NATIONAL QUALITY FORUM

Measure Evaluation 4.1
January 2010

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments and will appear if your cursor is over the highlighted area (or in the margin if your Word program is set to show revisions in balloons). Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** OT2-019-09 **NQF Project:** Patient Outcomes
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Functional Assessment of Cancer Therapy-General Version (FACT-G)
De.2 **Brief description of measure***:* The FACIT Measurement System is a collection of QOL questionnaires targeted to the management of chronic illness. "FACIT" (Functional Assessment of Chronic Illness Therapy) was adopted as the formal name of the measurement system in 1997 to portray the expansion of the more familiar "FACT" (Functional Assessment of Cancer Therapy) series of questionnaires into other chronic illnesses and conditions. Thus, FACIT is a broader, more encompassing term that includes the FACT questionnaires under its umbrella. The measurement system, under development since 1987, began with the creation of a generic CORE questionnaire called the Functional Assessment of Cancer Therapy-General (FACT-G). The FACT-G (now in Version 4) is a 27-item compilation of general questions divided into four primary QOL domains: Physical Well-Being, Social/Family Well-Being, Emotional Well-Being, and Functional Well-Being. It is considered appropriate for use with patients with any form of cancer, and has also been used and validated in other chronic illness conditions (e.g., HIV/AIDS and multiple sclerosis) and in the general population (using a slightly modified version).
1.1-2 **Type of Measure:** Outcome De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** «composite_paired»
De.4 **National Priority Partners Priority Area:** Population health, Palliative and End of Life De.5 **IOM Quality Domain**: patient-centered De.6 **Consumer Care Need:** Living with illness

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed.	**A**

Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available. **A.1 Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes **A.2 Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** proprietary measure **A.3 Measure Steward Agreement:** agreement signed and submitted **A.4 Measure Steward Agreement attached:** «agreement_attach»	**Y☐** **N☐**
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** **Y☐** **N☐**
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ►**Purpose:** public reporting, quality improvement	**C** **Y☐** **N☐**
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 24 months of endorsement. **D.1Testing:** Yes, fully developed and tested **D.2 Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** **Y☐** **N☐**
(for NQF staff use**)** **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** **Y☐** **N☐**
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	**Eval Rating**
(for NQF staff use) **Specific NPP goal**:	
1a.1 Demonstrated High Impact Aspect of Healthcare: Yes **1a.2** affects large numbers, a leading cause of morbidity/mortality, severity of illness, patient/societal consequences of poor quality, frequently performed procedure, high resource use **1a.3 Summary of Evidence of High Impact: Cancer and chronic illnesses are widely cited in medical literature as high resource diseases which impact large numbers of people.** **1a.4 Citations for Evidence of High Impact:** **Cella D., et al. The Functional Assessment of Cancer Therapy Scale: Development and validation of the General measure. Journal of Clinical Oncology (1993) Vol (11): 570-579**	**1a** **C☐** **P☐** **M☐** **N☐**
1b. Opportunity for Improvement	**1b** **C☐**

1b.1 Benefits (improvements in quality) envisioned by use of this measure: «improvement_benefits»

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
n/a

1b.3 Citations for data on performance gap:
n/a

1b.4 Summary of Data on disparities by population group:
The disparities in cancer care by population group is widely published. In fact the NIH and AHRQ have whole funding initiatives specifically identified to address these issues. Dr. Cella's group has considerable expertise in researching and measuring response differences across groups, including publications on literacy and cross cultural assessment issues.

1b.5 Citations for data on Disparities:
ISOQOL Article of the Year 2007, Hahn, E., et al The impact of literacy on health-related quality of life measurement and outcomes in cancer outpatients. Quality of Life Research, 16(3), 495-507.

Hahn E, Cella D. Health outcomes assessment in vulnerable populations: measurement challenges and recommendations. Archives of Physical Medicine and Rehabilitation 2003; 84(Suppl 2):S35-S42.

P☐ M☐ N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** FACT-G (now version 4) is a 27-item compilation of general questions divided into 4 primary QOL domains: physical well-being, social/family well-being, emotional wellbeing, and functional well-being. As described previously and in the literature citations, it provides a multidisciplinary measure of a patient's well being including emotional and physical. It is a measure that's responsive to change over time. It can also be used to measure response to treatment (it measures change to clinical status). It can be used to demonstrate that a change in treatment plan is warranted, and it can be used to demonstrate the effectiveness of palliative care.

