Psychometric Evaluation of the FACT Colorectal Cancer Symptom Index (FCSI-9): Reliability, Validity, Responsiveness, and Clinical Meaningfulness

HILARY H. COLWELL,^a SUSAN D. MATHIAS,^b MICHELLE P. TURNER,^a JOHN LU,^c NICOLA WRIGHT,^c MARC PEETERS,^d DAVID CELLA,^e GIOVANNA DEVERCELLI^f

^aICON Clinical Research, Lifecycle Sciences Group, San Francisco, California, USA; ^bHealth Outcomes Solutions, Winter Park, Florida, USA; ^cAmgen, Inc., Thousand Oaks, California, USA; ^dGent University Hospital, Gastroenterology, Gent, Belgium; ^eCenter on Outcomes Research and Education (CORE), Evanston Northwestern Healthcare, Evanston, Illinois, USA; ^fBayer HealthCare Pharmaceuticals Inc., Montville, New Jersey, USA

Key Words. Quality of life • Reliability and validity • Questionnaire • Patient outcomes • Colorectal cancer

Disclosures: Hilary H. Colwell: Research funding/contracted research: Amgen; Susan D. Mathias: Research funding/contracted research: Amgen; Michelle P. Turner: Research funding/contracted research: Amgen; John Lu: Employment/leadership position: Amgen; Nicola Wright: Employment/leadership position: Amgen; Marc Peeters: Research funding/contracted research: Amgen; David Cella: None; Giovanna Devercelli: Employment/leadership position: Amgen (at the time of the study).

The content of this article has been reviewed by independent peer reviewers to ensure that it is balanced, objective, and free from commercial bias. No financial relationships relevant to the content of this article have been disclosed by the independent peer reviewers.

ABSTRACT

Background. Patient-reported outcomes (PROs) are essential for evaluating treatment effects on health-related quality of life and symptoms from the patient's perspective. This study sought to evaluate the psychometric properties of the nine-item Functional Assessment of Cancer Therapy/National Comprehensive Cancer Network Colorectal Cancer Symptom Index (FCSI-9) in a metastatic colorectal cancer (mCRC) population.

Methods. The FCSI-9 and EQ-5D were administered every 2–4 weeks to mCRC subjects in a phase III clinical trial. Three hundred ninety-one mCRC subjects completed the questionnaires at baseline and at least one follow-up assessment. Internal consistency reliability, test–retest reliability, construct validity, known groups validity, responsiveness, and the minimum important difference (MID) of the FCSI-9 were evaluated.

Results. The internal consistency and test–retest reliability of the FCSI-9 were acceptable (0.81 and 0.76, respectively). Construct validity was supported based on moderate correlations with the EQ-5D. Known groups validity was evaluated by examining the FCSI-9 scores of subjects categorized by their Eastern Cooperative Oncology Group performance status (PS) score. Subjects with better PS scores reported significantly higher FCSI-9 scores than those with lower PS scores at both baseline and week 8. Responsiveness, as measured by Guyatt's statistic, was 0.77 from baseline to week 8 and 0.60 from week 4 to week 12. Considering all data together, the MID of the FCSI-9 is estimated to be in the range of 1.5–3.0 points.

Conclusion. Results provide preliminary evidence of the reliability, validity, and responsiveness of the FCSI-9. *The Oncologist* 2010;15:308–316

Correspondence: Hilary H. Colwell, M.P.H., ICON Clinical Research, Lifecycle Sciences Group, 188 Embarcadero, Suite 200, San Francisco, California 94105, USA. Telephone: 415-371-2100; Fax: 415-856-0840; e-mail: hilary.colwell@iconplc.com Received February 4, 2009; accepted for publication January 18, 2010; first published online in *The Oncologist Express* on February 26, 2010. ©AlphaMed Press 1083-7159/2010/\$30.00/0 doi: 10.1634/theoncologist.2009-0034

INTRODUCTION

In the U.S., colorectal cancer is the third most common cancer in both men and women [1]. It was estimated that >106,100 cases of colon and 40,870 cases of rectal cancer would be diagnosed in 2009, and >49,920 deaths attributed to colorectal cancer were expected in the U.S. [1]. Approximately 19% of all newly diagnosed cases are diagnosed at a distant stage [2]. The 1- and 5-year survival rates of persons diagnosed with colorectal cancer are 83% and 64%, respectively [1]. The 5-year survival rates decrease with progression to metastatic disease, ranging from 68% for cancers that have spread to nearby organs to 11% for distant metastases [1]. However, with the introduction of newer cytotoxic chemotherapies and biological agents for the treatment of metastatic colorectal cancer (mCRC), the median survival time has increased from 10 months to >20 months within just a few years [3].

