for the Melanoma Subscale, and 1 to 2 points for the Melanoma Surgery Subscale. Each method produced similar but not identical ranges of MIDs.

Conclusions: The properties of the anchor instrument employed to derive MIDs directly affect resulting MID ranges and point values. When MIDs are offered as supportive evidence of a clinically meaningful change, the anchor instrument used to derive clinically meaningful thresholds of change should be clearly stated along with information supporting the choice of anchor instrument as the most appropriate for the domain of interest.

Keywords: Functional Assessment of Cancer Therapy, melanoma, minimal important differences, patient reported outcomes, quality of life.

Introduction

Quality of life (QOL) measures in cancer research have been shown to be independent predictors of both survival and response to therapy [1–5], and for melanoma in particular, QOL has been shown to be an independent predictor of survival for patients with advanced disease [2,6]. Patient reported outcomes such as QOL can serve multiple purposes, including acting as validation measures of treatment efficacy in the context of clinical trials and serving as reference points for clinical decision-making when modest differences in survival are anticipated among various treatment modalities [7–9].

With respect to measuring and reporting patient reported outcomes in the context of clinical trials, the US Food and Drug Administration published 2006 draft guidance on the methods to derive and interpret Minimal Important Differences (MIDs) for QOL instruments [7]. In this document, which was primarily intended to serve as a guidance for industry related to product/drug labeling, MIDs are defined as the minimum change observed in a patient reported outcome measure (e.g., QOL score) between treatment groups that can be correlated with or interpreted as a treatment benefit. With the increasingly acknowledged importance of patient reported outcomes in the context of clinical trials [7,10], the assessment of MIDs for QOL instruments is central to interpreting study results, in that they can both establish benchmarks for interpreting mean differences and inform the discussion of what constitutes a clinically meaningful change. A variety of techniques have been employed to determine MIDs [7], including distribution-based methods [11–13], anchor-based methods [14,15], empirical rules [16], or combinations of methods [17–23]. These techniques have been applied to many cancer-related QOL instruments, including the Functional Assessment of Cancer Therapy (FACT)-Lung, FACT-Prostate, FACT-Colorectal, and the FACT-Breast [17–22,24], but to date, there has been no comprehensive assessment of MIDs for the FACT-Melanoma (FACT-M) QOL questionnaire.

The Functional Assessment of Chronic Illness and Therapy is a patient reported outcome measurement system composed of a general health-related QOL component for patients with chronic disease coupled with disease-specific modules [25,26]. This hybrid approach for measuring patient reported outcomes has the advantage of allowing for both a more focused assessment of disease-specific profiles and symptoms while retaining comparability across populations due to the common core items [27]. For example, for patients undergoing treatment for cancer, the FACT-General (FACT-G) serves as the general component, and for those with specific malignancies such as melanoma, the general items of the FACT-G are supplemented with the melanoma-specific items of the FACT-M. The FACT-M was developed at the University of Texas M.D. Anderson Cancer Center and has been validated as a patient-reported QOL measure for melanoma patients with American Joint Committee on Cancer stages I through IV disease [28,29].

The objectives of this study were to derive MIDs for the FACT-M using anchor-based methods, to conduct a comparative distribution-based analysis, and to assess the correspondence

Address correspondence to: Janice N. Cormier, Department of Surgical Oncology, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Box 0444, Houston, TX 77030, USA. E-mail: jcormier@mdanderson.org

10.1111/j.1524-4733.2009.00570.x

1098-3015/09/1144 1144–1150

among MID ranges derived from each of the employed methods.

Methods

Clinical, demographic, and QOL data were collected prospectively at the Melanoma and Skin Center of the University of Texas M.D. Anderson Cancer Center in Houston, Texas with the approval of the institutional review board for the protection of human subjects. From 2004 to 2005, new patients and those within 3 years of melanoma diagnosis were recruited when they presented for scheduled appointments, with recruitment efforts targeting an equal proportion of patients with local, regional, and distant metastatic disease. With 80 patients per disease stage grouping (I and II, III, IV), a statistically significant difference in effect sizes of 0.45 or greater could be detected with 80% statistical power. Initial sampling targets were set at 300 participants to account for attrition, while assuring sufficient sample size at analysis. Study participants had to be at least 18 years of age and fluent in English, and enrollment required pathological confirmation of melanoma. Exclusion criteria included inability to consent due to other medical illnesses and disorientation to person, place, or time. Written informed consent was obtained for each of the study participants, and questionnaires were completed at baseline and during regularly scheduled follow-up visits. Clinical data were abstracted from patient files at baseline and for respective follow-up time points.