Item content was determined by combined expert and cancer patient input, ensuring that clinically important issues relevant to patients were included. Content validity has been ensured by use of a rigorous, peer-reviewed procedure for determining the relevance and relative importance of each of the many issues raised by cancer patients as having a bearing upon their HRQOL. There are over 25 published reports detailing its performance. Thus, there is a solid reference literature to which one can compare results.

1c.2-3. Type of Evidence:
cohort study, evidence based guideline, expert opinion, observational study, randomized controlled trial, systematic synthesis of research

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
«outcomes_relationship_evidence_summary»

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
See answer to 1c.4

1c.6 Method for rating evidence: See answer to 1c.4

1c.7 Summary of Controversy/Contradictory Evidence: See answer to 1c.4

1c
C☐ P☐ M☐ N☐

1c.8 Citations for Evidence (*other than guidelines***):** «evidence_guidelines_other» **1c.9 Quote the Specific guideline recommendation (***including guideline number and/or page number***):** «guideline_quote» **1c.10 Clinical Practice Guideline Citation:** See answer to 1c.9 **1c.11 National Guideline Clearinghouse or other URL:** n/a **1c.12 Rating of strength of recommendation (***also provide narrative description of the rating and by whom***):** n/a **1c.13 Method for rating strength of recommendation** (*If different from USPSTF system, also describe rating and how it relates to USPSTF)***:** n/a **1c.14 Rationale for using this guideline over others:** This questionnaire is widely used by clinicians, clinical trialists and cooperative group trials. It has been well validated and is in widespread use for assessing cancer patient's QOL and in clinical decision-making. It is available in over 45 languages.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Eval Rating**
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	**2a-specs** **C☐** **P☐** **M☐** **N☐**
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** Individual items ask cancer patients about how true certain symptoms have been for them. The composite score of all the items gives a Quality of Life (QOL) score which can be used by clinicians and in clinical trials to determine certain clinical indicators. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** Respondents are requested to look back on the previous 7 days. **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** The FACT-G scores are individual respondent scores. Responses are "Not at All", A Little Bit", "Somewhat", "Quite a Bit", and "Very Much". Each items is scored as being either a positive or negative item, depending on if the response would be positive or negative to the patient's quality of life (ie, "I feel tightness in my chest" Response: Very much isc considered a negatively scored items vs. "My thinking is clear" Response: Very muchh is considered a positively worded item.	
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being*	

measured):
n/a

2a.5 **Target population gender:** Female, Male
2a.6 **Target population age range:** pediatric and adult patients

2a.7 **Denominator Time Window (***The time period in which cases are eligible for inclusion in the denominator***):**
n/a

2a.8 **Denominator Details (***All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
n/a

2a.9 **Denominator Exclusions (***Brief text description of exclusions from the target population***):** n/a

2a.10 **Denominator Exclusion Details (***All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
n/a

2a.11 **Stratification Details/Variables (***All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
n/a

2a.12-13 **Risk Adjustment Type:** no risk adjustment necessary

2a.14 **Risk Adjustment Methodology/Variables (***List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
n/a

2a.15-17 **Detailed risk model available Web page URL or attachment:**

2a.18-19 **Type of Score:** Other (specify) Individuals items are scored so that a higher score represents a better QOL. Total (summed) scores are also calculated and higher scores represent better QOL.
2a.20 **Interpretation of Score:** Better quality = higher score
2a.21 **Calculation Algorithm (***Describe the calculation of the measure as a flowchart or series of steps***):**
FACT-G Scoring Guidelines (Version 4)

Instructions:* 1. Record answers in "item response" column. If missing, mark with an X
2. Perform reversals as indicated, and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.
4. Add subscale scores to derive total FACT-G score. ***The higher the score, the better the QOL.***

Subscale	Item Code	Reverse item?	Item response	Item Score
PHYSICAL	GP1	4	- _____	=_____
WELL-BEING	GP2	4	- _____	=_____
(PWB)	GP3	4	- _____	=_____
score range: 0-28	GP4	4	- _____	=_____
	GP5	4	- _____	=_____
	GP6	4	- _____	=_____
	GP7	4	- _____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=**PWB subscale score**

SOCIAL/FAMILY	GS1	0	+	_____	=_____
WELL-BEING	GS2	0	+	_____	=_____
(SWB)	GS3	0	+	_____	=_____
score range: 0-28	GS4	0	+	_____	=_____
	GS5	0	+	_____	=_____
	GS6	0	+	_____	=_____
	GS7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=**SWB subscale score**