Patient-reported outcomes (PROs), such as health-related quality of life (HRQoL) and disease-related symptoms, are critical components for evaluating treatment effects from the patient's perspective [4]. There are several colorectal cancer-specific HRQoL measures available and widely used, including the Functional Assessment of Cancer Therapy-Colorectal Cancer (FACT-C) [5] and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C38) [6]. The nine-item FACT/National Comprehensive Cancer Network-Colorectal Symptom Index (FCSI-9) was derived from the FACT-C to be a briefer, more targeted symptom-specific measure [7]. Given its brevity, it can easily be administered with minimal patient burden during clinical studies or as part of routine care to evaluate symptoms of colorectal cancer.

The objective of this paper is to assess the reliability, validity, responsiveness, and minimum important difference (MID) of the FCSI-9 in subjects with mCRC.

MATERIALS AND METHODS

A phase III, open-label, randomized, multicenter clinical trial was conducted to evaluate the efficacy and safety of panitumumab plus best supportive care (BSC) compared with BSC alone in mCRC patients with documented disease progression after treatment with fluoropyrimidine-, irinotecan-, and oxaliplatin-containing chemotherapy regimens. Subjects who met the following criteria were eligible: pathologic diagnosis of colorectal adenocarcinoma; physician-rated Eastern Cooperative Oncology Group (ECOG) performance status (PS) score of 0, 1, or 2 (a detailed description of the ECOG PS score is included in Table 1) [8];

documented evidence of disease progression during or following treatment for mCRC; age ≥ 18 years; tumor expressing epidermal growth factor receptor by immunohistochemistry ($\geq 1\%$); and have received at least two, but no more than three, prior chemotherapy regimens for mCRC. Panitumumab was administered every 2 weeks. This study received institutional review board approval and all subjects provided informed consent.

PRO Questionnaire

Subjects completed the FCSI-9 and the EQ-5D at baseline (week 1, prior to administration of panitumumab), every 2 or 4 weeks during the study, and at the safety follow-up visit at least 4 weeks after the last treatment administration (for the panitumumab plus BSC group) or at any time within 4 weeks upon withdrawal (for the BSC alone group). A description of these questionnaires is included in Table 1. The FCSI-9 was included to assess symptoms of colorectal cancer, whereas the EQ-5D was included to assess subjects' overall HRQoL and to calculate quality-adjusted life years (QALYs). For this particular analysis, the EQ-5D was used primarily to assess the construct validity of the FCSI-9. The FCSI-9 was derived from the FACT-C as a set of brief, clinically relevant, colorectal cancer symptoms for assessing symptomatic response. It comprises the most important symptoms associated with colorectal cancer, including energy, pain, weight, diarrhea, nausea, swelling or cramps in the stomach area, appetite, ability to enjoy life, and overall quality of life [7]. (The item content of the FACT-C was determined based on interviews with patients and experienced oncologists [9]. These cognitive debriefing interviews on each item of the FACT-C provide assurance that its derivative, the FCSI, contains a set of questions that are clear and consistent with their intended meaning [9].) The FCSI-9 raw score ranges from 0 to 36 (0, severely symptomatic on all symptoms assessed; 36, symptom free on all symptoms assessed). The FCSI-9 score was calculated for each subject if at least half of the items within the scale were nonmissing (i.e., at least five items were completed) [7]. No attempt was made to impute responses to items when more than half of the items were missing (only one patient reported missing five or more items on the questionnaire). The EQ-5D visual analog scale (VAS) ranges from 0 to 100. The EQ-5D Index Score was scored according to published guidelines, using the U.K. valuation set [10, 11]. Although the resulting EQ-5D Index Score could range from -0.594 to 1.0 [10, 11], for this study, no subject scored <0.0 . For all measures, a higher value represents better functioning or well-being/fewer symptoms.