Instruments

The FACT-M is composed of items from the FACT-G, melanoma-specific items, and items related to melanoma surgery [28,29]. Four constructs comprise the FACT-G scale, with seven items assessing Physical Well-Being (PWB), seven items assessing Social/Family Well-Being, six items assessing Emotional Well-Being, and seven items assessing Functional Well-Being (FWB). With the addition of the 16-item Melanoma Subscale (MS) and the 8-item Melanoma Surgery Subcale (MSS)—collectively known as the Melanoma Combined Scale (MCS), the number of items of the FACT-M totals 51. Each of these subscales has been shown to have high levels of internal consistency (Cronbach's α: 0.71–0.95) and high test–retest reliability (r: 0.71–0.90) [29]. Higher scores on any subscale or total score indicate higher levels of QOL [26]. The Trial Outcome Index (TOI), which is often the most appropriate single patient-reported end point in clinical trials [22], has been defined as the summed score of the seven PWB and seven FWB items from the FACT-G and the disease-specific subscale items.

The Karnofsky Performance Scale (KPS) is designed to assess a patient's status in terms of functional impairment. With a range of 0 (dead) to 100 (normal—no evidence of disease), this scale assists clinicians and caretakers in gauging a patient's ability to perform activities basic to daily living [30,31]. The KPS was administered at the time of study entry and 3 months later. The Eastern Cooperative Oncology Group Performance Status (ECOG-PS) scale is a measure of disease progression and its effect on a patient's daily living [32,33]. The range of the scale is from 0 (representing fully active, pre-disease performance status) to 5 (death). Clinicians assessed patient performance status at baseline and at 3 months to correspond in timing with the other assessments.

Statistical Analysis

Baseline demographic and clinical data were summarized, and FACT-M TOI, MCS, MS, and MSS scores were calculated using STATA statistical software (v. SE9.2, StataCorp, College Station, TX, USA). Mean FACT-M scores were compared across groups based on relevant clinical indicators previously assessed for statistically significant score differences (e.g., disease stage and treatment status) [29]. Patient performance status was assessed at the time of study entry (baseline) and again after 3 months. For the KPS, an increase or decrease of 10 or more units on a 100-unit scale over time was considered clinically meaningful. The ECOG-PS scale is inverted, and an increase or a drop of one or more units on the six-unit scale was considered clinically meaningful. Patient groups (improving, remaining stable, and declining) were created based on these respective performance measures, and mean per-patient change in QOL scores were calculated for each of these groups. The differences between the QOL change scores of the various patient performance groups were identified and served as minimal units of clinically meaningful change in patient QOL [22].

To facilitate a patient-visit level analysis and to mitigate the effect of missingness that is common to studies involving QOL outcomes, a pattern mixture model analysis was employed following previously established methods [23,34,35]. Using SAS version 9.1 (SAS Institute, Cary, NC, USA) for Windows, separate analyses were conducted with the clinical anchors serving as independent variables, the FACT-M subscales serving as dependant variables, and several clinical variables with group-level differences in QOL such as disease stage, treatment status, and drop out status serving as covariates in the model.

Several distribution-based methods were also included for comparison, though the authors recognize that distribution-based methods provide no direct information about differences that are minimal and important [36,37]. From the methods of Wyrwich et al, the standard error of measurement (SEM) method was used to approximate minimum relevant thresholds of change, as the SEM has been shown to align with MID estimates for other important QOL instruments [13]. The SEM method employs the formula

$$SEM = \sigma_x (1 - rel_{xx})^{1/2} \quad (1)$$

where σ_x is the standard deviation of the sample and rel_{xx} is an appropriate measure of reliability for the instrument—in this case, Cronbach's alpha. Separate SEMs were calculated for each time point (baseline and 3 months) and served as comparative reference points for the MID. Measures of effect size have also served as distribution-based reference points for MIDs [38–40]; however, effect sizes are distributionally standardized and thus not in the form of a scale score. Based on the correspondence of modified standard deviation measures (1/3 SD and 1/2 SD) to small and moderate effect sizes, respectively [38], the standard deviation measures of the FACT-M scale scores were divided by two and separately by three to reflect standards of change in the form of scale scores. This step was repeated for each measurement time point (i.e., baseline and 3 months), and again for the standard deviation of per-patient change.