EMOTIONAL	GE1	4	-	_____	=_____
WELL-BEING	GE2	0	+	_____	=_____
(EWB)	GE3	4	-	_____	=_____
score range: 0-24	GE4	4	-	_____	=_____
	GE5	4	-	_____	=_____
	GE6	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 6: _____
Divide by number of items answered: _____=**EWB subscale score**

FUNCTIONAL	GF1	0	+	_____	=_____
WELL-BEING	GF2	0	+	_____	=_____
(FWB)	GF3	0	+	_____	=_____
score range: 0-28	GF4	0	+	_____	=_____
	GF5	0	+	_____	=_____
	GF6	0	+	_____	=_____
	GF7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=**FWB subscale score**

TOTAL SCORE: _____ + _____ + _____ + _____=_____=**FACT-G Total score**
(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score)

score range: 0-108

*For additional guidelines please refer to the Administration and Scoring Guidelines in the manual or at www.facit.org.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:
«performance_discrimination_method»

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
The sample size for the study in question is dependent on how the scale will be used. It can be used with single patients for clinical decision-making. Or it can be used for clinical trial QOL scores to be tied to a clinical response. If IRT (item response theory)/Rasch analyses will be used, that will also impact the sample size.

The sample can be any individual or group of patients being treated, or having previously been treated for cancer.

The questionnaire can be administered by RN's or research personnel directly instructing the participants,

or it can be administered electronically online or via telephone CATI (computer adaptive telephone interview). Each assessment method will impact the sample in terms of accessibility.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Registry data

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
The FACT-G questionnaire is currently being used by investigators from medical and educational institutions, industry sponsors, and cooperative clinical trial groups. Application includes use in Phase I, II, and III, clinical trials, in health practice, for symptom management, for psychological interventions, and in other cancer treatment evaluations. The FACT-G is most commonly used in the clinical trial setting, but has also been used in screening, survivorship, and end-of-life evaluations.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: Attachment Published Evidence on Reliability and Validity of the FACT.docx

2a.29-31 Data dictionary/code table web page URL or attachment: Attachment FACT-G_ENG_Final_Ver4_16Nov07.doc

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Can be measured at all levels

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Ambulatory Care: Amb Surgery Center, Ambulatory Care: Office, Ambulatory Care: Clinic, Ambulatory Care: Emergency Dept, Ambulatory Care: Hospital Outpatient, Assisted Living, Behavioral health/psychiatric unit, Dialysis Facility, Group homes, Home, Hospice, Hospital, Long term acute care hospital, nursing home (NH) /Skilled Nursing Facility (SNF), Rehabilitation Facility

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)***:** The data used in the original validation report were derived from two samples. The first sample was tested twice over a 2-month period and was therefore used for the sensitivity to change analysis. Because this sample also completed a QL instrument, the Functional Living Index-Cancer (FLIC), 2 ° data from this sample were used to assess concurrent validity as well.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)***:**
The FACT-G total score provides a useful summary of overall quality of life across a diverse group of patients.

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
The FACT-G was administered to a previously untested sample of 70 outpatients with mixed cancer diagnoses. A second administration was planned within 3 to 7 days. Timing was prearranged to avoid chemotherapy treatment between administrations of the FACT-G. Of the 70 patients who completed administration 1, 60 (86%) completed administration 2 within 3 to 7 days. Test-retest correlation coefficients for these 60 patients were as follows: physical well-being, .88; functional well-being, .84; social well-being, .82; emotional well-being, .82; relationship with doctor, .83; and total score, .92.

Sensitivity to change is an important capability of any QL instrument that is proposed to evaluate treatment- or illness-related differences in a clinical trial. To a great extent, the performance of an instrument in the field will document its sensitivity. However, it was decided to obtain an early indication