Table 1. Overview of the study assessments

Measure	Administration	Description/rationale for inclusion	Source	Administration schedule
ECOG PS score	Physician reported	<ul style="list-style-type: none"> ● These scales and criteria are used by doctors and researchers to assess how a patient's disease is progressing, assess how the disease affects the daily living abilities of the patient, and determine appropriate treatment and prognosis ● Helps to determine patient fitness for therapy ● Is widely used to quantify general functional status ● In many solid tumors is associated with disease prognosis ● Categorizes patients as: <ul style="list-style-type: none"> 0–Fully active, able to carry on all predisease performance without restriction 1–Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, i.e., light housework or office work 2–Ambulatory and capable of all self-care but unable to carry out any work activities; up and about >50% of waking hours 3–Capable of only limited self-care; confined to bed or chair >50% of waking hours 4–Completely disabled; cannot carry on any self-care; totally confined to bed or chair 5–Dead 	The ECOG PS is in the public domain [8]	Baseline (wk 1), every 2 wks during the first 8 wks of treatment, monthly thereafter until disease progression, and at the safety follow-up visit (4 wks after the last assigned treatment) or upon withdrawal
NCCN/FACT Colorectal Symptom Index (FCSI-9)	Patient self-administered	<ul style="list-style-type: none"> ● Nine-item disease-specific questionnaire ● Assesses common symptoms of colorectal cancer (e.g., energy, pain, weight loss, nausea, diarrhea, swelling or cramps in the stomach area, appetite, ability to enjoy life, and overall QoL) ● Each item has a 5 response category ranging from 0 (not at all) to 4 (very much) ● The questionnaire is scored to produce a single total score (FCSI); as scored, a higher score indicates less symptomatology 	Subset of the FACT-C questionnaire [5]	Baseline (wk 1), every 2 wks during the first 8 wks of treatment, monthly thereafter until disease progression, and at the safety follow-up visit (4 wks after the last assigned treatment) or upon withdrawal
EQ-5D	Patient self-administered	<ul style="list-style-type: none"> ● Measures overall HRQoL ● Contains 5 dimensions assessing mobility, self-care, usual activities, pain/discomfort, and anxiety/depression (3 response categories) [10] ● Scores from each dimension yield an overall weighted health state index based on values derived from general population samples [10], ranging from 0 (death) to 1 (full health) ● Also includes a 0–100 VAS (EQ-5D VAS) assessing current health state, with 0 representing worst imaginable health state and 100 representing best imaginable health state [10] 	All components of the EQ-5D were administered [12]	Baseline (wk 1), monthly thereafter until disease progression, and at the safety follow-up visit or upon withdrawal

Abbreviations: ECOG, Eastern Cooperative Oncology Group; FACT-C, Functional Assessment of Cancer Therapy–Colorectal Cancer; HRQoL, health-related quality of life; NCCN, National Comprehensive Cancer Network; PS, performance status; QoL, quality of life; VAS, visual analog scale.

Statistical Methods

All subjects, regardless of treatment, who completed the baseline and at least one follow-up assessment were included in the validation analysis. To the extent possible, the U.S. Food and Drug Administration's PRO guidance document was followed for evaluating the psychometric properties that should be established for a PRO questionnaire to support labeling claims or for promotional purposes [4].

Subject Demographic and Clinical Characteristics

Demographic characteristics were described by computing frequency distributions of the categorical variables for each characteristic.

Variability

We examined the variability of the FCSI-9 individual items and overall score at each assessment. Floor and ceiling effects were assessed by examining the percentage of subjects who had the lowest and the highest possible scale scores across assessments.

Reliability

Internal consistency reliability and test-retest reliability were measured. Internal consistency, the extent to which items or scales are measuring the same concept, was calculated using Cronbach's α [12, 13]. A Cronbach's α was computed using data from each assessment and also using data pooled from all assessments. An α coefficient ≥ 0.70 was considered to represent acceptable reliability [14].

Test-retest reliability, the extent to which a measure yields consistent scores over a short period of time (assuming there is no underlying change), was evaluated using the intraclass correlation coefficient (ICC). An ICC ≥ 0.70 was considered acceptable [12, 14]. The ICC for the FCSI-9 was computed using two approaches for the week 4 to week 8 interval as follows: (a) including subjects whose EQ-5D Index Score remained exactly the same and (b) including subjects whose physicians rated their ECOG PS score as the same.