As missing data is a common issue for studies assessing QOL [41], it was important to examine how missingness may have affected the MID estimates. Baseline demographic and clinical characteristics of study participants who completed 3 months of follow-up were compared to those for whom no follow-up data were available. Subsequent analyses included tests of association between missingness and each of the demographic and clinical variables, while the influence of potential covariates was examined.

Table 1 Demographic and clinical profile

	n = 273	%
Median age in years (range)	52 (20–79)	
Race/ethnicity		
White, non-Hispanic	268	98.1
Hispanic	3	1.1
African-American	1	0.4
Other	1	0.4
Gender		
Male	159	58.2
Female	114	41.8
Marital status		
Married	218	79.8
Never married	27	9.9
Separated or divorced	21	7.7
Widowed	7	2.6
Education		
<High school	40	14.6
Some college	84	30.8
College graduate	92	33.7
Graduate school	53	19.4
Missing/Unknown	4	1.5
Primary tumor site		
Head/Neck	41	15.0
Trunk/Back	69	25.3
Upper extremity	54	19.8
Lower extremity	79	28.9
Unknown primary	30	11.0
AJCC disease stage		
Stages I and II	102	37.4
Stage III	100	36.6
Stage IV	71	26.0
Treatment status		
Active treatment	75	27.5
Follow-up surveillance	198	72.5
Surgical procedures		
Complete node dissection	133	48.7
Wide local excision and sentinel node biopsy	105	38.5
Wide local excision only	31	11.4
Fine needle aspiration	4	1.5

AJCC, American Joint Committee on Cancer.

Results

A total of 273 patients were enrolled in the prospective study, and 163 completed assessments at 3 months. Relatively even patient distributions were observed across levels of primary tumor site and education (Table 1). More than 25% of patients were undergoing active treatment at the time of study enrollment with the remainder in post-treatment follow-up surveillance. Stratifying change in QOL by baseline clinical characteristics uncovered significant group differences in QOL scale scores between patients with local-regional and advanced disease ($P < 0.001$) and between those in active treatment or in follow-up surveillance ($P < 0.001$). Differences in TOI scale scores ranged from 12 to 14 points, and for the MCS, they ranged from 6 to 7 points. For the MS and the MSS, differences ranged from 4 to 5 points and 1 to 3 points, respectively. These point differences were expected to provide theoretical upper limits to the MID estimates, as these patient groups were derived from clinically observed distinctions among patients and their corresponding treatment and QOL outcomes.

Table 2 outlines patient QOL as measured by mean FACT-M subscale scores. The mean TOI increased from baseline (130.8) to 3 months (136.9), and on the patient level, the mean per-patient change in TOI score was 5.6. Similar trends were seen for the subscales. As QOL scores improved over time, patient-reported performance status improved, as reflected in both the KPS and ECOG-PS scores. With increasing scores reflecting improvement in patient performance status, the baseline mean KPS score of 93.0 at baseline rose to 95.4 at 3 months with a mean per-patient change of 1.7. Similarly, performance status improved over time on the ECOG-PS with decreasing scale scores representing better health states. The mean ECOG-PS scale score at baseline was 0.3 and at 3 months was 0.2 with a mean per-patient change of –0.1 points.

Table 3 summarizes mean per-patient change in FACT-M subscale scores stratified by the anchor-based performance-change groups. Specifically, changes in QOL scores for patients whose performance status improved were compared with scores for patients whose performance scores decreased and with patients whose performance status remained stable. Differences in QOL change scores between the KPS performance groups ranged from 1.4 to 1.8 for the MSS, 1.9 to 3.4 for the MS, 3.7 to 4.8 for the MCS, and 5.1 to 7.1 for the TOI. Similarly, for the ECOG-PS performance groups, differences in QOL change scores ranged from 0.9 to 2.3 for the MSS, 2.6 to 4.6 for the MS, 4.9 to 5.5 for the MCS, and 8.3 points for the TOI.

Pattern mixture model analysis resulted in statistically significant MID estimates for each of the QOL subscales while controlling for the effect of covariates (Table 4). With the KPS serving as predictor, MID point estimates for the TOI, MCS, MS, and MSS were 9.0, 7.1, 4.6, and 2.6, respectively. When the ECOG-PS served as the predictor, MID point estimates were 12.2, 9.6, 6.1, and 3.5 for the respective subscales.

