ARTICLE

Patient Self-Reports of Symptoms and Clinician Ratings as Predictors of Overall Cancer Survival

Chantal Quinten, John Maringwa, Carolyn C. Gotay, Francesca Martinelli, Corneel Coens, Bryce B. Reeve, Henning Flechtner, Eva Greimel, Madeleine King, David Osoba, Charles Cleeland, Jolie Ringash, Joseph Schmucker-Von Koch, Martin J. B. Taphoorn, Joachim Weis, Andrew Bottomley

Manuscript received November 30, 2010; revised October 12, 2011; accepted October 17, 2011.

Correspondence to: Chantal Quinten, MSc, Quality of Life Department, European Organization for Research and Treatment of Cancer, Ave E. Mounierlaan, 83/11, Brussels 1200 Belgium (e-mail: chantal.quinten@eortc.be; chantal.quinten@ecdc.europa.eu).

Background	The National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI-CTCAE) reporting system is widely used by clinicians to measure patient symptoms in clinical trials. The European Organization for Research and Treatment of Cancer's Quality of Life core questionnaire (EORTC QLQ-C30) enables cancer patients to rate their symptoms related to their quality of life. We examined the extent to which patient and clinician symptom scoring and their agreement could contribute to the estimation of overall survival among cancer patients.
Methods	We analyzed baseline data regarding six cancer symptoms (pain, fatigue, vomiting, nausea, diarrhea, and constipation) from a total of 2279 cancer patients from 14 closed EORTC randomized controlled trials. In each trial that was selected for retrospective pooled analysis, both clinician and patient symptom scoring were reported simultaneously at study entry. We assessed the extent of agreement between clinician vs patient symptom scoring using the Spearman and kappa correlation statistics. After adjusting for age, sex, performance status, cancer severity, and cancer site, we used Harrell concordance index (C-index) to compare the potential for clinician-reported and/or patient-reported symptom scores to improve the accuracy of Cox models to predict overall survival. All P values are from two-sided tests.
Results	Patient-reported scores for some symptoms, particularly fatigue, did differ from clinician-reported scores. For each of the six symptoms that we assessed at baseline, both clinician and patient scorings contributed independently and positively to the predictive accuracy of survival prognostication. Cox models of overall survival that considered both patient and clinician scores gained more predictive accuracy than models that considered clinician scores alone for each of four symptoms: fatigue (C-index = .67 with both patient and clinician data vs C-index = .63 with clinician data only; P <.001), vomiting (C-index = .64 vs .62; P = .01), nausea (C-index = .65 vs .62; P < .001), and constipation (C-index = .62 vs .61; P = .01).
Conclusion	Patients provide a subjective measure of symptom severity that complements clinician scoring in predicting overall survival.
	J Natl Cancer Inst 2011;103:1851–1858

Physicians use the Common Terminology Criteria for Adverse Events (CTCAE), as defined by the US National Cancer Institute (NCI), as a standard classification system for reporting adverse events in cancer clinical trials (1). An adverse event is defined as an unfavorable and unintended sign, symptom, or disease temporarily associated with the use of a medical treatment or procedure; adverse events can include abnormal laboratory findings. Developed initially in 1982 by the NCI's Cancer Therapy Evaluation Program, this rating system is currently in its fourth version. The rapid pace of revisions reflects the changing needs of investigators for usable adverse events criteria and the increasing complexity of cancer clinical trials. In version 4 of the NCI-CTCAE, there are 790 specific items, with approximately 10% of the adverse events representing symptoms such as fatigue, pain, and nausea.

Recently, it has become common practice in major American and European adult cancer clinical trial cooperative group studies to use patient-reported assessments of adverse events alongside clinician-reported assessments (2–4). This trend reflects the increasing importance of the patient's voice in cancer treatment, especially because studies (5–7) have shown that patient-reported outcomes can improve the accuracy and efficiency of symptomatic adverse event data collection.

Several articles (8–10) have reported that clinician-reported assessments of symptoms tend to underestimate, or in some cases overestimate, the symptom burden compared with patient ratings

CONTEXT AND CAVEATS

Prior knowledge
Clinicians regularly use the National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI-CTCAE) system to report cancer patients' symptoms in clinical trials. It was suggested that the inclusion of patients' self-assessments of their symptoms might improve models to predict their overall survival.

Study design
Clinician-reported data regarding six common patient symptoms (NCI-CTCAE) and similar patient-reported data from a European Organization for Research and Treatment of Cancer Quality of Life questionnaire (EORTC QLQ-C30) were collected at baseline from 14 EORTC randomized controlled trials involving 2279 cancer patients. The authors examined the extent to which responses from clinicians and patients were correlated and concordant. They also assessed whether the inclusion of patients' responses improved Cox models to predict overall survival.