Validity

Construct validity was determined by comparing correlations of the FCSI-9 and EQ-5D Index and VAS score using Pearson correlation coefficients. Convergent validity is demonstrated when scales or items thought to measure the same construct have high correlation coefficients, and divergent validity is demonstrated when items or scales thought to measure different constructs have low correlation coefficients. The directionality and strength of correla-

tions were evaluated pooling data from all assessments. Pearson correlations >0.70 were considered strong, values of $0.40-0.70$ were considered moderate, values of $0.20-0.40$ were considered weak, and values <0.20 were considered zero order [15].

Known-groups validity evaluates the ability of the measure to discriminate between groups known to be clinically different. Known-groups validity was assessed by comparing the FCSI-9 scores of subjects rated with an ECOG PS score of 0, 1, or 2. ECOG PS is commonly used to quantify general functional status ranging from fully ambulatory without symptoms to bedridden. It helps determine patient fitness for therapy and in many solid tumors is associated with disease prognosis [8]. Our hypothesis was that subjects with a better performance status (i.e., lower ECOG PS rating) would report fewer symptoms on the FCSI-9 than those with a worse performance status (higher ECOG PS rating). These analyses used the two-sample *t*-test and included data from the baseline and week 8 assessments.

Responsiveness

Responsiveness, which is how effectively a measure detects change over time, was evaluated based on the paired *t*-test and Guyatt's statistic over two 8-week time intervals (between baseline and week 8 and between week 4 and week 12). A paired *t*-test analysis was conducted on the subset of subjects whose ECOG PS score worsened between the two time points in each 8-week interval. For each time interval, FCSI-9 change scores were tested against zero using the one-sample *t*-test. Responsiveness was demonstrated if the change score was statistically significantly less than zero.

In addition, responsiveness was evaluated using Guyatt's statistic. For each 8-week interval, the numerator of the Guyatt's statistic was the mean change in FCSI-9 score in the corresponding 8-week interval for the cohort of subjects who experienced change (i.e., worsening ECOG PS rating). The denominator was the standard deviation (SD) of the change in score in the time interval for the clinically stable group (i.e., those with no change in ECOG PS rating). A Guyatt's statistic with an absolute value of 0.20 was considered minimally adequate, whereas an absolute value >1.00 indicates a highly responsive measure [16]. It was assumed that those who worsened on the FCSI-9 would have a change in score less than or equal to negative one (-1), those who improved on the FCSI-9 would have a change greater than or equal to one ($+1$), and those who remained stable on FCSI-9 would have a change in score greater than negative one (-1) but less than one ($+1$).

MID

The MID is the smallest difference in score on the FCSI-9 that would be considered clinically meaningful [17]. The MID is typically used as a benchmarking tool to assist in interpreting differences between treatment arms. To compute MID estimates, we used an anchor-based method, which maps the changes in the measure of interest to the changes in a clinically relevant independent standard or anchor that is itself interpretable and at least moderately correlated with changes in the measure.

The ECOG PS score served as the anchor for measuring MID in this population. The MID was calculated using the β coefficients from an analysis of variance model in which the dependent variable was the FCSI-9 change score from baseline to week 8 and the independent variable was the change in ECOG PS score between baseline and week 8. This model was run once for the FCSI-9, and the MID was calculated as the β coefficient. Because a smaller reduction in ECOG PS may indicate deterioration, we also computed the β coefficient for a half-grade change in ECOG PS score. As suggested by Wiebe et al. [18], we computed r^2 to assess the strength of the relationship between the ECOG PS score and the FCSI-9 to confirm that an association exists between the measure and the anchor.

All statistical analyses were generated using SAS® software (version 9.1.3, SAS System for Windows; SAS Institute, Inc., Cary, NC).

RESULTS

Subject Demographic and Clinical Characteristics

Table 2 shows the demographic and clinical characteristics of the 391 subjects that provided consent before completing the assessments and completed the baseline and at least one follow-up assessment (the total clinical trial population included 463 subjects, so the validation cohort represents 84% of the total trial population. With the exception of gender, $p < .05$, there were no statistically significant differences in demographics between the clinical trial population and the validation cohort). The 72 individuals who were excluded provided consent after completing the baseline assessment and were therefore excluded. The majority of the validation cohort was male (66%), white (99%), <65 years of age (60%), and from western Europe (76%). Most of the subjects were rated by their physician with an ECOG PS score of 0 or 1 (88%), indicating that most subjects did not have considerable physical restrictions at baseline. Physicians rated 42% as fully active, with no restrictions (ECOG PS score, 0), whereas 47% had limitations in physically