MID reference points from the standard deviation methods were similar with ranges of 1.3 to 3.0 for the MSS, 1.7 to 3.8 for the MS, 2.5 to 5.9 for the MCS, and 4.1 to 9.7 for the TOI (Table 5). Likewise, the SEM method yielded ranges of 1.5 to 2.3 for the MSS, 2.0 to 2.9 for the MS, 2.6 to 4.1 for the MCS, and 4.1 to 6.5 for the TOI. Figure 1 illustrates the ranges of MIDs for each of the FACT-M subscales stratified by method of analysis. Pattern mixture model estimation resulted in higher MID ranges than the other methods with the confidence interval acting as the

Table 2 Mean baseline, 3 months, and per-patient change in performance and quality of life scores

	Baseline n = 273 Mean (SD)		3 months n = 163 Mean (SD)		Per-patient difference n = 163 Mean (SD)	
Quality of Life measures						
Trial Outcome Index (PWB + FWB + MS + MSS)	130.8	(19.5)	136.9	(16.4)	5.6	(12.3)
Melanoma Combined Subscale (MS + MSS)	82.9	(11.7)	86.7	(9.2)	3.7	(7.3)
Melanoma Subscale (MS)	56.1	(7.6)	58.3	(6.1)	2.0	(5.2)
Melanoma Surgery Subscale (MSS)	26.9	(5.9)	28.4	(4.8)	1.7	(3.8)
Performance measures						
Karnofsky Performance Scale	93.0	(9.0)	95.4	(8.0)	1.7	(6.8)
Eastern Cooperative Oncology Group Performance Status	0.3	(0.5)	0.2	(0.5)	–0.1	(0.4)

Table 3 Mean per-patient change in FACT-M scale scores stratified by performance-change groups

	Change in FACT-M TOI			Change in Melanoma Combined Subscale			Change in Melanoma Subscale			Change in Melanoma Surgery Subscale		
	Mean	SD	Diff	Mean	SD	Diff	Mean	SD	Diff	Mean	SD	Diff
Change in Karofsky Performance Status at 3 months												
Improved (n = 40)	11.4	14.8		7.6	8.4		4.7	6.3		2.9	4.3	
Stable (n = 110)	4.3	10.3		2.8	6.2		1.3	4.1		1.5	3.5	
Declined (n = 13)	−0.9	14.1		−0.9	8.5		−0.6	6.6		−0.3	4.4	
Differences												
(Improved ↔ Stable)			7.1			4.8			3.4			1.4
(Stable ↔ Declined)			5.1			3.7			1.9			1.8
Change in ECOG Performance Status at 3 months												
Improved (n = 22)	13.1	16.6		8.1	10.2		4.4	6.6		3.7	5.8	
Stable (n = 135)	4.8	10.6		3.2	6.4		1.8	4.6		1.4	3.3	
Declined (n = 6)	−3.5	19.2		−2.3	8.5		−2.8	7.2		0.5	3.6	
Differences												
(Improved ↔ Stable)			8.3			4.9			2.6			2.3
(Stable ↔ Declined)			8.3			5.5			4.6			0.9

Diff, difference; ECOG, Eastern Cooperative Oncology Group; FACT-M, Functional Assessment of Cancer Therapy-Melanoma; TOI, Trial Outcome Index.

upper and lower bounds of the suggested MID range. Average MID estimates from the anchor-based methods were higher than those from the distribution-based methods, and on nearly every scale, the MIDs from the ECOG-PS were higher than those from the KPS.

Table 4 Pattern mixture model estimates* stratified by performance-anchors

	Estimate	95% CI	*P* value
Karnofsky Performance Scale			
Trial Outcome Index	8.95	(7.67–10.24)	<0.001
Melanoma Combined Scale	7.12	(6.14–8.11)	<0.001
Melanoma Subscale	4.61	(3.94–5.28)	<0.001
Melanoma Surgery Subscale	2.58	(2.04–3.13)	<0.001
Eastern Cooperative Oncology Group-Performance Scale (ECOG-PS)†			
Trial Outcome Index	12.17	(9.99–14.36)	<0.001
Melanoma Combined Scale	9.59	(7.90–11.27)	<0.001
Melanoma Subscale	6.11	(4.97–7.26)	<0.001
Melanoma Surgery Subscale	3.52	(2.62–4.42)	<0.001

*Model accounts for effects of disease stage, treatment status, surgery, patient drop-out, and interactions.
†ECOG-PS scale is inverted, but absolute values are reported in the table for consistency.

