

Original Investigation

Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE)

Amylou C. Dueck, PhD; Tito R. Mendoza, PhD; Sandra A. Mitchell, PhD, CRNP, AOCN; Bryce B. Reeve, PhD; Kathleen M. Castro, RN, MS, AOCN; Lauren J. Rogak, MA; Thomas M. Atkinson, PhD; Antonia V. Bennett, PhD; Andrea M. Denicoff, MS, RN, ANP; Ann M. O'Mara, PhD, RN, FAAN; Yuelin Li, PhD; Steven B. Clauser, PhD, MPA; Donna M. Bryant, MSN, ANP-BC, OCN, CCRC; James D. Bearden III, MD, FACP; Theresa A. Gillis, MD; Jay K. Harness, MD; Robert D. Siegel, MD, FACP; Diane B. Paul, AAS; Charles S. Cleeland, PhD; Deborah Schrag, MD, MPH; Jeff A. Sloan, PhD; Amy P. Abernethy, MD, PhD; Deborah W. Bruner, RN, PhD, FAAN; Lori M. Minasian, MD, FACP; Ethan Basch, MD, MSc; for the National Cancer Institute PRO-CTCAE Study Group

IMPORTANCE To integrate the patient perspective into adverse event reporting, the National Cancer Institute developed a patient-reported outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE).

OBJECTIVE To assess the construct validity, test-retest reliability, and responsiveness of PRO-CTCAE items.

DESIGN, SETTING, AND PARTICIPANTS A total of 975 adults with cancer undergoing outpatient chemotherapy and/or radiation therapy enrolled in this questionnaire-based study between January 2011 and February 2012. Eligible participants could read English and had no clinically significant cognitive impairment. They completed PRO-CTCAE items on tablet computers in clinic waiting rooms at 9 US cancer centers and community oncology practices at 2 visits 1 to 6 weeks apart. A subset completed PRO-CTCAE items during an additional visit 1 business day after the first visit.

MAIN OUTCOMES AND MEASURES Primary comparators were clinician-reported Eastern Cooperative Oncology Group Performance Status (ECOG PS) and the European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire (QLQ-C30).

RESULTS A total of 940 of 975 (96.4%) and 852 of 940 (90.6%) participants completed PRO-CTCAE items at visits 1 and 2, respectively. At least 1 symptom was reported by 938 of 940 (99.8%) participants. Participants' median age was 59 years; 57.3% were female, 32.4% had a high school education or less, and 17.1% had an ECOG PS of 2 to 4. All PRO-CTCAE items had at least 1 correlation in the expected direction with a QLQ-C30 scale (111 of 124, $P < .05$ for all). Stronger correlations were seen between PRO-CTCAE items and conceptually related QLQ-C30 domains. Scores for 94 of 124 PRO-CTCAE items were higher in the ECOG PS 2 to 4 vs 0 to 1 group (58 of 124, $P < .05$ for all). Overall, 119 of 124 items met at least 1 construct validity criterion. Test-retest reliability was 0.7 or greater for 36 of 49 prespecified items (median [range] intraclass correlation coefficient, 0.76 [0.53-.96]). Correlations between PRO-CTCAE item changes and corresponding QLQ-C30 scale changes were statistically significant for 27 prespecified items (median [range] $r = 0.43$ [0.10-.56]; all $P \leq .006$).

CONCLUSIONS AND RELEVANCE Evidence demonstrates favorable validity, reliability, and responsiveness of PRO-CTCAE in a large, heterogeneous US sample of patients undergoing cancer treatment. Studies evaluating other measurement properties of PRO-CTCAE are under way to inform further development of PRO-CTCAE and its inclusion in cancer trials.

JAMA Oncol. doi:10.1001/jamaoncol.2015.2639
Published online August 13, 2015.

 [Invited Commentary](#)

 [Supplemental content at jamaoncol.com](#)

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: National Cancer Institute PRO-CTCAE Study Group members are listed at the end of the article.

Corresponding Author: Amylou C. Dueck, PhD, Mayo Clinic, 13400 E Shea Blvd, Scottsdale, AZ 85259 (dueck@mayo.edu).

In cancer clinical trials, adverse events (AEs) are collected and reported using the US National Cancer Institute's (NCI's) Common Terminology Criteria for Adverse Events (CTCAE).¹ The CTCAE is a library of items representing 790 discrete AEs, each graded using an ordinal severity scale.² Approximately 10% of AEs in the CTCAE are symptoms (eg, nausea, sensory neuropathy), which in trials have historically been reported by clinical investigators.³ However, there is empirical evidence that collection of this information directly from patients improves the precision and reliability of symptomatic AE detection in trials⁴⁻⁹ and is feasible.^{10,11} Moreover, there is substantial evidence that clinical investigators may miss up to half of patients' symptomatic AEs.^{5,6,12,13}

To improve precision and patient-centeredness in the capture of symptomatic AEs, the NCI developed a library of patient-reported outcome (PRO) items to supplement the CTCAE, called the PRO-CTCAE,¹⁴ as has been previously described.¹⁵ Of the 790 AEs in the CTCAE, 78 were identified as amenable to patient self-report. For each of these AEs, PRO items were created reflecting the attributes of frequency, severity, interference with usual or daily activities, amount, or presence or absence. For any given AE, 1 to 3 attributes were selected depending on the content of the CTCAE criteria for that AE and the nature of that particular AE. In total, 124 individual items represent the 78 symptomatic AEs currently in the PRO-CTCAE item library.

The generic structure for PRO-CTCAE items and response options is shown in Table 1. Each item includes a plain language term for the AE, the attribute of interest, and the standard recall period of "the past 7 days." Cognitive interviews previously determined a high level of patient understanding and meaningfulness of the items.¹⁶ Software was developed for administering PRO-CTCAE items to patients either via World Wide Web or an automated telephone interactive voice response interface, and was refined through usability testing.^{15,17}

For any new measurement tool in clinical research (eg, biomarkers, imaging, diagnostic test), it is essential to establish that the new instrument accurately and reliably captures the underlying phenomenon that it is intended to measure. To accomplish this for the PRO-CTCAE, this study was designed to evaluate the measurement properties of the 124 items in the PRO-CTCAE item library including validity (degree to which an instrument accurately measures the underlying phenomenon), reliability (ability of an instrument to produce similar scores on repeated measurements under similar conditions), and responsiveness (capacity of an instrument to show a change when there has been a change in the underlying phenomenon). These properties were examined individually for each item because PRO-CTCAE items are individually reported in trials and not aggregated into a single score. Inclusion of patients with diversity with respect to cancer type, treatment modality, and sociodemographic characteristics was considered essential given the intended use of PRO-CTCAE across varying research contexts. To simultaneously evaluate the measurement properties of 124 items within a single study required us to use a varied set of comparators or "anchors" and warranted a larger and more diverse sample of respondents and settings than is typically used in most validation studies of fixed-length PRO measures.