Table 2. Demographic and clinical characteristics

Characteristic	n = 391, n (%)
Gender	
Female	135 (34.5)
Male	256 (65.5)
Race	
White	386 (98.7)
Asian	2 (0.5)
Hispanic or Latino	2 (0.5)
Japanese	1 (0.3)
Age	
<65	233 (59.6)
≥65	158 (40.4)
≥75	37 (9.5)
Region	
Western Europe	296 (75.7)
Central and eastern Europe	35 (9.0)
Rest of world	60 (15.3)
Baseline ECOG PS score	
0	163 (41.7)
1	182 (46.5)
2	45 (11.5)
3 ^a	1 (< 1%)
Primary diagnosis	
Colon	260 (66.5)
Rectum	131 (33.5)

^aOne patient with a PS score of 3 was admitted to the study as a protocol deviation, and was included in all analyses, with the exception of the known-groups analysis.
Abbreviation: ECOG PS, Eastern Cooperative Oncology Group performance status.

strenuous activities (ECOG PS score, 1). Only 12% reported significant impairments in daily activities (ECOG PS score ≥2). The majority of subjects (67%) reported a primary diagnosis of mCRC in the colon (as opposed to the rectum). Finally, approximately 300, 150, and 100 patients provided data at week 4, week 8, and week 12, respectively.

Variability

In general, the FCSI-9 scale and individual items demonstrated good variability. The mean (SD) values of the baseline, week 4, week 8, and week 12 FCSI-9 were 25.8 (5.1), 26.0 (5.1), 27.2 (5.2), and 27.6 (5.3), respectively. Less than 1% of subjects had scale scores at the ceiling (data not shown) and none of the subjects had scale scores at the floor across all visits.

Table 3. Construct validity (interscale Pearson correlations)

PRO scales	FCSI-9
EQ-5D Index Score	0.68
EQ-5D VAS	0.64

All available assessments were used to assess construct validity.
Abbreviations: FCSI, Functional Assessment of Cancer Therapy Colorectal Symptom Index; PRO, patient-reported outcome; VAS, visual analog scale.

Internal Consistency Reliability

Internal consistency reliability of the FCSI-9, as measured by Cronbach's α coefficient, was acceptable for baseline through week 25 (i.e., baseline [week 1] and weeks 3, 5, 7, 9, 13, 17, 21, and 25; range, 0.75–0.84) and when pooled over all assessments (0.805).

Test–Retest Reliability

Fifty subjects' (12.8%) health status remained stable on the EQ-5D Index Score between week 4 and week 8, and 112 subjects (28.6%) remained stable in ECOG PS score between week 4 and week 8. The ICC for the FCSI-9, based on stable EQ-5D Index Scores, was 0.76. Among subjects who had stable ECOG PS scores, the ICC for the FCSI-9 was also 0.76.

CONSTRUCT VALIDITY

Construct validity, as measured by interscale Pearson correlations, was moderate, as demonstrated by the relationship between the FCSI-9 and the EQ-5D (Table 3). The FCSI-9's convergent validity is demonstrated through the moderate correlation of the FCSI-9 with both the EQ-5D Index Score ($r = 0.68$) and the EQ-5D VAS ($r = 0.64$).

Known-Groups Validity

Subjects were categorized into three groups according to ECOG PS score. At baseline, subjects with a better performance status ($n = 157$; ECOG PS score, 0) reported better functioning/fewer symptoms on the FCSI-9 than those with a worse performance status ($n = 175$; ECOG PS score, 1; $n = 43$, ECOG PS score, 2)— 28.2 ± 3.91 for ECOG PS score 0 versus 25.0 ± 4.66 for ECOG PS score 1 versus 20.8 ± 5.58 for ECOG PS score 2; overall $p < .001$) (Table 4). At week 8, subjects with an ECOG PS score of 0 or 1 reported better FCSI-9 scores than subjects with an ECOG PS score of 2 (30.0 ± 4.00 for ECOG PS score 0 versus 26.3 ± 4.33 for ECOG PS score 1 versus 20.3 ± 6.24 for ECOG PS score 2; overall $p < .001$).

Responsiveness

Table 5A and 5B describe the responsiveness of the FCSI-9 using different statistics and time intervals. Between baseline and week 8, 27 (6.9%) subjects had a decline in ECOG PS score of at least one grade. FCSI-9 scores worsened over this interval. A one-sample t -test indicated that this difference was statistically significant (26.1 ± 5.32 at baseline versus 23.9 ± 5.86 at week 8; $p = .008$).