Because a nearly 40% attrition rate for this study was observed at 3 months, group comparisons between patients with and without follow-up data were made for all baseline demographic and clinical characteristics. Although no statistically significant group differences emerged from comparisons of treatment status, a significant difference was found for baseline disease stage ($P < 0.001$), with a larger proportion of patients with metastatic disease (regional and distant) among the missing (71.8%) than among those with follow-up data at 3 months (56.4%). Furthermore, those with metastatic disease at baseline were nearly twice as likely to have missing data at 3 months (OR 1.97; 95% CI [1.17–3.30], $P = 0.01$), suggesting a statistical association. Subsequent examination of the relationship between disease stage and QOL scores revealed that baseline QOL scores differed for those with and without metastatic disease ($P < 0.001$). However, when controlling for disease stage, no association was observed between baseline QOL and missingness.

Discussion

A comprehensive assessment of MIDs makes use of multiple approaches to triangulate on a range of MID estimates [37], and

Table 5 Distribution-based ranges for minimal important differences

Standard deviation (SD) method	SD baseline	SD 3 months	SD per-patient change
FACT-Melanoma Trial Outcome Index (TOI)	19.46	16.41	12.31
Medium effect size (TOI/2)	9.73	8.21	6.16
Small effect size (TOI/3)	6.49	5.47	4.10
Melanoma Combined Subscale (MCS)	11.73	9.22	7.35
Medium effect size (MCS/2)	5.86	4.61	3.67
Small effect size (MCS/3)	3.91	3.07	2.45
Melanoma Subscale (MS)	7.60	6.11	5.16
Medium effect size (MS/2)	3.80	3.06	2.58
Small effect size (MS/3)	2.53	2.04	1.72
Melanoma Surgery Subscale (MSS)	5.91	4.79	3.84
Medium effect size (MSS/2)	2.95	2.39	1.92
Small effect size (MSS/3)	1.97	1.60	1.28
Standard error of measurement (SEM) method	**SEM baseline**	**SEM 3 months**	**SEM per-patient change**
FACT-Melanoma Trial Outcome Index	6.45	5.44	4.08
Melanoma Combined Subscale	4.06	3.19	2.55
Melanoma Subscale	2.94	2.37	2.00
Melanoma Surgery Subscale	2.29	1.86	1.49

FACT, Functional Assessment of Cancer Therapy.