Contribution
For each of the six symptoms studied (pain, fatigue, vomiting, nausea, diarrhea, and constipation), patient-reported scores did differ somewhat from clinician-reported scores. Both scores made a positive and additive contribution to the predictive accuracy of survival models.

Implication
Patient-reported data can complement clinician-reported data to achieve more reliable measurements of clinical outcomes.

Limitations
Relative to other factors, both patient and clinician scores of symptoms make a small contribution to the Cox models. Furthermore, the NCI-CTCAE and EORTC QLQ-30 were intended for different purposes and so it is difficult to exactly compare them. Last, patients with severe symptom burdens would be unlikely to be among this study population, which consisted of new enrollees in randomized trials.

From the Editors

and that substantial variability might exist between physician and patient ratings, particularly for more subjective symptoms such as fatigue [11,12]. Therefore, Varricchio and Sloan [13] have advocated the use of self-reported measures for many symptoms because of their subjective nature.

Bruner [14] reported several flaws with regard to clinician-reports on the CTCAE, such as high variability in scoring among clinicians and institutions and a lack of standardization in the implementation and recording of the tool. Moreover, these authors pointed out that there is no psychometric testing on the NCI-CTCAE and for most adverse events, no formal evidence exists as to what separates grade 1 toxicity from grade 2 and 3 and whether this separation represents clinical significance for the patient. The collection of patient-reported outcomes using validated psychometric questionnaires on a patient's first visit might augment these data to provide a better assessment tool for healthcare professionals who must choose the right treatments for their patients and must assess whether they are eligible to enter clinical trials in which toxic effects might become a too-high burden for some individuals.

In addition to more accurate assessments of symptoms, patient-reported outcomes might also provide some valuable information with regard to overall survival. A large prospective study by Christakes and Lamont [15] demonstrated that clinicians are often inaccurate in their prognosis for terminally ill patients; in that study, they overestimated survival by a factor of 5.3. In an extensive review of more than 39 selected trials, Gotay et al. [16] reported that patient-reported outcomes have prognostic value of overall survival in 36 of the 39 studies; however, the trials that they analyzed often used a wide variety of different patient-reported outcome measurements. Several large pooled studies [17,18] modeled the prognostic value of patient-reported outcomes using a single standard of measure and showed that incorporating patient-reported outcomes alongside clinician data improved the accuracy of predicting overall survival.

The aim of this study was to assess the level of agreement between clinician and patient reports of symptoms collected at baseline and to examine the value of each data source, alone and in combination, for predicting overall survival. The added value of patient-reported outcomes data was quantified by the degree of accuracy conferred by using both ratings to contribute to the prediction of overall survival in cancer patients. The analysis assessed the extent to which patient perspectives, clinician ratings, and the extent of agreement between these ratings contributes to estimating cancer survival.

Methods

Patients and Data

This study was conducted using a secondary dataset collected from 14 closed phase III European Organization for Research and Treatment of Cancer (EORTC) randomized controlled trials initiated at the EORTC between 1990 and 2002. The 14 EORTC randomized controlled trials, including a total of 3719 patients, were individually selected based on the availability of both clinician- and patient-reported symptom scoring at baseline and the use of overall survival as the primary outcome. Patients for whom a patient or clinician score was missing for a given symptom were excluded from further analysis, leaving a total of 2279 patients for whom data were collected for one or more symptoms. Nine of the 14 trials provided both patient and clinician scores for pain; five trials provided both patient and clinician scores for vomiting; six trials provided both patient and clinician scores for nausea; six trials provided both patient and clinician scores for diarrhea; five trials provided both patient and clinician scores for fatigue; four trials provided both patient and clinician scores for constipation.

All of the 14 selected trials measured symptom burden using the EORTC Quality of Life Core questionnaire (QLQ-C30) that was developed to assess the health-related quality of life of cancer patients [[19]; see Supplementary Materials, available online]. It surveys 15 health-related quality of life parameters; in addition to assessing the symptoms pain, fatigue, appetite loss, nausea and/or vomiting, diarrhea, constipation, sleep disturbance, and financial impact, it was also designed to measure the physical, psychological, and social functioning of cancer patients. The measures of functioning will not be discussed in this article because of the absence of comparable ratings by clinicians. In the EORTC QLQ-C30

questionnaire, the patient rated his or her symptoms on a 4-point ordinal scale in which a score of 1 meant "not at all," a score of 2 meant "a little," a score of 3 meant "quite a bit," and a score of 4 meant "very much" (20). Standard EORTC QLQ-C30 scoring of three symptoms uses multi-item subscales: for fatigue, there were three items scored; for nausea and vomiting, two items scored; and for pain, two items scored. In this study, nausea and/or vomiting was scored as two separate items to correspond with the NCI-CTCAE criteria, vomiting and nausea.