Table 5B shows the responsiveness from week 4 to week 12. A total of 22 subjects had a decline in ECOG PS score by at least one grade between week 4 and week 12. Those subjects also had a decrease in FCSI-9 score during this interval, which was significant (28.0 ± 4.20 at week 4 versus 25.3 ± 6.73 at week 12; $p = .040$).

Guyatt's statistic results are also presented for the two intervals, from baseline to week 8 (Table 5A) and from week 4 to week 12 (Table 5B). That analysis included subjects who worsened by at least one ECOG PS grade over each 8-week interval. All scales were sensitive to change using either 8-week interval, because the absolute value of

Table 4. Known-groups validity: mean scale score by ECOG PS

PRO scale ^b	ECOG PS = 0 ^a		ECOG PS = 1 ^a		ECOG PS = 2 ^a		<i>p</i> -value ^c
	<i>n</i>	Mean (SD)	<i>n</i>	Mean (SD) ^c	<i>n</i>	Mean (SD)	
FCSI-9 at Baseline ^d	157	28.2 (3.91)	175	25.0 (4.66)	43	20.8 (5.58)	<.001
FCSI-9 at wk 8 ^e	68	30.0 (4.00)	57	26.3 (4.33)	9	20.3 (6.24)	<.001

^aSubjects categorized by ECOG PS (0 = Fully active, able to carry on all predisease performance without restriction, 1 = Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, i.e., light housework or office work, 2 = Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about > 50% of waking hours).

^bThe FCSI-9 raw score ranges from 0–36, where a higher score represents fewer symptoms.

^c*p*-value from *F*-test evaluating the overall difference in mean scale score by ECOG PS score.

^dFCSI-9 mean (SD) classified by ECOG PS score at baseline.

^eFCSI-9 mean (SD) classified by ECOG PS score at week 8.

Abbreviations: ECOG PS, Eastern Cooperative Oncology Group performance status; FCSI, Functional Assessment of Cancer Therapy Colorectal Symptom Index; PRO, patient-reported outcome; SD, standard deviation.

Table 5. Responsiveness: Change scores and Guyatt's statistic from baseline to week 8^a

A. Baseline to week 8							
PRO scale^c	Baseline <i>n</i> = 27 Mean (SD)	Week 8 <i>n</i> = 24 Mean (SD)	Change score <i>n</i> = 24 Mean (SD)	<i>p</i>-value^b	Numerator <i>n</i> = 24 Mean of the change cohort	Denominator <i>n</i> = 90 SD of the stable cohort	Guyatt's statistic
FCSI-9	26.1 (5.32)	23.9 (5.86)	−2.8 (4.76)	0.008	−2.8	3.70	−0.767
B. Week 4 to week 12							
PRO scale^c	Week 4 <i>n</i> = 21 Mean (SD)	Week 12 <i>n</i> = 21 Mean (SD)	Change score <i>n</i> = 20 Mean (SD)	<i>p</i>-value^b	Numerator <i>n</i> = 20 Mean of the change cohort	Denominator <i>n</i> = 63 SD of the stable cohort	Guyatt's Statistic
FCSI-9	28.0 (4.20)	25.3 (6.73)	−2.3 (4.58)	0.040	−2.3	3.76	−0.599

^aThe numerator is the mean of the scale for the subset of subjects whose ECOG PS worsens score by at least one grade between baseline and week 8 (A) or between week 4 and week 12 (B). The denominator is the standard deviation of the scale for the subset of subjects whose ECOG PS score remains stable between baseline and week 8 (A) or between week 4 and week 12 (B).

^b*p*-value from one-sample *t*-test.

^cThe FCSI-9 raw score ranges from 0 to 36, where a higher score represents fewer symptoms.

Abbreviations: ECOG PS, Eastern Cooperative Oncology Group performance status; FCSI, Functional Assessment of Cancer Therapy Colorectal Symptom Index; PRO, patient-reported outcome; SD, standard deviation.

the Guyatt's statistics for the FCSI-9 was 0.77 from baseline to week 8 and 0.60 for the interval from week 4 to week 12.