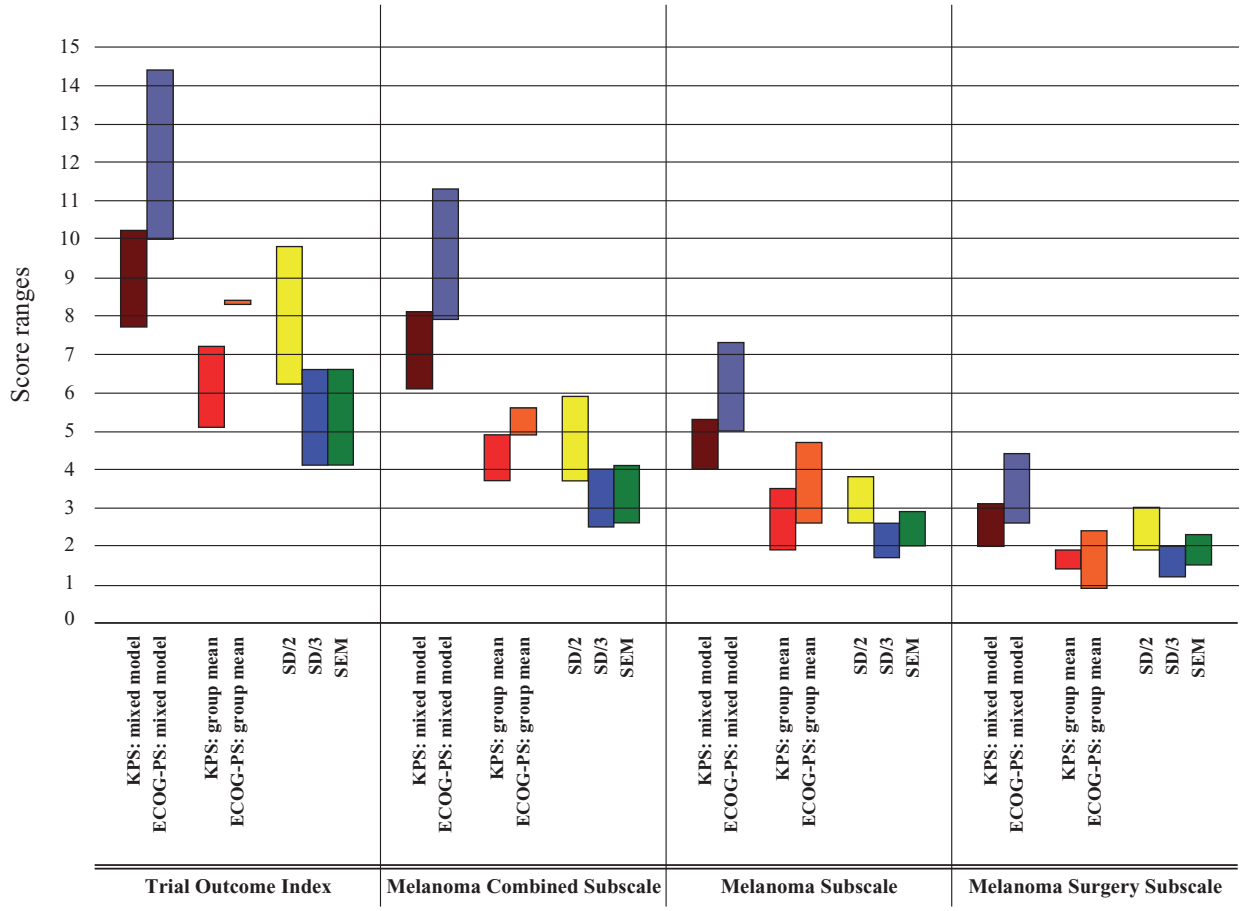

Figure 1 Ranges of Minimal Important Differences (MID) for FACT-Melanoma subscales stratified by method of analysis (, MID range from KPS mixed model; , MID range from ECOG-PS mixed model; , MID range from KPS group mean; , MID range from ECOG-PS group mean; , MID range from Medium Effect Size (SD/2); , MID range from Small Effect Size (SD/3); , MID range from standard error of measurement [SEM]). ECOG-PS, Eastern Cooperative Oncology Group Performance Status; KPS, Karnofsky Performance Scale.

the results of this study indicate that ranges for MIDs of the FACT-M scales are between 5 and 9 points for the TOI, 4 and 6 points for the MCS, 2 and 4 points for the MS, and 1 and 2 points for the MSS. These estimates are similar in range to those of other cancer-specific FACT subscales [18,21,22,42]. Although the anchor- and distribution-based methods employed to assess MIDs resulted in similar ranges, it is important to note that they were not the same, so the strengths and weaknesses of each method warrant further discussion.

One strength of the distribution-based methods is that they can be derived from a single sampling time point (e.g., cross-sectional study design), and the analysis requires few statistical elements to derive the ranges (e.g., the standard deviation of the sample and a measure of the instrument's reliability). However, distribution-based methods lack clinical reference points, and because of this, a range of MIDs derived from distribution-based methods employed in isolation are likely to be viewed with skepticism in regards to their clinical meaning. In contrast, anchor-based methods for deriving MIDs make use of external reference points, and it is from these anchors that the MID ranges derive their clinical meaning. However, anchor-based methods generally require assessments of *change* in patient health status, and measures of change require multiple assessments of a study group over time along with the administration of multiple instruments.

Pattern mixture modeling is an anchor-based method that minimizes the effect of missingness by accounting for it in the model. Additionally, it allows visit-level analysis resulting in increased statistical power when multiple visits are present in the dataset. However, on several of the scales in this analysis, mixed model estimates were higher than group estimates derived from clinical distinctions that served as upper-bounds for the MIDs (i.e., treatment status, and disease stage). Our findings that patients with advanced disease were more likely to have missing data at 3 months and that they were the group expected to experience larger changes in QOL over time were not surprising. Because mixed modeling methods incorporate all visits and covariates, it is logical that MID estimates from mixed modeling were higher than those of the other anchor-based methods. Because the study objective was to ascertain the thresholds of minimal difference that are clinically meaningful, when mixed model estimates were above the upper thresholds set by clinical group distinctions, the estimates derived from mixed models were given less weight when triangulating the final suggested range for each subscale. There is still potential for bias from the lack of follow-up data for patients with rapidly progressing disease [41]; however, larger differences in change scores would likely result in larger MIDs, not smaller ones. Furthermore, the apparent association between baseline QOL and missingness was not

significant when controlling for covariates, and as such, further adjustment of the MID ranges to account for missingness may not be warranted.