At a Glance

- Symptomatic adverse events (AEs) in cancer trials are currently graded by clinicians using the National Cancer Institute's (NCI's) Common Terminology Criteria for Adverse Events (CTCAE).
- This study assessed the measurement properties (validity, reliability, and responsiveness) of the newly developed NCI Patient-Reported Outcomes version of the CTCAE (PRO-CTCAE).
- A total of 940 adults with cancer undergoing outpatient cancer treatment provided PRO-CTCAE and other patient-reported and clinical data.
- Most of the PRO-CTCAE items (119 of 124) met at least a validity criterion.
- The PRO-CTCAE provides a valid and reliable assessment of symptomatic toxic effects from the patient's perspective and is encouraged for use in oncology trials to enhance the accuracy of AE reporting.

Table 1. Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) Item Formats^a

Please think back over the past 7 days:	Example
Frequency (25 symptomatic AE terms): How often did you have _____? Never/Rarely/Occasionally/Frequently/Almost constantly	Vomiting
Severity (51 symptomatic AE terms): What was the severity of your _____ at its worst? None/Mild/Moderate/Severe/Very severe	Pain
Interference (25 symptomatic AE terms): How much did _____ interfere with your usual or daily activities? Not at all/A little bit/Somewhat/Quite a bit/Very much	Sudden urges to urinate
Presence (21 symptomatic AE terms): Did you have any _____? No/Yes	Unusual darkening of the skin
Amount (2 symptomatic AE terms): Did you have any _____? Not at all/A little bit/Somewhat/Quite a bit/Very much	Hair loss

Abbreviation: AE, adverse event.

^a See Basch et al¹⁵ for a complete listing of PRO-CTCAE items.

Methods

Patients

Adult patients initiating or undergoing outpatient chemotherapy, radiation therapy, or both at 1 of 9 US cancer centers or community oncology practices were approached in clinical waiting areas and invited to participate in this study. Participating sites with number of patients enrolled included Dana-Farber Cancer Institute, Boston, Massachusetts (n = 40); Hartford Hospital-Helen and Harry Gray Cancer Center, Hartford, Connecticut (n = 104); Helen F. Graham Cancer Center and Research Institute at Christiana Care Health System, Newark, Delaware (n = 105); Mayo Clinic, Rochester, Minnesota (n = 9); Memorial Sloan Kettering Cancer Center, New York, New York (n = 280); Our Lady of the Lake and Mary Bird Perkins Cancer Center, Baton Rouge, Louisiana (n = 133); Gibbs Cancer Center, Spartanburg, South Carolina (n = 113); St Joseph Hospital of Orange, Orange, California (n = 104); and University of Texas M. D. Anderson Cancer Center, Houston (n = 52).

Eligibility criteria required that all participants be able to read and comprehend English, be without clinically significant cognitive impairment on the basis of site investigator judgment, have a cancer diagnosis, and be actively undergoing cancer treatment or be initiating treatment within the next 7 days. Patients with any cancer type were eligible, but an accrual strategy was used to enrich for specific cancer types to facilitate planned comparisons between groups based on cancer type in the validity analysis, including breast, aerodigestive tract (head/neck and esophageal cancer), genitourinary (prostate and bladder), lung, colorectal, and lymphoma or myeloma. An enrichment strategy was also used to ensure that a minimum of 15% of participants had impaired performance status (PS), defined as Eastern Cooperative Oncology Group (ECOG) PS of at least 2.

Study sites were selected to encompass geographic, racial/ethnic, economic, and educational diversity reflective of the US population with the understanding that the requirement to be English speaking would limit the enrollment of Hispanic patients (a separate study evaluating the Spanish language version of the PRO-CTCAE has been conducted¹⁸). Race/ethnicity was self-reported by patients.

Institutional review board approval was obtained at all sites and at the NCI, and all patients provided written informed consent. Each participant received a \$20 gift card or parking voucher.

Questionnaire

The previously developed PRO-CTCAE item library consists of 78 symptomatic AEs represented by 124 distinct items.^{14,15} To limit burden, a maximum of 58 symptomatic AEs (82 items) was presented to each participant. Seven electronic surveys targeted toward different cancer types (eTable 1 in the [Supplement](#)) were created in the central PRO-CTCAE web survey administration platform. As part of the registration process, the site coordinator selected a single survey on the basis of the patient's diagnosis, and that survey was then automatically scheduled for completion at each visit. All surveys included a set of 20 "core" symptomatic AEs,¹⁵ predetermined on the basis of high prevalence across cancer types in prior NCI-sponsored clinical trials.¹⁹ Remaining symptomatic AEs were classified a priori as likely to be prevalent or nonprevalent in specific cancer types on the basis of expert consultation, patient representative input, and literature review. These items were included on surveys for selected cancer types to facilitate planned comparisons between groups based on cancer type. When 80% of accrual was reached, to increase sample size for the 58 symptomatic AEs that were not systematically administered to all patients, a new survey containing exactly these 58 symptomatic AEs was administered to all subsequently enrolled patients.

Procedure

The PRO-CTCAE items were completed by participants prior to clinic appointments on tablet computers via the PRO-CTCAE measurement system hosted on a secure server at the NCI.¹⁷ To optimize usability by individuals with disabilities, PRO-CTCAE software is compliant with Section 508 of the US Rehabilitation Act. The PRO-CTCAE measurement system uses conditional branching for AEs that contain more than a single attribute, such that subsequent items about severity or inter-

ference are skipped if respondents indicate that they are not experiencing a specific symptomatic AE. Participants were required to answer questions without assistance but could request technical assistance with using the tablet computer from study staff.