2b
C☐
P☐
M☐
N☐

of whether subtests and the overall score would fluctuate as expected in patient groups that are known to change over time. A common (albeit global) parameter of change is PSR, and it was predicted that the physical and functional subtests would show the most significant sensitivity to change in this parameter, whereas other subtests might show marginal sensitivity to the related-but-distinct concept of PSR. The FACT-G was administered to an additional previously untested sample of 104 patients currently receiving chemotherapy for advanced breast, lung, or colon cancer. A second administration occurred 2 months later. Patient-reported PSRs32 were also generated in an interview conducted before completion of the FACT. Patients were then categorized into three groups, according to change in PSR over time: those whose PSR declined (n = 27); those whose PSR improved (n = 17), and those whose PSR remained unchanged (n = 60). Multivariate analysis of variance confirmed a significant overall effect (*P* = .002), indicating that the FACT-G can clearly distinguish the three groups. Results of the follow-up Univariate tests are listed in Table 7. These indicate, as expected, that the strongest contributors to sensitivity to change in PSR were the physical (*P* < .001) and functional (*P* < .01) subscales. Also sensitive to change in PSR was the emotional subscale (*P* < .05), but not the social or relationship with doctor subscales.	
2c. Validity testing 2c.1 **Data/sample** *(description of data/sample and size)*: **Items were generated using semi-structured interview input from 45 patients with cancer and 15 oncology specialists. Patients currently receiving treatment for advanced breast (n = 15), lung (n = 15), and colorectal (n = 15) cancer were approached to participate in a brief interview, "to help develop a measure of quality of life for people with your illness." To increase the likelihood that patients in this phase would have had sufficient experience with cancer symptoms and treatment side effects, eligibility was restricted to patients currently receiving treatment (including hormone therapy) for advanced (stage III or IV) cancer. It was also required that patients be able to read and speak English and that they have no known evidence of brain metastasis, delirium, psychosis, or severe depression. All 45 patients who were approached agreed to participate. Median age of the patients was 60 years (range, 27 to 76).** 2c.2 **Analytic Method** *(type of validity & rationale, method for testing)*: 2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*: **Initial evidence for convergent and discriminant validity was evaluated using data from the 316 patients who completed the full validation packet. Convergent validity is evaluated by examining the association between scores on the FACT-G and those of other similar measures completed at the same time. Relatively high correlation coefficients are expected in these comparisons. Divergent validity is evaluated by examining the association between scores on the FACT-G and dissimilar measures completed at the same time. Low correlations are expected in these comparisons. As was expected, the Pearson correlation with the FLIC was high (.79). Also, correlations with measures of mood distress were also rather high: rFACT/TMAS = .58; rTFACT/BriefPOMS = -. 65. The correlation with activity level as measured by the ECOG 5-point PSR was moderately high (*r* = -. 56), within the expected range. Also as expected, the correlation with social desirability, as measured by the shortened M-CSDS, was rather low (*r* = .22), supporting divergent validity.**	**2c** **C☐** **P☐** **M☐** **N☐**
2d. Exclusions Justified 2d.1 **Summary of Evidence supporting exclusion(s):** n/a 2d.2 **Citations for Evidence:** 2d.3 **Data/sample** *(description of data/sample and size)*: 2d.4 **Analytic Method** *(type analysis & rationale)*: 2d.5 **Testing Results** *(e.g., frequency, variability, sensitivity analyses)*:	**2d** **C☐** **P☐** **M☐** **N☐** **NA☐**

2e. Risk Adjustment for Outcomes/ Resource Use Measures **2e.1 Data/sample** *(description of data/sample and size)*: n/a **2e.2 Analytic Method** *(type of risk adjustment, analysis, & rationale)*: n/a **2e.3 Testing Results** *(risk model performance metrics)*: n/a **2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:** n/a	**2e** C☐ P☐ M☐ N☐ NA☐
2f. Identification of Meaningful Differences in Performance **2f.1 Data/sample from Testing or Current Use** *(description of data/sample and size)*: See 2.b.2 and full report at end of this submission for full description of published data on FACT-G **2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: See 2.b.2 and full report at end of this submission on different analytic methods used with FACT-G **2f.3 Provide Measure Scores from Testing or Current Use** *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*: See 2.b.3 and full report at end of this submission for full description of different scores and analytic methods with FACT-G	**2f** C☐ P☐ M☐ N☐
2g. Comparability of Multiple Data Sources/Methods **2g.1 Data/sample** *(description of data/sample and size)*: The FACT-G is widely used in different studies in many different patient populations in and many different ways. Data sample characteristics from published data are listed in 2.6.1 with a more full description in the FACT-G report attached at the end of this submission. **2g.2 Analytic Method** *(type of analysis & rationale)*: See 2.b.2 and full report at end of this submission on different analytic methods used with FACT-G **2g.3 Testing Results** *(e.g., correlation statistics, comparison of rankings)*: See 2.b.2 and full report at end of this submission for full description of different scores and analytic methods with FACT-G	**2g** C☐ P☐ M☐ N☐ NA☐
2h. Disparities in Care **2h.1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: n/a **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** The FACT-G can help identify disparities in care/treatment regime as outlined in the literature. We have also done significant work in identifying challenges for low-literacy patients and in cross-cultural populations. We have also assessed different methods of administration to reduce patient burden, all with the hope of reducing assessment burden across all populations. All FACT scales are designed for patient self-administration, but can also be administered by interview format. Interview administration is considered appropriate after adequate training of interviewers so as to elicit non-biased patient responses. Technical (mode of administration) and statistical equivalence of similar scales in our measurement system have been demonstrated, providing the user with some flexibility as to mode of assessment (self versus interviewer administration) literacy level (high versus low) and language	**2h** C☐ P☐ M☐ N☐ NA☐