At issue is what the MID ranges derived from each method actually represent. One previously noted distinction between the distribution- and anchor-based methods for deriving MIDs is that they reflect separate underlying types of differences—the former reflecting a minimal statistically quantifiable difference in scores and the later reflecting a minimal clinically meaningful difference in scores [43]. With this distinction in mind, comments by de Vet et al. underscore the importance of maintaining an integrated approach, in that by demonstrating that the clinically meaningful differences (e.g., anchor-based MIDs) are above the thresholds set by minimum detectable differences (e.g., distribution-based MIDs), the choice of the anchor instrument is supported [43], as was found in this study. However, minor discordance was observed among the MID ranges derived from the anchor instruments, which warranted further examination of the properties of each anchor instrument.

Although both the KPS and the ECOG-PS scale scores have been strongly correlated with patient survival [44], there remains an important distinction in their general mode of administration, as the ECOG-PS is typically a physician-assessed measure, while the KPS is often patient-assessed. Although others have suggested that the ECOG-PS is a viable substitute for the KPS given its demonstrated validity and reliability [45], the findings of this study illustrate that substitution yields different point ranges for MIDs. In the FACT-M validation study [29], FACT-M scores were found to be more sensitive to change in performance status when performance status was assessed by the KPS. The observed increase in sensitivity with the KPS may be in-part due to issues of instrument granularity as the KPS operates on a scale of 0 to 100 while the ECOG-PS operates on scale of 0 to 5. However, differences in granularity are unlikely to account for all of the observed differences in MID estimates, as interrater variability generally increases with the increasing number of score increments available to raters [45]. Given that the acknowledged limitations of this study include a relatively small sample size, missingness of data, and positively skewed patient status measures, it is possible that the MIDs would realign with a larger population of patients.

Evidence that patients provide valuable and reliable evaluations of their own performance status and QOL that is better correlated than that of their clinicians [45–47] supports the view that MID estimates may be more appropriately derived from patient-assessed anchor instruments—particularly when using QOL instruments like the FACT-M that are also patient-assessed. This point is particularly relevant to the clinical trial setting, as the definition of a responder depends on more than just statistical precision [7]: it depends upon the validity of the reference instruments imparting clinical meaning to the observed mean differences. Our findings that the properties of the anchor instrument directly affect the resulting MIDs support the view that when MIDs are cited, the source anchor instrument should be included as part of the evaluation process.

Source of financial support: National Cancer Institute, Clinical Oncology Research Development Program, 5-K12-CA088084.

References

1 Coates A, Gebski V, Signorini D, et al. Prognostic value of quality-of-life scores during chemotherapy for advanced Breast Cancer. Australian New Zealand Breast Cancer Trials group. J Clin Oncol 1992;10:1833–8.
2 Coates A, Thomson D, McLeod GR, et al. Prognostic value of quality of life scores in a trial of chemotherapy with or without interferon in patients with metastatic malignant melanoma. Eur J Cancer 1993;29A:1731–4.
3 Jerkeman M, Kaasa S, Hjermstad M, et al. Health-related quality of life and its potential prognostic implications in patients with aggressive lymphoma: a Nordic Lymphoma Group Trial. Med Oncol 2001;18:85–94.
4 Maisey NR, Norman A, Watson M, et al. Baseline quality of life predicts survival in patients with advanced colorectal cancer. Eur J Cancer 2002;38:1351–7.
5 Roychowdhury DF, Hayden A, Liepa AM. Health-related quality-of-life parameters as independent prognostic factors in advanced or metastatic bladder cancer. J Clin Oncol 2003;21: 673–8.
6 Butow PN, Coates AS, Dunn SM. Psychosocial predictors of survival in metastatic melanoma. J Clin Oncol 1999;17:2256–63.
7 Burke L, Stifano T, Dawisha S. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79.
8 Brady MJ, Cella DF, Mo F, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. J Clin Oncol 1997;15:974–86.
9 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 Manag 1997;13:63–74.
10 Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res 2000;9:887–900.
11 Rejas J, Pardo A, Ruiz MA. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. J Clin Epidemiol 2008;61:350–6.
12 Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999;37:469–78.
13 Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999;52:861–73.
14 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–15.
15 Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. J Clin Epidemiol 1994;47:81–7.
16 Ringash J, O'Sullivan B, Bezjak A, Redelmeier DA. Interpreting clinically significant changes in patient-reported outcomes. Cancer 2007;110:196–202.
17 Yost KJ, Sorensen MV, Hahn EA, et al. Using multiple anchor- and distribution-based estimates to evaluate clinically meaningful change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) instrument. Value Health 2005;8:117–27.
18 Yost KJ, Cella D, Chawla A, et al. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches. J Clin Epidemiol 2005;58:1241–51.
19 Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes 2007;5:70.
20 Kulkarni AV. Distribution-based and anchor-based approaches provided different interpretability estimates for the Hydrocephalus Outcome Questionnaire. J Clin Epidemiol 2006;59:176–84.
21 Cella D, Nichol MB, Eton D, et al. Estimating clinically meaningful changes for the functional assessment of cancer therapy-prostate: results from a clinical trial of patients with metastatic