Anchors

Anchors are measurable criteria prespecified as comparators in an instrument validation study. Examples of anchors relevant in PRO validation studies include well-validated patient- and clinician-reported outcomes and clinical variables such as disease site or concurrent medication use. For this study, anchors selected a priori included both generic measures (eg, patient-reported global health-related quality of life [HRQOL] or clinician-reported PS) and more specific clinical variables (eg, antiemetic use or receipt of taxane chemotherapy). These anchors were selected on the basis of literature review, expert consensus, and patient representative input.

The PRO anchors were administered to participants using a paper booklet containing the European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ-C30),²⁰ a 30-item instrument that produces an HRQOL summary score,^{21,22} a global health status/quality of life (QOL) scale score, 5 functioning (physical, role, emotional, social, cognitive) scale scores, and 9 selected symptom item/scale scores. There are 28 items measured on a 1 to 4 scale (1 = not at all; 4 = very much), with the remaining 2 items (overall health and QOL) scored on a 1 to 7 scale (1 = very poor; 7 = excellent). Like PRO-CTCAE, the recall period for the QLQ-C30 is "the past week." Patients also completed 3 Global Impression of Change (GIC)^{23,24} items at the primary follow-up visit. These items asked patients to rate their changes in overall QOL, physical condition, and emotional state on a 7-point scale ranging from "very much better," "moderately better," "a little better," "about the same," "a little worse," "moderately worse," to "very much worse."

Clinician-reported ECOG PS was collected at each visit via a case report form. Other clinical anchors were abstracted from medical records and included whether the participant had received radiation therapy, surgery, and/or chemotherapy in the prior 2 weeks; type of chemotherapy; and use of specific medication classes, including hormonal therapy, narcotic analgesics, laxatives or stool softeners, antiemetics, sleep aids, antiarrhythmic medications, antacids, bronchodilators or inhaled corticosteroids, anxiolytics, and/or antidepressants.

Study Visits

Participants were assigned to 1 of 3 groups with differing questionnaire schedules based on cancer type and clinic visit schedule, to avoid the necessity of extra clinic visits in this symptomatic population (eFigure 1 in the [Supplement](#)). Group A included patients undergoing daily radiation or chemoradiation therapy to enable analyses of test-retest reliability and varying recall periods (recall period analyses will be reported separately).²⁵ Group B included patients with at least 4 planned consecutive weekly clinic visits. Group C included participants whose planned clinic visits precluded participation in group B but who did have a return clinic visit planned within

1 to 6 weeks. Irrespective of group assignment, all patients completed PRO-CTCAE items and QLQ-C30 at 2 visits that were spaced approximately 1 to 6 weeks apart. At each visit, ECOG PS and other clinical anchors were recorded on case report forms. The PRO-CTCAE surveys administered to patients in group A on the business day following study day 1 were used for the analysis of test-retest reliability, and included 49 prespecified PRO-CTCAE items.

Statistical Analysis

Construct validity reflects the association between a new measurement tool and an established measure of the underlying concept(s) of interest. Construct validity is often investigated through convergent validity, which determines whether the new measure moves in the same direction as an established instrument, and known-groups validity, which determines whether the measurement tool can distinguish between groups of patients who are thought to be distinct with respect to the underlying concept being measured. To assess convergent validity, Pearson correlations were computed between each PRO-CTCAE item and QLQ-C30 HRQOL summary and other functioning/symptom scale scores. To aid interpretation, QLQ-C30 HRQOL summary and functioning/global scales were reverse scored such that higher scores represent inferior outcomes, matching the direction of PRO-CTCAE items. Pearson correlation values of 0.1, 0.3, and 0.5 were interpreted as small, medium, and large.^{26(pp19-108)} To assess known-groups validity, 2-sample *t* tests for ordinal 0 to 4 scales and χ^2 tests for binary scales were used to compare each PRO-CTCAE item between patients with high and low PS (ECOG PS 0-1 vs 2-4). Additional known-groups analyses were prespecified for PRO-CTCAE items that were expected to be higher in 1 group of patients vs another on the basis of cancer type, treatment, or other clinically relevant characteristic (eg, pain in the abdomen in patients with gastrointestinal vs lung cancers). Effect sizes (computed as the difference between group means divided by the pooled standard deviation [Cohen *d*], or difference between twice the arcsine of the square root of each sample proportion [Cohen *h*]) of 0.2, 0.5, and 0.8 were interpreted as small, medium, and large, respectively.^{26(pp19-108)}

Test-retest reliability was estimated using the intraclass correlation coefficient (ICC) based on a 1-way analysis of variance model²⁷ with an ICC of 0.7 or greater interpreted as high.^{28(pp264-265)} Responsiveness of items was investigated by comparing change from first to second visit in 27 PRO-CTCAE items selected a priori. Comparisons were made using a 1-sided Jonckheere-Terpstra test across respondents who reported their GIC to be worse (“a little worse,” “moderately worse,” or “very much worse”), unchanged (“about the same”), or improved (“a little better,” “moderately better,” or “very much better”).²⁹ Standardized response means (SRMs) were computed as the mean change score divided by the standard deviation of the change scores within each change category (worse vs no change vs improved) for each PRO-CTCAE item. Pearson correlations were also computed between PRO-CTCAE item changes and QLQ-C30 scale changes. One GIC item and 1 QLQ-C30 scale were specified a priori for each of the 27 PRO-CTCAE items. See eTable 2 in the Supplement for symptomatic AEs included in each analysis.

To accommodate conditional branching in the PRO-CTCAE software, values for automatically skipped items were assumed to be zero. $P < .05$ was considered statistically significant. To take into consideration potential collinearity and multiplicity, sensitivity analyses used a stricter cutoff of $P < .001$ and the Hochberg step-up procedure³⁰ across construct validity analyses within each item. An item was considered valid if statistical significance ($P < .05$) along with a meaningful effect size (Pearson $r \geq 0.1$ or group difference effect size d or $h \geq 0.2$) was observed for at least 1 convergent or known-groups validity analysis.