MID

Using a one-unit change in ECOG PS score as the anchor, the MID for the FCSI-9 raw score was 2.83. However, it is expected that, from the patient's perspective, a one-unit increase in ECOG PS score over an 8-week period is a significant clinical change, not a minimal one. Therefore, a one-half-unit change in ECOG PS score may be a more appropriate anchor for detecting a minimal change. Using the one-half-unit change in ECOG PS score as the anchor, the MID for the FCSI-9 was 1.42. The ECOG PS score was moderately correlated with the FCSI-9 ($r^2 = 0.15$; $r = 0.39$).

DISCUSSION

The objective of this study was to evaluate the psychometric properties of the FCSI-9 administered to mCRC patients in a phase III clinical trial. The results provide preliminary evidence to suggest that the FCSI-9, a brief, symptom-specific measure, is reliable, valid, and responsive in this population.

The results indicate that the FCSI-9 has acceptable internal consistency and test-retest reliability. The results of the construct validity testing demonstrated convergent validity. One of the limitations of the evaluation of construct

validity was that the EQ-5D was the only available instrument to conduct this analysis. Ideally, in the future, construct validity should be evaluated using another well-validated colorectal cancer-specific questionnaire.

The FCSI-9 is able to distinguish between subjects categorized according to their ECOG PS score. Subjects with an ECOG PS score of 2 reported lower scores on the FCSI-9 at baseline and at week 8 than subjects with an ECOG PS score of 0 or 1.

The FCSI-9 was responsive during the first assessment period (e.g., baseline to week 8) whereas the sensitivity to change was slightly lower from week 4 to week 12. Similarly, using Guyatt's statistic to evaluate responsiveness, the FCSI-9 showed a slightly lower sensitivity from week 4 to week 12, compared with baseline to week 8. This is not altogether surprising and may be a result of regression to the mean. In addition, these results should be interpreted cautiously given the small sample sizes.

The MID analysis indicated that the FCSI-9, a symptom-specific scale, required an MID of approximately 1.5–3.0 points. Therefore, a change of approximately 1.5–3.0 on the FCSI-9 would be considered clinically important. However, the correlation ($r = 0.39$) of the FCSI-9 scale with the ECOG PS score indicates that the change in clinician-rated ECOG PS is not a strong predictor of change. The narrow ECOG PS range allowed in this study (0, 1, or 2) may contribute to the fact that the change in ECOG PS score explained a smaller proportion of the variance in the change.

Further work should consider measures other than the ECOG PS score to anchor MID calculations.

The data collected from the clinical trial allowed for the evaluation of the psychometric properties of the FCSI-9. However, a number of limitations arise when using clinical trial data for this purpose. For instance, the study population was fairly homogenous, composed primarily of white subjects from western Europe. Additionally, some analyses were limited by the number of subjects included in the analyses. We were also limited by the clinical measure used for calculating responsiveness. A clinical measure more sensitive than the ECOG PS score that allows for more variability may provide better resolution for detecting the responsiveness of the FCSI-9. Finally, only data from subjects whose clinical condition worsened were used to compute responsiveness and the MID. Ideally, these calculations would have included subjects who worsened and subjects who improved. Therefore, the MID that we obtained of 1.5–3.0 points corresponds only to those subjects who worsened. It is typical to calculate the MID for the group that improves as well, and the MID may differ for the improved and worsened groups. Unfortunately, given the subject population (i.e., mCRC patients), we did not have an adequate sample size to calculate the MID for an improved group.

Given the sound psychometric properties of this instrument, we recommend that it be used to assess colorectal cancer symptoms in future studies. Given its length, it can be easily implemented in studies with minimal patient burden. The questionnaire is scored to produce a single summary score, and therefore comparisons among different treatments or across studies can be easily accomplished. In addition to administering the FCSI-9, we suggest also considering the EQ-5D, particularly when it is important to provide an overall picture of the patient's health and to calculate QALYs for a cost-utility analysis.

CONCLUSION

The primary goal of this study was to present our findings on the psychometric properties of FCSI-9 in a population of mCRC subjects. The results show preliminary evidence of the questionnaire's reliability, validity, and responsiveness, and an MID estimate was provided. Because the psychometric properties of PROs should be based on accumulating evidence, we encourage others to confirm and expand on our findings in future studies.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of Kimberly Miller, Gladys Tom, and Jennifer Welle of ICON Clinical Research in preparing this manuscript; and Carolyn Roberts for administrative support.