hormone-refractory prostate cancer. Value Health 2009;12:124–29.
22 Cella D, Eton DT, Fairclough DL, et al. What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. J Clinical Epidemiol 2002;55:285–95.
23 Barrett B, Brown R, Mundt M. Comparison of anchor-based and distributional approaches in estimating important difference in common cold. Qual Life Res 2008;17:75–85.
24 Eton DT, Cella D, Yost KJ, et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. J Clin Epidemiol 2004;57:898–910.
25 FACIT.org. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System Overview. Available from: http://www.facit.org/about/overview_measure.aspx [Accessed October 3, 2008].
26 Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. Health Qual Life Outcomes 2003;1:79.
27 Cella D, Nowinski CJ. Measuring quality of life in chronic illness: the functional assessment of chronic illness therapy measurement system. Arch Phys Med Rehabil 2002;83(12 Suppl. 2):S10–7.
28 Cormier JN, Davidson L, Xing Y, et al. Measuring quality of life in patients with melanoma: development of the FACT-melanoma subscale. J Support Oncol 2005;3:139–45.
29 Cormier JN, Ross MI, Gershenwald JE, et al. Prospective assessment of the reliability, validity, and sensitivity to change of the Functional Assessment of Cancer Therapy-Melanoma questionnaire. Cancer 2008;112:2249–57.
30 Centers AEaT. Clinical Manual for Management of the HIV-Infected Adult. AETC National Resource Center, Newark, NJ. 2006.
31 Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherpeutic agents in cancer. In: Macleod CM, ed. Evaluation of Chemotherapeutic Agents in Cancer. New York: Columbia University Press, 1949.
32 Group ECO. Eastern Cooperative Oncology Group Performance Status. 2008; Available from: http://ecog.dfci.harvard.edu/general/perf_stat.html [Accessed October 3, 2008].
33 Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5:649–55.
34 Hedeker D, Gibbons R. application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychol Methods 1997;2:64–78.
35 Brant R, Sutherland L, Hilsden R. Examining the minimum important difference. Stat Med 1999;18:2593–603.
36 Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. COPD 2005;2:63–7.
37 Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102–9.
38 Cohen J. Statistical Power Analysis for the Behavioral Sciences (Rev. ed.). New York: Academic Press, 1977.
39 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27(3 Suppl.):S178–89.
40 Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. J Eval Clin Pract 2000;6:39–49.
41 Donaldson GW, Moinpour CM. Learning to live with missing quality-of-life data in advanced-stage disease trials. J Clin Oncol 2005;23:7380–4.
42 Steel JL, Eton DT, Cella D, et al. Clinically meaningful changes in health-related quality of life in patients diagnosed with hepatobiliary carcinoma. Ann Oncol 2006;17:304–12.
43 de Vet HC, Terwee CB, Ostelo RW, et al. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes 2006;4:54.
44 Taylor AE, Olver IN, Sivanthan T, et al. Observer error in grading performance status in cancer patients. Support Care Cancer 1999;7:332–5.
45 de Borja MT, Chow E, Bovett G, et al. The correlation among patients and health care professionals in assessing functional status using the Karnofsky and Eastern Cooperative Oncology Group performance status scales. Support Care Cancer 2004;2:59–63.
46 Lipscomb J, Gotay CC, Snyder C. Outcomes Assessment in Cancer: Measures, Methods, and Applications. Cambridge, UK; New York: Cambridge University Press, 2005.
47 Fromme EK, Eilers KM, Mori M, et al. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol 2004;22:3485–90.