Results

Between January 2011 and February 2012, 975 patients initiating or undergoing chemotherapy and/or radiation therapy were enrolled, with 940 of 975 (96.4%) eligible patients completing PRO-CTCAE items at visit 1 and 852 of 940 (90.6%) completing PRO-CTCAE items at visit 2 (eFigure 1 in the Supplement). Characteristics of the 940 participants included in this analysis are presented in Table 2. Median (range) age was 59 (19-91) years, 539 (57.3%) were female, 161 (17.1%) had impaired PS (ECOG 2-4), and 305 (32.4%) had no more than a high school education.

Almost all participants (938 of 940) reported the presence of at least 1 symptom (ie, a score >0) during the 2 primary visits, with 768 of 940 (81.7%) reporting at least 1 symptom as frequent, severe, and/or interfering “quite a bit” with daily activities. Patients were broadly symptomatic, reporting presence of a median (range) of 23 (0-91) symptoms, with 904 of 940 (96.2%) reporting presence of 5 or more symptoms at the first visit. Of the 124 PRO-CTCAE items, 118 (95.2%) were reported as present by at least 10% of respondents at both primary visits, with 82 of 124 (66.1%) items having at least 25% prevalence. The distribution of item scores for the set of 20 “core” symptomatic AEs appears in eFigure 2 in the Supplement.

Detailed results related to construct validity of PRO-CTCAE items using all anchors are provided in eTable 3 in the Supplement. With respect to convergent validity, 122 of 124 (98.4%) PRO-CTCAE items were associated in the expected direction with the QLQ-C30 HRQOL summary score (102 of 124, $P < .05$ for all; 87 of 124, $P < .001$ for all) (Figure 1); 107 of 124 items demonstrated meaningful correlation (Pearson $r \geq 0.1$). When all QLQ-C30 functioning/global scales were considered, all 124 PRO-CTCAE items were associated in the expected direction with 1 or more scales, with 114 of 124 demonstrating meaningful correlation (Pearson $r \geq 0.1$), and 111 of 124 coefficients were statistically significant ($P < .05$ for all; 90 of 124, $P < .001$ for all). The PRO-CTCAE items that were likely to affect physical functioning had the strongest correlations with the QLQ-C30 physical functioning scale (eg, shortness of breath severity: Pearson $r = 0.47$, $P < .001$), whereas items likely to affect cognitive functioning had the strongest correlations with the QLQ-C30 cognitive functioning scale (eg, problems with concentration severity: Pearson $r = 0.71$, $P < .001$; problems with memory severity: Pearson $r = 0.69$, $P < .001$). Simi-

Table 2. Patient Characteristics

Characteristic	No. (%) (N = 940)
Age at enrollment, median (range), y	59 (19-91)
Age group, y, No. (%)	
<30	23 (2.5)
30-64	597 (63.5)
65-74	235 (25.0)
≥75	85 (9.0)
Sex, No. (%)	
Female	539 (57.3)
Male	401 (42.7)
Ethnicity, No. (%)	
Hispanic or Latino	56 (6.0)
Not Hispanic or Latino	832 (88.5)
Missing	52 (5.5)
Race, No. (%)	
White	675 (71.8)
Black or African American	203 (21.6)
Asian	42 (4.5)
Other or multiple races reported	8 (0.9)
Missing	12 (1.3)
Education, No. (%)	
High school or less	305 (32.4)
Some college	199 (21.2)
College graduate or more	415 (44.1)
Missing	21 (2.2)
Cancer type, No. (%)	
Lung, head, or neck	329 (35.0)
Breast	260 (27.7)
Genitourinary or gynecologic	172 (18.3)
Gastrointestinal	95 (10.1)
Hematologic	47 (5.0)
Other or unknown	37 (3.9)
ECOG performance status at first visit, No. (%)	
0-1	779 (82.9)
2-4	161 (17.1)
Cancer treatment in prior 2 wk, No. (%)	
Chemotherapy	522 (55.5)
Radiation	424 (45.1)
Surgery	35 (3.7)

Abbreviation: ECOG, Eastern Cooperative Oncology Group.

lar results were seen between PRO-CTCAE items and conceptually related QLQ-C30 emotional, role, and social functioning scales. For those PRO-CTCAE items with a parallel QLQ-C30 symptom scale/item (eg, fatigue), large correlations between analogous items (all Pearson $r > 0.69$, $P < .001$) were consistently observed.

In the known-groups comparison between patients with low and high PS, 94 of 124 PRO-CTCAE items had higher mean scores in the ECOG PS 2 to 4 group vs 0 to 1 group (58 of 124, $P < .05$ for all; 37 of 124, $P < .001$ for all; shown for 37 PRO-CTCAE items in eFigure 3 in the Supplement).

In 127 a priori known-groups comparisons involving 87 PRO-CTCAE items based on cancer type, treatment, or other

clinically relevant characteristic, 110 of 127 comparisons demonstrated higher PRO-CTCAE scores in the group expected to have worse symptom experience (85 of 127, $P < .05$ for all; 53 of 127, $P < .001$ for all) (eTable 3 in the Supplement).

Most PRO-CTCAE items (119 of 124) had a statistically significant and meaningful effect size on 1 or more construct validity criteria. The 5 items that did not exhibit at least 1 statistically significant and meaningful effect had low prevalence in this sample, thereby limiting our analysis. These items were nosebleeds (prevalence, 15% [frequency] and 14% [severity]); pain, swelling, or redness at site of drug injection or intravenous therapy (prevalence, 13%); pain during vaginal sex (prevalence, 21%); and rash (prevalence, 17%). Most PRO-CTCAE items (99 of 124 and 101 of 124) remained statistically significant under stricter criteria ($P < .001$ and Hochberg $P < .05$, respectively) in sensitivity analyses (eTable 3 in the Supplement).

In the subset of 80 respondents who completed PRO-CTCAE on consecutive business days (median [range], 1 [1-3] days), the test-retest reliability for the 49 prespecified items ranged from 0.53 to 0.96 (median ICC, 0.76) with 36 of 49 items having an ICC of at least 0.7 (eTable 4 in the Supplement).