Baseline clinician assessments were performed in these 14 EORTC trials using version 2 of the NCI-CTCAE questionnaire (21), the accepted version at the time the 14 trials started (1990–2002). In this assessment instrument, clinicians were asked to assess the patient in terms of a predefined set of symptoms, as specified in the protocol, and to report their score on a case report form. For each symptom, a description was provided to facilitate physician ratings, which in this case were on a 5-point scale. In version 2.0, "grades" 0 through 4 were associated with unique clinical descriptions of the severity of adverse events based on the following general criteria: a score of 0 meant "none or normal." a score of 1 meant "mild," a score of 2 meant "moderate," a score 3 meant "severe," and a score of 4 meant "life threatening or disabling" (see questionnaire in Supplementary Materials, available online).

Statistical Methods

The mean scores and variability of the selected baseline symptoms were calculated for both the EORTC QLQ-C30 and NCI-CTCAE measures. Spearman correlation coefficients (ρ) were calculated as a simple exploratory statistic to study the magnitude of agreement between patient and clinician scoring; ρ ranged from −1 to 1, whereby a higher absolute value indicated a stronger correlation. Cohen (22) suggested that a correlation coefficient that is less than .30 indicates a weak relationship, a value between .30 and .50 indicates a moderate relationship, and a value greater than .50 indicates a strong relationship.

"Exact agreement" was defined as the proportion of identical paired responses between the EORTC QLQ-C30 and NCI-CTCAE assessment instruments. We considered each of the following pairs to be identical responses: EORTC QLQ-C30 score 1 vs NCI-CTCAE score 0; EORTC QLQ-C30 score 2 vs NCI-CTCAE score 1; EORTC QLQ-C30 score 3 vs NCI-CTCAE score 2; EORTC QLQ-C30 score 4 vs NCI-CTCAE score 3 and 4 combined. The level of exact agreement between patient vs clinician symptom rating was assessed using the Cohen kappa coefficient (κ) (23). This measure is used to assess the degree to which two or more raters, examining the same data, agree when it comes to assigning the data to categories. Landis and Koch (24) proposed categories for judging kappa values: a κ value less than 0.00 means poor agreement; κ between .00 and .20, slight agreement; κ between .21 and .40, fair agreement; κ between .41 and .60, moderate agreement; κ between .61 and .80, substantial agreement; and κ between .81 and 1.00, almost perfect agreement.

We evaluated the accuracy of clinician and patient scoring to predict overall survival, alone and in combination, after adjusting for age (≤60 vs >60 years) and sex (male vs female) (both prognostic sociodemographic variables) and metastasis (yes vs no), World Health Organization (WHO) performance status (WHO 0–1 vs WHO 2–3), and cancer site (all clinical variables). We calculated the ability of four Cox proportional hazards models to predict overall survival. The assumption of proportionality in the four models was assessed by investigating the weighted Shoenfeld residuals for each predictor. Model 1 included the sociodemographic and clinical variables alone; model 2 included the sociodemographic and clinical variables and the patient symptom ratings; model 3 included the sociodemographic and clinical variables and the clinician symptom ratings; and model 4 included the sociodemographic and clinical variables and both the patient and the clinician symptom ratings.

The predictive accuracy of four Cox models was investigated using Harrell concordance index (C-index; C), which is the area under the receiver-operator curve (ROC curve) adapted for survival data. The C-index estimates the percentage of correct predictions: that is, the percentage of patient pairs in which the predicted (according to the model) and observed (according to the data) survival order are in agreement (concordance), with $C = .5$ for a random model with no variables and $C = 1$ for a perfect order concordance (25). Differences between the C-indices were assessed using the jackknife method (26). All P values less than .05 were considered statistically significant, and all P values were from two-sided tests. In addition to the C-index, we also performed the Concordance Probability Estimate (CPE) measure of discrimination. This measure is based on the method of Gönen and Heller (27) and measures discriminative accuracy in the presence of censoring. SAS (SAS Institute Inc, Cary, NC) software was used for statistical analysis.