Financial support for this study was provided by Amgen, Inc. Nicola Wright and John Lu are employees of Amgen. Giovanna Devercelli was an employee of Amgen, Inc. at the time of the study. Hilary Colwell and Michelle Turner are employees of ICON Clinical Research, a contract research organization, which received financial support for undertaking this study. Susan Mathias was an employee of ICON Clinical Research at the time of the study. Marc Peeters was a principal investigator in this study and received financial support for his participation. David Cella is the developer of the FCSI and did not receive financial support for his participation in this work.

AUTHOR CONTRIBUTIONS

Conception/Design: Hilary Colwell, Susan Mathias, John Lu; Giovanna Devercelli
Financial support: John Lu, Giovanna Devercelli, Nicola Wright
Provision of study material or patients: Giovanna Devercelli, Marc Peeters
Collection and/or assembly of data: John Lu, Giovanna Devercelli; Nicola Wright; Marc Peeters
Data analysis and interpretation: Hilary Colwell, Susan Mathias, Michelle Turner, Nicola Wright, David Cella; John Lu
Manuscript writing: Hilary Colwell, Susan Mathias, Michelle Turner, John Lu, Giovanna Devercelli, David Cella
Final approval of manuscript: Hilary Colwell, Susan Mathias, Michelle Turner, John Lu, Giovanna Devercelli, Nicola Wright, Marc Peeters, David Cella

REFERENCES

- 1 American Cancer Society. American Cancer Society, Cancer Facts and Figures, 2009. Available at <http://www.cancer.org/downloads/STT/500809web.pdf>, accessed February 5, 2010.
- 2 Cancer Statistics Branch, National Cancer Institute, DCCPS Surveillance Research Program. Surveillance Epidemiology and End Results (SEER) Program (<http://www.seer.cancer.gov>). SEER*Stat Database: Incidence—SEER 17 Regs Public-Use, November 2005, Sub (1973–2003 varying), 2006 based on the November 2005 submission. Available at <http://seer.cancer.gov/statfacts/html/colorect.html>, accessed February 5, 2010.
- 3 Meulenbeld HJ, Creemers GJ. First-line treatment strategies for elderly patients with metastatic colorectal cancer. *Drugs Aging* 2007;24:223–238.
- 4 U.S. Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. FDA, 2009. Available at <http://www.fda.gov/ohrms/dockets/dockets/06d0044/06D-0044-EC30-Attach-1.pdf>, accessed February 5, 2010.
- 5 Ward WL, Hahn EA, Mo F et al. Reliability and validity of the Functional Assessment of Cancer Therapy–Colorectal (FACT-C) quality of life instrument. *Qual Life Res* 1999;8:181–195.
- 6 Sprangers MA, te Velde A, Aaronson NK. The construction and testing of the EORTC colorectal cancer-specific quality of life questionnaire module (QLQ-CR38). European Organization for Research and Treatment of Cancer Study Group on Quality of Life. *Eur J Cancer* 1999;35:238–247.
- 7 Cella D, Paul D, Yount S et al. What are the most important symptom targets when treating advanced cancer? A survey of providers in the National

- Comprehensive Cancer Network (NCCN). *Cancer Invest* 2003;21:526–535.
- 8 Oken MM, Creech RH, Tormey DC et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982;5:649–655.
 - 9 Cella DF, Tulsky DS, Gray G et al. The Functional Assessment of Cancer Therapy scale: Development and validation of the general measure. *J Clin Oncol* 1993;11:570–579.
 - 10 Dolan G, Kind P, Williams A. A Social Tariff for the EuroQol: Results from a UK General Population Survey. Centre for Health Economics Discussion Paper. Heslington, York, UK: Centre for Health Economics, University of York, 1995:138.
 - 11 EQ-5D: A measure of health-related quality of life. User Guide. The Netherlands: The Euroqol Group, 2000;8:1–44.
 - 12 Lohr KN, Aaronson NK, Alonso J et al. Evaluating quality-of-life and health status instruments: Development of scientific review criteria. *Clin Ther* 1996;18:979–992.
 - 13 Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–333.
 - 14 Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res* 2002;11:193–205.
 - 15 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press, 1977:1–531.
 - 16 Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171–178.
 - 17 Schünemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: The clinician's perspective. *Health Qual Life Outcomes* 2006;4:62.
 - 18 Wiebe S, Matijevic S, Eliasziw M et al. Clinically important change in quality of life in epilepsy. *J Neurol Neurosurg Psychiatry* 2002;73:116–120.