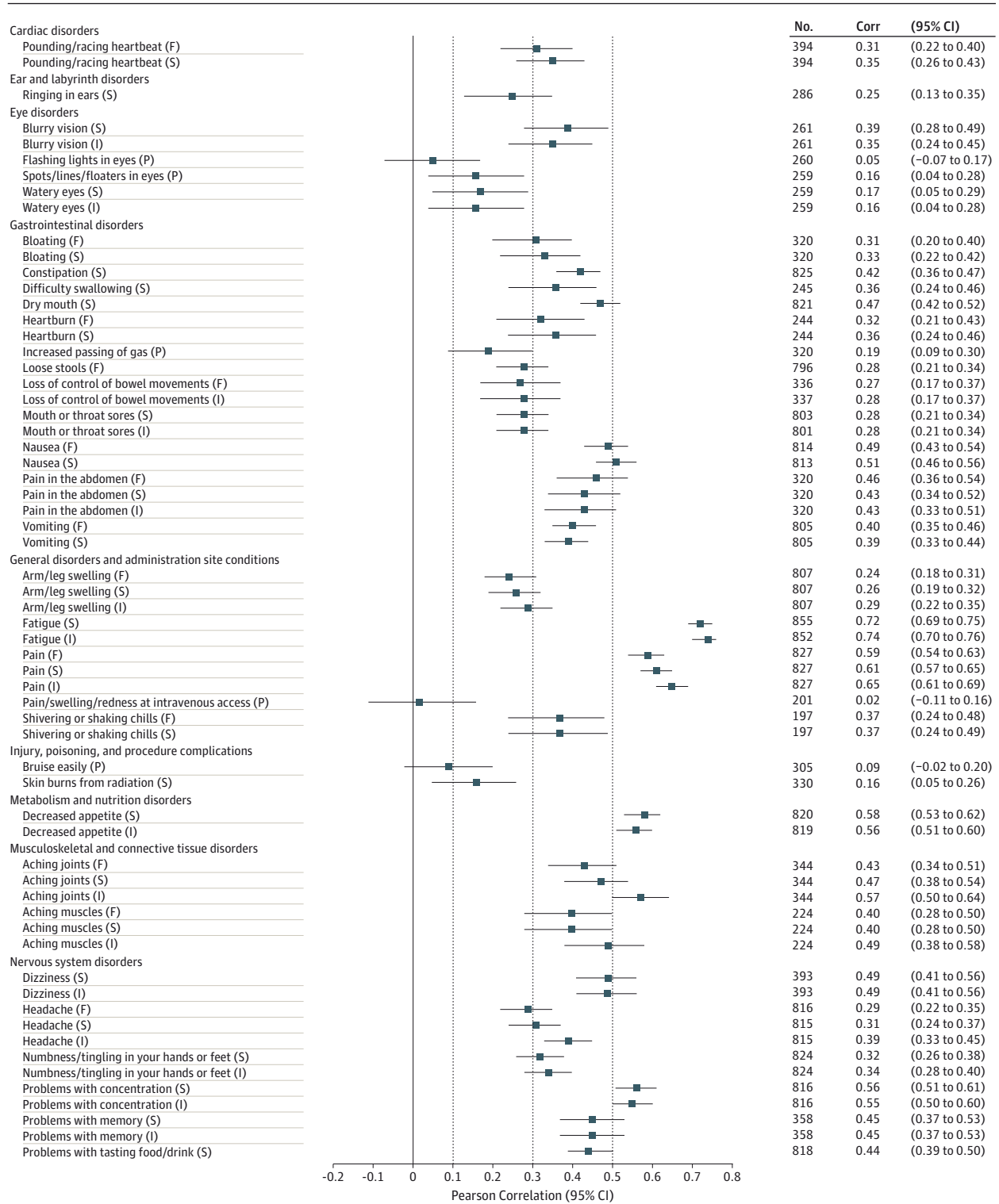
In the analysis of responsiveness (Figure 2), statistically significant ($P < .05$) monotonically decreasing mean PRO-CTCAE change scores were observed for 23 of 27 prespecified items ($P < .001$ for 13 items). The median (range) SRM in patients reporting worsening was 0.19 (0.03-0.40), whereas that in patients reporting improvement was -0.14 (-0.30 to 0.09). Statistically significant correlations were observed between PRO-CTCAE item changes and corresponding QLQ-C30 scale changes for all 27 prespecified items (median [range] r , 0.43 [0.10-0.56]; all $P \leq .006$).

Discussion

This large-scale multicenter study in adults undergoing active cancer therapy provides evidence supporting the validity, reliability, and responsiveness of the items in the PRO-CTCAE library. The PRO-CTCAE is unique in its intended use to complement the CTCAE by providing comprehensive data on symptomatic AEs in cancer clinical trials from the patient perspective.