Results

Both clinician and patient assessments at baseline could be retrieved for the following six symptoms: pain, fatigue, vomiting, nausea, diarrhea, and constipation. In all of the 14 selected randomized controlled trials, NCI-CTCAE version 2.0 was used for clinician assessments and either EORTC QLQ-C30 version 1, 2, or 3 was used for patient assessments. The three versions of EORTC QLQ-C30 differed with regard to scoring of three parameters that were not our focus (the role functioning scale, the physical functioning scale, and the overall health status scale) and therefore, the version used did not influence the analysis of the symptoms that we studied (28,29). Among the 14 available randomized controlled trials, nine trials, including a total of 1467 patients, reported both clinician and patient assessments of pain. Five of the 14 available randomized controlled trials, involving 1237 patients, reported clinician and patient assessments of fatigue; five trials including 824 patients provided the necessary information on vomiting; six trials including 813 patients had information on nausea; six trials including 815 patients had information on diarrhea; and four trials including 751 patients had information on constipation. The distribution of age, sex, WHO performance status, and distant metastasis status among the participants in the trials evaluated for each symptom were comparable. For each of the symptoms, the percentage of patients who had WHO 0–1 performance status was much higher than the percentage of patients who had a low WHO 2–3 performance status.

This might be explained by the inclusion criteria of the selected trials; in general, patients with low performance status were excluded from participation in EORTC randomized controlled trials; however, the cancer sites varied among the trials that could be used to evaluate each symptom (Table 1).

First, we examined the mean scores and 95% confidence intervals (CIs) from the patient assessments (EORTC QLQ-C30) vs clinician assessments (NCI-CTCAE) for each of the selected symptoms (Table 2). We noticed that reporting of fatigue was highly variable between patient and clinician assessments (patient score = 2.10 vs clinician score = 1.36). However, there was low variability in the reporting of vomiting (patient score = 1.11 vs clinician score = 1.18) and constipation (patient score = 1.50 vs clinician score = 1.11).

For exploratory purposes, we calculated the Spearman correlation coefficient (ρ) and the kappa coefficient (κ) to quantify the level of agreement between the patient- and clinician-reported assessments of adverse event symptoms (Table 3). The Spearman correlation between the clinician and patient scorings was low for diarrhea (ρ = .20) and pain (ρ = .58), indicating weak to moderate relationships between the two scorings. Low kappa coefficients were also reported for some symptoms, for example, fatigue (κ = .07) and pain (κ = .29). According to Landis and Koch (24), these κ values approximate low to fair agreement between the two assessors.

Next, we investigated whether patient and clinician ratings, either alone or together, might predict survival after adjustment for the prognostic sociodemographic variables, age and sex, and

the clinical variables, metastasis, performance status, and cancer site. The C-index, which calculates the accuracy of the four Cox models to predict overall survival (Table 4), shows that for all six symptoms, the relative gain in predictive accuracy of survival estimation compared with a random model with no variables (C = 0.5) is the highest when taking into consideration the prognostic sociodemographic and clinical variables (model 1), indicating their strong predictive value. Models to which patient and clinician symptom ratings were added, either alone or together, had improved predictive accuracy compared with a similar model that used only sociodemographic and clinical variables. When comparing how adding patient-reported scoring (model 2) vs clinician-reported scoring (model 3) next to clinical and sociodemographic variables affected the predictive accuracy of the model, we found statistically significant differences for four of the six symptoms. Patient-reported outcomes were more predictive than clinician-reported outcomes for fatigue (C-index for model 2 vs model 3 = .66 vs .63; P <.001), vomiting (C-index = .64 vs .62; P = .01), nausea (C-index = .65 vs .62; P <.001), and constipation (C-index = .62 vs .61; P = .03). When comparing the effect of fortifying the model with patient plus clinician scoring (model 4) vs the clinician scoring alone (model 3), we found statistically significant differences for the inclusion of the same four symptoms: fatigue (C-index for model 4 vs model 3 = .67 vs .63; P < .001), vomiting (C-index = .64 vs .62; P = .01), nausea (C-index = .65 vs .62; P <.001), and constipation (C-index = .62 vs .61; P = .01). In data not shown, similar results were noticed between the two methods for the CPE measure (27).

Table 1. The clinical and sociodemographic variables for the included patients and trials for the symptoms pain, fatigue, vomiting, nausea, diarrhea, and constipation*