(English versus Spanish). One of the aims of a recently completed large multicenter study of cancer (n = 2356) patients was to test the psychometric properties and statistical equivalence of the English and Spanish language versions of the FACT subscale across literacy level (low vs. high) and mode of administration (self vs. interview). Technical equivalence across mode of administration was demonstrated in the high literacy patients; there were no differences in data quality or in mean QOL scores, after adjustment for performance status rating, socioeconomic status, gender and age. Technical equivalence between modes of administration with the FACT permits unbiased assessment of the impact of chronic illnesses and their treatments on patients from diverse backgrounds (Hahn & Cella, 1997). We have additional data to support the appropriateness of computer-administered versions of the questionnaire, including a multimedia touch screen program (Hahn & Cella, 2003). We are currently developing other novel administration methods such as computer-assisted telephone and web-based administration. Across these modes of administration, our preliminary data suggest that while there are small differences in the way people respond based on mode of administration, these alternate formats are essentially equivalent, particularly when deriving group statistics (e.g., means and variances.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* **C☐** **P☐** **M☐** **N☐**
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** in use 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** The FACIT items are currently being used in several NIH-funded initiatives which are being used in public and general health status assessments. Included in these initiatives are PROMIS (U01 AR 052 177), NeuroQOL (HHSN 265200436), Toolbox (AG-260-06-01) and others. 3a.3 **If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years***):** The FACT-G (and other FACIT questionnaires) are widely used in clinical trials and clinics to improve the quality of clinical care for cancer patients. In addition to the aforementioned PROMIS, NeuroQOL and Toolbox projects, the use of these questionnaires is mainstream in cooperative group oncology trials for assessing the impact of treatment on patients' QOL. Most noteably the PROMIS project's Assessment Center (www.nihpromis.org) is now available for widespread public use. Assessment Center is an online publicly available system which clinicians and researchers can use to capture patient-reported data. It allows for CAT and contains specific items and item parameters (including the FACT and FACIT items. To date there are over 13 different item banks (questions/items in domains such as Social Well Being, Fatigue, Pain, etc), the measurement characteristics of which have already been calculated by Dr. Cella and colleagues in the PROMIS initiative. Dr. Cella is also one of the founding members of the PROMIS Health Organization, a non-profit organization developed to support the ongoing PROMIS initiative. Other participants include faculty from the NIH, researchers from academic institutions, clinicians and representatives of the pharmaceutical industry.	**3a** **C☐** **P☐** **M☐** **N☐**

Dr. Cella has granted the PROMIS, Toolbox and NeuroQOL item banking projects permission to use all FACIT system items. **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) 3a.4 **Data/sample** *(description of data/sample and size)*: The data samples and publications on FACT-G data in previous sections of this submission as well as the full FACT-G report attached at the end of this submission demonstrate the widespread use and acceptance of this questionnaire by clinicians and researchers. 3a.5 **Methods** *(e.g., focus group, survey, QI project)*: Data from the FACT-G has been used and found to be valid and interpretable in all the projects listed in question 2. 3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*: Qualitative and quantitative results were described in question 2. More details can be found in the full FACT-G report attached at the end of this submission.	
3b/3c. Relation to other NQF-endorsed measures 3b.1 **NQF # and Title of similar or related measures:** No similar/related endorsed or submitted measures	
(for NQF staff use) **Notes on similar/related endorsed or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): 3b.2 **Are the measure specifications harmonized? If not, why?**	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value 3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** 5.1 **Competing Measures If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), describe why it is a more valid or efficient way to measure quality:**	**3c** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	**Eval Rating**
4a. Data Generated as a Byproduct of Care Processes 4a.1-2 **How are the data elements that are needed to compute measure scores generated?** Survey	**4a** C☐ P☐ M☐ N☐