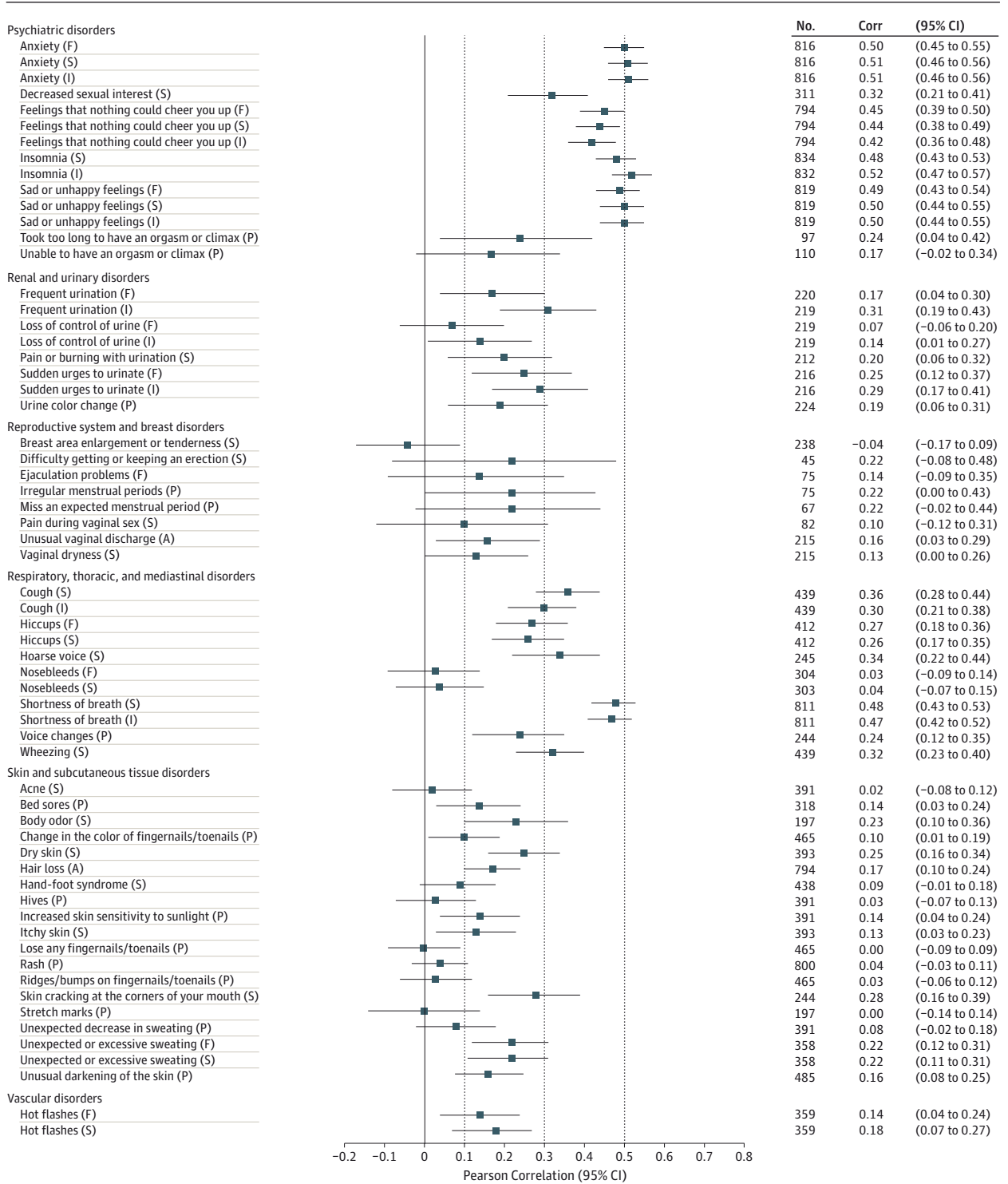
The design of this study posed a unique methodological challenge, due to the goal of assessing, within a single investigation, the measurement properties of 124 individual items representing a broad spectrum of symptomatic toxic effects. Typically, PRO validation studies will test the properties of a single composite index score or a small number of domains that encompass related concepts. For the assessment of validity in the present study, the primary strategy to address this challenge was inclusion of both broad generic anchors (eg, global HRQOL, ECOG PS) and more specific clinical variables (eg, receipt of specific medication classes such as antiemetics). Interestingly, all of the PRO-CTCAE items were associated in the expected direction with at least 1 generic functioning measure, suggesting the impact that even a single toxic effect may have on the patient experience.

Figure 1. Pearson Correlations Between 124 Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) Item Scores and European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ-C30) Health-Related Quality of Life Summary Score at Visit 1



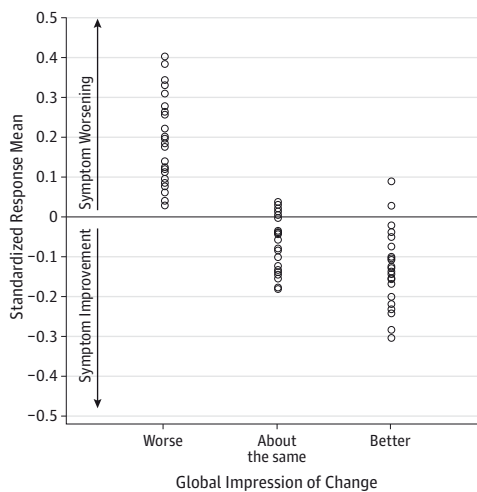
(continued)

Figure 1. Pearson Correlations Between 124 Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) Item Scores and European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ-C30) Health-Related Quality of Life Summary Score at Visit 1 (continued)



See eTable 3 in the Supplement for all computed Pearson correlations between PRO-CTCAE items and EORTC QLQ-C30 functioning, global, and symptom scales. Dashed lines at 0.1, 0.3, and 0.5 correspond to small, medium, and large effect sizes according to Cohen.²⁶ A indicates amount; Corr, correlation; F, frequency; I, interference with usual or daily activities; P, presence/absence; S, severity.

Figure 2. Standardized Response Means Across 27 Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) Items by Patient-Reported Global Impression of Change Category



Symbols show 27 frequency, severity, and interference items selected prior to initiation of the responsiveness analysis. The set of 20 "core" symptomatic adverse events (AEs) was reviewed and symptomatic AEs were selected if they had high potential to be meaningfully related to global changes in quality of life, physical condition, and/or emotional state (ie, the Global Impression of Change items that were administered at the second visit). Of the 20 reviewed symptomatic AEs, 13 were included on the basis of this criterion (see eTable 2 in the Supplement). The symptomatic AEs that were excluded were believed to be related to initiation or changes in specific treatments (dry mouth, problems with tasting food/drink, rash) so may not exhibit change in a heterogeneously treated sample of patients, may require a longer duration of follow-up to exhibit change (arm/leg swelling, hair loss), or may be related to cognitive condition (headache, problems with concentration), which was not assessed in the Global Impression of Change items.

Strengths of this study include the diverse sample, reflecting a wide range of cancer types and treatment modalities, and enrichment for less common cancer types. The sample was also successfully enriched for patients with impaired PS (ECOG PS ≥ 2), enabling demonstration of the meaningfulness of PRO-CTCAE among those with substantial symptom burdens, as well as the feasibility of survey administration in debilitated patients. Moreover, participants were accrued at both academic and community sites across the United States, including rural and urban settings, and reflected a range of educational and racial backgrounds.