Clinical and sociodemographic variables	Pain, No. (%) (nine RCTs; n = 1467)	Fatigue, No. (%) (five RCTs; n = 1237)	Vomiting, No. (%) (five RCTs; n = 824)	Nausea, No. (%) (six RCTs; n = 813)	Diarrhea, No. (%) (six RCTs; n = 815)	Constipation, No. (%) (four RCTs; n = 751)
Age, y						
≤60	585 (39.9)	827 (66.9)	491 (59.6)	481 (59.2)	487 (59.8)	476 (63.4)
>60	882 (60.1)	410 (33.1)	333 (40.4)	332 (40.8)	328 (40.2)	275 (36.6)
Performance status						
WHO 0-1	1,213 (82.7)	1,118 (90.4)	753 (91.4)	742 (91.3)	745 (91.4)	696 (92.7)
WHO 2-3	254 (17.3)	118 (9.5)	71 (8.6)	71 (8.7)	70 (8.6)	55 (7.3)
Sex						
Male	1,018 (69.4)	731 (59.1)	501 (60.8)	501 (61.6)	497 (61)	430 (57.3)
Female	449 (30.6)	506 (40.9)	323 (39.2)	312 (38.4)	318 (39)	321 (42.7)
Metastasis						
No	183 (12.5)	107 (8.6)	108 (13.1)	107 (13.2)	108 (13.3)	108 (14.4)
Yes	1,214 (82.8)	633 (51.2%)	715 (86.8%)	705 (86.7%)	706 (86.6%)	642 (85.5%)
Cancer site						
Brain	0	496 (40)	0	0	0	0
Colon or rectum	371 (25.3)	370 (29.9)	378 (45.9)	377 (46.4)	375 (46)	376 (50)
Lung	343 (23.4)	0	0	0	0	0
Ovary	69 (4.7)	0	0	0	0	0
Prostate	542 (36.9)	0	69 (8.4)	69 (8.5)	68 (8.3)	0
Breast	82 (5.6)	8 (0.7)	8 (0.9)	0	8 (1)	8 (1.1)
Melanoma	60 (4.1)	304 (24.6)	308 (37.4)	306 (37.6)	303 (37.2)	306 (40.8)
Pancreas	0	59 (4.8)	61 (7.4)	61 (7.5)	61 (7.5)	61 (8.1)

* Individual patient data for the variable "performance status" were unavailable for the symptom fatigue (n = 1), and data for the variable "metastasis" were unavailable for the symptoms pain (n = 70), fatigue (n = 497), vomiting (n = 1), nausea (n = 1), diarrhea (n = 1), and constipation (n = 1). For this reason, percentages do not always add up to 100%. RCT = randomized controlled trial; WHO = World Health Organization.

Table 2. The mean scores and 95% confidence intervals (CIs) for the symptoms pain, fatigue, vomiting, nausea, diarrhea, and constipation*

Clinical symptom	Patient score (EORTC QLQ-C30)†, mean (95% CI)	Clinician score (NCI-CTCAE)‡, mean (95% CI)
Pain	2.31 (2.26 to 2.36)	2.13 (2.07 to 2.18)
Fatigue	2.10 (2.05 to 2.15)	1.36 (1.33 to 1.40)
Vomiting	1.11 (1.08 to 1.14)	1.18 (1.15 to 1.21)
Nausea	1.38 (1.35 to 1.41)	1.20 (1.16 to 1.24)
Diarrhea	1.27 (1.23 to 1.31)	1.10 (1.08 to 1.12)
Constipation	1.50 (1.44 to 1.56)	1.11 (1.09 to 1.14)

* For purposes of comparison, we considered each of the following pairs to be identical responses: EORTC QLQ-C30 score 1 vs NCI-CTCAE score 0; EORTC QLQ-C30 score 2 vs NCI-CTCAE score 1; EORTC QLQ-C30 score 3 vs NCI-CTCAE score 2; EORTC QLQ-C30 score 4 vs NCI-CTCAE scores 3 and 4 combined. EORTC QLQ-C30 = European Organization for Research and Treatment of Cancer's Quality of Life core questionnaire; NCI-CTCAE = National Cancer Institute's Common Terminology Criteria for Adverse Events.

† In the EORTC QLQ-C30 questionnaire, the patient rated his or her symptoms on a 4-point ordinal scale in which a score of 1 meant "not at all," a score of 2 meant "a little," a score of 3 meant "quite a bit," and a score of 4 meant "very much."

‡ In the NCI-CTCAE scoring, the clinician rated the patient's symptoms on a 5-point scale: a score of 0 meant "none or normal," a score of 1 meant "mild," a score of 2 meant "moderate," a score 3 meant "severe," and a score of 4 meant "life threatening or disabling."

Discussion

The goal of this study was to determine whether estimation of overall survival could be improved by including both baseline patient- and/or clinician-reported scores of clinical symptoms in a large heterogeneous dataset of cancer patients. Our results show weak agreement between the patient and clinician ratings and that both ratings separately can make a positive contribution to the predictive accuracy of overall survival prognostication. More specifically, our study demonstrates that including only patient scores (model 2) or including patient scores combined with clinician scores (model 4) provides a statistically significant relative gain in predictive accuracy for the symptoms fatigue, vomiting, nausea,

and constipation compared with including clinician scores alone (model 3).