4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**4b** C☐ P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences **4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** Perhaps the biggest source of inaccuracies in QOL data is missing data in the questionnaires. Until recently most data was collected via paper and pencil, resulting in missed responses which were then imputed during data analysis. Recent developments in use of electronic collection of health status assessments has reduced missing data, however, those methods are subject to the budgetary constraints of the study sponsor. In the past several years, Dr. Cella and his colleagues have made impressive advances in IRT and CAT (computerized adaptive testing) which significantly reduces patient/respondent burden by lowering the number of items/questions required to produce a QOL score. This type of assessment requires access to a computer and/or the internet, which is also dependent on sponsor funding. It also reduces the likelihood of including low socio-economic participants.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** As stated above, prior to 3-4 years ago, QOL data was largely collected via paper and pencil which resulted in missing data. The missing data is dealt with via several different widely published statistical analyses methods (Bernhard, J., Cella, D., Coates, A., Fallowfield, L., Ganz, P.A., Moinpour, C., Mosconi, P., Osoba, D., Simes, J., & Hurny, C. (1998). Missing quality of life data in cancer clinical trials: Serious problems and challenges. Statistics in Medicine, 17, 517-532.) The timing and frequency of data collection is dependent on the type of disease, treatment or symptom being assessed. Patient confidentiality is handled differently according to type of assessment: if electronic, there are encryption and password protections required by HIPAA which are implemented in the database development; If paper and pencil, study coordinators are responsible for ensuring files are locked and monitored, again according to HIPAA guidelines. The largest cost of data collection for paper and pencil is the Research Assistant or questionnaire training staff time, as well as the data entry and management time. These costs are largely bypassed by ePRO (electronic Patient Reported Outcomes)assessments, however for ePRO, there are significant computer programming costs. When IRT is included, there are also significant psychometrician and biostatistician algorithm development costs. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:**	**4e** C☐ P☐ M☐ N☐

There is no cost for the use of any of the English versions of the FACIT measures. Licensing costs for use of the non-English multilingual versions are $1500 per subscale, per language, per trial for Roman font alphabet languages (ie French, German, Spanish) and $2000 per subscale, per language, per trial for non-Roman font languages (ie Greek, Hebrew, Russian).

4e.3 Evidence for costs:
The evidence of these costs is 15 years' experience in NIH-funded research with these scales (including cooperative group oncology trials) as well as consulting with pharmaceutical companies who use the FACIT scales in their trials.
It should be noted for the FACIT Fatigue that it is a short-form (only 13 items). Short forms/symptom indices allow for a more brief assessment which is less expensive to put into clinical practice or clinical trials. However, such a short form does not provide a full QOL measure since the other domains (such as social/family well being) are not assessed.

4e.4 Business case documentation:
The clinical trials industry uses QOL endpoints as a secondary endpoint for label claims. NIH-funded initiatives (noteably AHRQ) are including patient perspective of treatment burden for comparitive effectiveness research initiatives.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, met? **Rationale:**	**4** **C☐** **P☐** **M☐** **N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐** **N☐** **A☐**

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
David Cella at FACIT.org | 381 S. Cottage Hill Avenue | Elmhurst | Illinois | 60126

Co.2 Point of Contact
Lauren | Lent, M.S. | l-lent@northwestern.edu | 630-531-7959

Measure Developer If different from Measure Steward
Co.3 Organization
David Cella at FACIT.org | 381 S. Cottage Hill Avenue | Elmhurst | Illinois | 60126

Co.4 Point of Contact
Lauren | Lent, M.S. | l-lent@northwestern.edu | 630-531-7959

Co.5 Submitter If different from Measure Steward POC
Lauren | Lent, M.S. | l-lent@northwestern.edu | 630-531-7959

Co.6 Additional organizations that sponsored/participated in measure development

«developer_other_orgs»

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 **Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.**

Ad.2 **If adapted, provide name of original measure:**

Ad.3-5 **If adapted, provide original specifications URL or attachment**

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 **Year the measure was first released:** 1997
Ad.7 **Month and Year of most recent revision:** 2005-04
Ad.8 **What is your frequency for review/update of this measure?** Due to the item banking work in PROMIS, all the FACIT items are under continual review

Ad.9 **When is the next scheduled review/update for this measure?** 2010-03

Ad.10 **Copyright statement/disclaimers:** Copyright 1987,1997

Ad.11 -13 **Additional Information web page URL or attachment:**

Date of Submission (*MM/DD/YY*)**:** «date_submitted»