Several caveats should be considered. First, our study was conducted in an English-speaking US-residing patient population. Ongoing research is evaluating linguistic adaptations

of PRO-CTCAE, and the measurement properties of both the English and other language versions in settings outside the United States.³¹ Linguistic validation of a Spanish language translation of PRO-CTCAE is being reported elsewhere.¹⁸ Second, we assessed reliability in a subset of 49 items; thus, future studies to examine the test-retest reliability of the remaining PRO-CTCAE items are warranted. Third, a small number of highly specific symptomatic AEs were uncommon in the study sample and received low endorsement rates, thus limiting our ability to evaluate their measurement properties. Specifically, 5 items reflecting 4 symptomatic AEs (nosebleeds; pain, swelling, and/or redness at site of drug injection or intravenous therapy; pain during vaginal sex; rash) did not exhibit a statistically significant and meaningful effect on at least 1 construct validity criterion. These items are being evaluated in other clinical trial contexts. Whereas the large number of items and anchors evaluated in this study raises the possibility of inflated type I error, in sensitivity analyses using more stringent significance thresholds, the majority of items retained statistical significance. Last, notwithstanding inclusion of participants with diverse cancer types in this study, results may not fully generalize to populations with rare tumor types. However, a prior cognitive interviewing study¹⁶ affirms that PRO-CTCAE items were well understood by respondents with varying disease sites and receiving diverse anti-cancer treatments. Continued evaluation of PRO-CTCAE is currently under way in a variety of trial contexts to support the interpretability and value of patient reporting of symptomatic treatment-related toxic effects.

Conclusions

The CTCAE has historically enabled clinicians to describe the toxicity burden of cancer treatments using a consistent standard language allowing comparisons across trials. The value of patients' input in describing their own experiences is well recognized. Having a measurement system that integrates the patient perspective into AE reporting and that fosters consistency, transparency, and comparability across trials is similarly an important objective. The results of this validation study suggest that PRO-CTCAE can achieve its intended aim of integrating the patient experience into routine clinical trial AE reporting, thereby augmenting the capacity for informed decision making. In conclusion, this large-scale multicenter validation study in individuals undergoing active cancer therapy provides robust evidence for the validity, reliability, and responsiveness of items in the PRO-CTCAE library.

ARTICLE INFORMATION

Accepted for Publication: June 11, 2015.

Published Online: August 13, 2015.

doi:10.1001/jamaoncol.2015.2639.

Author Affiliations: Department of Health Sciences Research, Mayo Clinic, Scottsdale, Arizona (Dueck); Department of Symptom Research, University of Texas MD Anderson Cancer Center, Houston (Mendoza, Cleeland); Division of Cancer

Control and Population Sciences, National Cancer Institute, Rockville, Maryland (Mitchell, Castro, Clauser); Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill (Reeve, Bennett, Basch); Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York (Rogak, Basch); Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, New York (Atkinson, Li); Division of Cancer

Treatment and Diagnosis, National Cancer Institute, Rockville, Maryland (Denicoff, Minasian); Division of Cancer Prevention, National Cancer Institute, Rockville, Maryland (O'Mara); Department of Clinical Research, Cancer Program of Our Lady of the Lake and Mary Bird Perkins, Baton Rouge, Louisiana (Bryant); Gibbs Cancer Center and Research Institute, Spartanburg, South Carolina (Bearden); Helen F. Graham Cancer Center and Research Institute, Christiana Care Health System,

Newark, Delaware (Gillis); Center for Cancer Prevention and Treatment, St. Joseph Hospital of Orange, Orange, California (Harness); Helen and Harry Gray Cancer Center, Hartford Hospital, Hartford, Connecticut (Siegel); patient advocate and cancer survivor, Brooklyn, New York (Paul); Division of Population Sciences, Dana-Farber Cancer Institute, Boston, Massachusetts (Schrag); Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota (Sloan); Department of Medicine, Duke University Medical Center, Durham, North Carolina (Abernethy); Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, Georgia (Bruner).

Author Contributions: Dr Dueck had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Dueck, Mendoza, Mitchell, Reeve, Rogak, Atkinson, Denicoff, O'Mara, Clauser, Bryant, Paul, Cleeland, Schrag, Sloan, Abernethy, Bruner, Minasian, Basch.

Acquisition, analysis, or interpretation of data: Dueck, Mendoza, Mitchell, Reeve, Castro, Rogak, Atkinson, Bennett, Denicoff, O'Mara, Li, Bryant, Bearden, Gillis, Harness, Siegel, Paul, Schrag, Abernethy, Bruner, Basch.

Drafting of the manuscript: Dueck, Mitchell, Rogak, Atkinson, Bennett, Li, Bryant, Schrag, Sloan, Abernethy, Bruner, Basch.

Critical revision of the manuscript for important intellectual content: Dueck, Mendoza, Mitchell, Reeve, Castro, Rogak, Atkinson, Bennett, Denicoff, O'Mara, Clauser, Bearden, Gillis, Harness, Siegel, Paul, Cleeland, Schrag, Abernethy, Minasian, Basch. **Statistical analysis:** Dueck, Mendoza, Mitchell, Rogak, Atkinson, Li, Sloan.

Obtained funding: Dueck, Castro, Abernethy. **Administrative, technical, or material support:** Dueck, Mendoza, Mitchell, Castro, Rogak, Atkinson, Li, Clauser, Bryant, Gillis, Cleeland, Schrag, Bruner, Basch.

Study supervision: Dueck, Mitchell, Castro, Rogak, Siegel, Abernethy, Minasian, Basch.

Conflict of Interest Disclosures: Drs Dueck, Mendoza, and Reeve, Ms Rogak, and Drs Atkinson, Bennett, Li, Bryant, Bearden, Gillis, Harness, Siegel, Cleeland, Schrag, Sloan, Abernethy, Bruner, and Basch served as investigators or study managers under contract or subcontract from the National Cancer Institute to develop the PRO-CTCAE. No other disclosures are reported.

Funding/Support: This study was supported by contracts from the National Cancer Institute: HHSN261200800043C, HHSN261201000063C, and HHSN261200800001E.