Although the inclusion of data regarding diarrhea contributed positively to survival estimation, low variability between patients at baseline might explain why including data on diarrhea confers low predictive value. There was relatively low baseline interpatient variability of diarrhea data in both patient-reported scores (SD = 0.58; 95% CI = 1.23 to 1.31) and clinician-reported scores (SD = 0.36; 95% CI = 1.08 to 1.12) compared with the other symptoms that we studied. Low interpatient variability at baseline was also reported for nausea and vomiting. Diarrhea, nausea, and vomiting are more strongly related to cancer treatment (eg, chemotherapy and/or radiotherapy) than to the presence of cancer itself, and therefore, baseline P values are lower for these symptoms compared with the other symptoms that we studied. It is possible that diarrhea, nausea, and vomiting are better predictors for survival when ratings are taken throughout the course of treatment.

We found low agreement between clinician and patient scores for diarrhea and nausea. An explanation might be that for the clinician using the NCI-CTCAE scoring system, these symptoms can be measured by an exact number of the amount of stools or incidences of emesis that a patient experienced. However, for patients, diarrhea and nausea may have broader effects on their well-being, which are reflected in their scores [30]. Low correlation between the clinician and patient ratings might be explained by the use of two different approaches to measurement; this may imply that scoring behavior is imposed by the specific questions asked by the measurements. This idea has been discussed by Gotay [31], and it is exemplified in our study by the discrepancy in the EORTC QLQ-C30 pain and fatigue scores. Upon examining our EORTC QLQ-C30 pain scores, we can see that the clinician score for cancer pain is more highly correlated with the question "Have you had pain?" referring to the pain intensity, than with the question "Did pain interfere with your daily activities?" which refers to the functional disability due to the pain. Upon examining our EORTC QLQ-C30 fatigue scores, we also notice that the agreement levels with clinician-reported scores are somewhat different depending on the specific questions asked of patients.

Table 3. Spearman (ρ) and kappa (κ) correlation coefficients for comparisons between clinician and patient assessments of pain, fatigue, vomiting, nausea, diarrhea, and constipation*

Clinician (NCI-CTCAE)	Patient (EORTC QLQ-C30)	ρ†	κ (95% confidence interval)‡
Pain	Have you had pain?	0.58	0.29 (0.26 to 0.33)
	Did pain interfere with your daily activities?	0.50	0.27 (0.23 to 0.30)
Fatigue	Did you need to rest?	0.30	0.07 (0.03 to 0.10)
	Have you felt weak?	0.28	0.07 (0.03 to 0.10)
	Were you tired?	0.30	0.08 (0.04 to 0.11)
Vomiting	Have you vomited?	0.32	0.22 (0.13 to 0.30)
Nausea	Have you felt nauseated?	0.32	0.14 (0.10 to 0.18)
Diarrhea	Have you had diarrhea?	0.20	0.14 (0.07 to 0.20)
Constipation	Have you been constipated?	0.38	0.16 (0.11 to 0.21)

* EORTC QLQ-C30 = European Organization for Research and Treatment of Cancer's Quality of Life core questionnaire; NCI-CTCAE = National Cancer Institute's Common Terminology Criteria for Adverse Events.

† As suggested by Cohen [22], a ρ value that is less than .30 indicates a weak relationship, a value between .30 and .50 indicates a moderate relationship, and a value greater than .50 indicates a strong relationship.

‡ As suggested by Landis and Koch [24], a κ value less than 0.00 mean poor agreement; κ between .00 and .20 slight agreement; κ between .21 and .40 fair agreement; κ between .41 and .60 moderate agreement; κ between .61 and .80 substantial agreement; and κ between .81 and 1.00 almost perfect agreement.

Table 4. C-index from four Cox models to compare the predictive accuracy for overall survival using patient- and/or clinician-reported symptoms and *P* values for comparisons between models for pain, fatigue, vomiting, nausea, diarrhea, and constipation*

Clinical symptom	C-index				$P\P_{\text{model 2 vs model 3}}$	$P\P_{\text{model 3 vs model 4}}$
	Model 1†	Model 2‡	Model 3§	Model 4‖		
Pain	.60	.62	.62	.63	.59	.44
Fatigue	.60	.66	.63	.67	<.001	<.001
Vomiting	.62	.64	.62	.64	.01	.01
Nausea	.62	.65	.62	.65	<.001	<.001
Diarrhea	.62	.62	.62	.62	.08	.52
Constipation	.60	.62	.61	.62	.03	.01

* Four different Cox models were analyzed for each symptom, as described below. The C-index calculates the predictive accuracy from the Cox model to predict overall survival. (The C-index estimates the percentage in which the correct predictions derived from the Cox model and the observed survival order are in agreement, with *C* = 0.5 for a random model with no variables and *C* = 1 with a perfect order concordance.)

† Model 1 includes the sociodemographic and clinical variables alone.

‡ Model 2 includes the sociodemographic and clinical variables and patient symptom rating.

§ Model 3 includes the sociodemographic and clinical variables and clinician symptom rating.

‖ Model 4 includes the sociodemographic and clinical variables and both patient and clinician symptom ratings.

¶ *P* values are from two-sided tests. Each *P* value tests the hypothesis that the predictions from one Cox model, indicated by the C-index is more concordant with the observed outcome than predictions from the other Cox model, indicated by the C-index, within paired predictions. In this analysis, we compared model 2 vs model 3 and model 3 vs model 4.

We found a statistically significant difference between patient vs clinician reporting of fatigue and pain. As Basch et al. (32) suggest, clinicians may refrain from using the high end of the pain rating continuum until a patient's disease progresses, in contrast with the patient, who has only his or her own current experience to use in making assessments of pain severity. On the other hand, patients may also minimize their symptoms, fearing that their active treatment might be stopped if their clinical condition appears worse (10).

Our study does have potential limitations. One might speculate that the statistically significant differences between the C-indices are driven by the large sample size because the symptom scores make a relatively small contribution to predict overall survival compared with clinical and sociodemographic variables. Therefore, statistical significant differences in the prediction accuracy of different Cox models might not always translate into meaningful changes between the predictive value of the models, and further validation studies may be needed. Second, the large sample of different cancer sites did not allow us to include cancer-specific prognostic variables that might be identified as key independent prognostic factors for survival (33–36). Another limitation to our study is that there is no evidence-based consensus regarding how to compare the scoring derived from the patient- vs clinician-reported measurements. Based on the NCI-CTCAE manual (37), the NCI-CTCAE criteria should be used to assess any unfavorable symptom, sign, or disease that is temporarily associated with the use of a medical treatment, and it should not be used to grade the cancer itself. As such, baseline data are normally used to establish toxicity at the earliest time point to measure changes due to treatment. So, the normal purpose of the NCI-CTCAE baseline assessment differs from the EORTC QLQ-C30, in which the patient is asked to assess his quality of life more in general, independent of treatment. These different purposes of assessment might provide some rationale as to why different scores and subsequently low levels of agreement are reported between patients and clinicians at baseline. Last, the fact that our data was collected from randomized controlled trials might have an impact on the generalization of our findings. Randomized controlled trials usually exclude elderly patients, patients with comorbidities, and patients who have severe symptom burdens at the time of study entry, and therefore our study was limited to a relative asymptomatic population. Therefore, we invite the replication of our findings in a broader population of cancer patients or in observational studies.

Because the use of NCI-CTCAE scoring by clinicians has been accepted as a tool to rate the symptom burden of cancer, this study highlights the importance of patient-reported outcomes as subjective measures of the severity of symptoms that a cancer patient may experience. The use of NCI-CTCAE criteria to assess the prognostic value of symptoms should be questioned, given the limitation of the NCI-CTCAE as a sound psychometric instrument to value the symptom of a cancer burden as well as its inconsistency in application and evaluation. In addition, patient-reported outcome tools like the EORTC QLQ-C30, which also assess psychosocial and functional aspects of the cancer patient's experience, allow for a broader assessment of the cancer burden than simple reports of symptoms. Patient-reported outcomes provide relevant information for cancer patients undergoing treatment, especially with regard to supportive and palliative care, symptom management, and new treatments that extend survival (38). They are also useful in assisting health-care professionals to distinguish between treatments, and if needed, to modify therapeutic regimens.

Patient-reported outcomes, as an approved psychometric test and valid tool, might also play a vital role in clinical research. Stratifying at baseline for patients at higher risk for developing symptoms might improve the reliability of trial findings. Patient-reported symptoms may also have prognostic value; this use for patient reports should be assessed in upcoming clinical studies. Our study has shown that patient-reported outcomes add value to prediction models when combined with clinicians' assessments. Also, patient-reported outcomes should be used to rate symptom adverse events when monitoring patients in study trials, because it will improve clinical decision making in defining the best treatment options for patients. Similar suggestions have been made by Gotay et al. (16).

Our study has indicated agreement at baseline; however, patient-reported outcomes should also be assessed on a longitudinal basis. We therefore encourage future work in patient-reported outcome assessment on a longitudinal scale to monitor adverse events and to include patient-reported outcomes and patient-reported outcome trajectories to support clinical decision making.