Role of the Funder/Sponsor: National Cancer Institute scientific staff (Mitchell, Castro, Denicoff, O'Mara, Clauser, Minasian) participated in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: Additional National Cancer Institute PRO-CTCAE Study Group members are Narre Heon, BA; Mary Shaw, BA; Sean Ryan, MS; Liora P. Stark, MPH; and Laura Sit, MS (Memorial Sloan Kettering Cancer Center, New York, New York); Amy Thomassie, RN, BSN, CCRP (Cancer Program of Our Lady of the Lake and Mary Bird Perkins, Baton Rouge, Louisiana); Wendy Pettus, MS, and Lucy Gansauer, RN, MSN, OCN (Gibbs Cancer Center, Spartanburg,

South Carolina); Barbara Marconi, RN; and Angela Steele-Tilton, MSN (Helen F. Graham Cancer Center and Research Institute, Christiana Care Health System, Newark, Delaware); Donna Malveaux (University of Texas M. D. Anderson Cancer Center, Houston); and Jennifer Wind, MA (Dana-Farber Cancer Institute, Boston, Massachusetts).

Additional Contributions: We thank the patients and clinicians who participated in this study. Gitana Davila and Kathy Alexander contributed study coordination, patient recruitment, and data management at the Center for Cancer Prevention and Treatment at St Joseph Hospital of Orange and Hartford Hospital—Helen and Harry Gray Cancer Center, respectively. Their institutions received funding to support their efforts. Neil Aaronson, PhD, provided methodological input on the EORTC QLQ-C30. He was not compensated for his contribution. The authors also wish to thank observers from the Food and Drug Administration who have provided valuable input during development of the PRO-CTCAE, including Ann Marie Trentacosti, MD; Virginia Kwitkowski, MS, RN, ACNP-BC; Laurie Burke, RPh, MPH (retired); Elektra Papadopoulos, MD, MPH; Ashley Slagle, PhD, MS; and Paul Kluetz, MD.

REFERENCES

- National Cancer Institute, National Institutes of Health, US Department of Health and Human Services. Common Terminology Criteria for Adverse Events (CTCAE), Version 4.0. NIH publication 09-7473. Published May 29, 2009; Revised June 14, 2010. http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf. Accessed March 16, 2015.
- Setser A. CTCAE v4.0: Frequently Asked Questions. http://evs.nci.nih.gov/ftp1/CTCAE/Archive/CTCAE_4.01_2009-07-14_FAQ.doc. Accessed March 16, 2015.
- Trotti A, Colevas AD, Setser A, Basch E. Patient-reported outcomes and the evolution of adverse event reporting in oncology. *J Clin Oncol*. 2007;25(32):5121-5127.
- Atkinson TM, Li Y, Coffey CW, et al. Reliability of adverse symptom event reporting by clinicians. *Qual Life Res*. 2012;21(7):1159-1164.
- Basch E. The missing voice of patients in drug-safety reporting. *N Engl J Med*. 2010;362(10):865-869.
- Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. 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(17):3485-3490.
- Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. *Am J Manag Care*. 2008;14(8):530-539.
- Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst*. 2009;101(23):1624-1632.
- Quinten C, Maringwa J, Gotay CC, et al. Patient self-reports of symptoms and clinician ratings as predictors of overall cancer survival. *J Natl Cancer Inst*. 2011;103(24):1851-1858.
- Meacham R, McEngart D, O'Gorman H, Wenzel K. Use and compliance with electronic patient

reported outcomes within clinical drug trials. Paper presented at: 2008 ISPOR (International Society for Pharmacoeconomics and Outcomes Research) Annual International Meeting; May 3-7, 2008; Toronto, Ontario, Canada.

- Judson TJ, Bennett AV, Rogak LJ, et al. Feasibility of long-term patient self-reporting of toxicities from home via the Internet during routine chemotherapy. *J Clin Oncol*. 2013;31(20):2580-2585.
- Di Maio M, Gallo C, Leigh NB, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. *J Clin Oncol*. 2015;33(8):910-915.
- Basch E, Iasonos A, McDonough T, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol*. 2006;7(11):903-909.
- Basch EM, Reeve BB, Mitchell SA, et al. Electronic toxicity monitoring and patient-reported outcomes. *Cancer J*. 2011;17(4):231-234.
- Basch E, Reeve BB, Mitchell SA, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst*. 2014;106(9):dju244.
- Hay JL, Atkinson TM, Reeve BB, et al; NCI PRO-CTCAE Study Group. Cognitive interviewing of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Qual Life Res*. 2014;23(1):257-269.
- Agarwal H, Gulati M, Baumgartner P, et al. Web interface for the patient-reported version of the CTCAE. Paper presented at: NCI caBIG Annual Meeting; July 20-22, 2009; Washington, DC.
- Arnold BJ, Mitchell SA, Lent L, et al. Linguistic validation of the Spanish translation of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events. Paper presented at: International Society for Quality of Life Research 21st Annual Conference; October 15-18, 2014; Berlin, Germany.
- Reeve BB, Mitchell SA, Dueck AC, et al. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. *J Natl Cancer Inst*. 2014;106(7):dju129.
- Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85(5):365-376.
- Gundy CM, Fayers PM, Groenvold M, et al. Comparing higher order models for the EORTC QLQ-C30. *Qual Life Res*. 2012;21(9):1607-1617.
- Kieffer JM, Giesinger JM, Fayers P, et al. Replication and relative validity of the mental and physical higher order model of the EORTC QLQ-C30. Paper presented at: International Society for Quality of Life Research 21st Annual Conference; October 15-18, 2014; Berlin, Germany.
- Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol*. 1998;16(1):139-144.

24. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR; Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002;77(4):371-383.
25. Mendoza TR, Bennett AV, Mitchell SA, et al. Impact of recall period on the accuracy of selected items from the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Poster presented at: International Society for Quality of Life Research 21st Annual Conference; October 15-18, 2014; Berlin, Germany.
26. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988.
27. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979; 86(2):420-428.
28. Nunnally JC, Berstein IH. *Psychometric Theory*. 3rd ed. New York: McGraw-Hill; 1994.
29. Jonckheere AR. A distribution-free k-sample test against ordered alternatives. *Biometrika.* 1954; 41:133-145.
30. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika.* 1988; 75:800-802.
31. Kirsch M, Mitchell SA, Dobbels F, et al. Linguistic and content validation of a German-language PRO-CTCAE-based patient-reported outcomes instrument to evaluate the late effect symptom experience after allogeneic hematopoietic stem cell transplantation. *Eur J Oncol Nurs.* 2015;19(1):66-74.