These elements have also been recognized by the NCI, which in 2008 launched the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) initiative to systematically collect and integrate patients' ratings of their symptoms for adverse event reporting in cancer trials (39), reconfirming the general belief that patients themselves provide a unique and needed perspective in rating their own symptom burden and quality of life. The PRO-CTCAE system provides a web-based platform to collect patient reports of symptoms they are experiencing while undergoing treatment. Currently, 81 symptoms are represented in the CTCAE (version 4.0) and have been identified to be amenable to patient reporting. This system, when complete, will enable the collection of complementary data from patients and clinicians (31) to provide information on symptom experience to guide cancer care.

References

1. National Cancer Institute. *Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0.* Bethesda, MD: U.S. Department of Health and Human Services; 2009.
2. Bruner DW, Movsas B, Konski A, et al. Outcomes research in cancer clinical trial cooperative groups: the RTOG model. *Qual Life Res.* 2004;13(6):1025–1041.
3. Moinpour CM, Lovato LC. Ensuring the quality of quality of life data: the Southwest Oncology. *Stat Med.* 1998;17(5–7):641–651.
4. Bottomley A. Developing clinical trials protocols for quality-of-life assessment. *Appl Clin Trials.* 2001;10(1):40.
5. Fromme EK, Eilers KM, Mori M, et al. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J Clin Oncol.* 2004;22(17):3485–3490.
6. Petersen M, Larsen H, Pedersen L, et al. Assessing health-related quality of life in palliative care: comparing patient and physician assessment. *Eur J Cancer.* 2006;42(8):1159–1166.
7. Greimel ER, Bjelic-Radisic V, Pfisterer J, et al. Toxicity and quality of life outcomes in ovarian cancer patients participating in randomized controlled trials. *Support Care Cancer.* 2010;19(9):1421–1427.
8. 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(12):903–909.
9. Blazeby JM, Williams MH, Alderson D, et al. Observer variation in assessment of quality of life in patients with oesophageal cancer. *Br J Surg.* 1995;82(9):1200–1203.
10. Wilson A, Dowling A, Abdolell M, et al. Perception of quality of life by patients, partners and treating physicians. *Qual Life Res.* 2000;9(9):1041–1052.
11. Von Essen L. Proxy ratings of patient quality of life. *Acta Oncol.* 2004;43(3):229–234.
12. Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. *J Clin Epidemiol.* 1992;45(7):743–760.
13. Varricchio CG, Sloan JA. The need for and characteristics of randomized, phase III trials to evaluate symptom management in patients with cancer. *J Natl Cancer Inst.* 2002;94(16):1194–1185.
14. Bruner DW. Should patient-reported outcomes be mandatory for toxicity reporting in cancer clinical trials? *J Clin Oncol.* 2007;25(34):5345–5347.
15. Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses for terminally ill patients: prospective cohort study. *BMJ.* 2000;320(7233):469–473; reprinted in *West J Med.* 172(5):310–313.
16. Gotay CC, Kawamoto CT, Bottomley A, et al. The prognostic significance of patient-reported outcomes in cancer clinical trials. *J Clin Oncol.* 2006;26(8):1355–1363.
17. Quinten C, Coens C, Mauer M, et al. Baseline quality of life as prognostic indicator of survival: a meta-analysis of individual patient data from EORTC clinical trials. *Lancet Oncol.* 2009;10(9):865–871.
18. Tan AD, Novotny PJ, Kaur JS, et al. A patient-level meta-analytic investigation of the prognostic significance of baseline quality of life (QOL) for overall survival (OS) among 3,704 patients participating in 24 North Central Cancer Treatment Group (NCCTG) and Mayo Clinic Cancer Center (MC) oncology clinical trials. *J Clin Oncol.* 2008;26(15S) (20 suppl). Abstract 9515.
19. 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):365–364.
20. Fayers PM, Aaronson NK, Bjordal K, et al. *The EORTC QLQ-C30 Scoring Manual.* 3rd ed. Brussels, Belgium: European Organisation for Research and Treatment of Cancer; 2001.
21. National Cancer Institute. *Common Toxicity Criteria, Version 2.0.* Bethesda, MD: U.S. Department of Health and Human Sciences; 1999.
22. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
23. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20(1):37–46.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):189–196.
25. Harrel F. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression and Survival Analysis.* New York, NY: Springer; 2001.
26. Kremers WK. *Concordance for Survival Time Data: Fixed and Time-Dependent Covariates and Possible Ties in Predictor and Time.* Technical Report Series no. 80. Rochester, MN: Department of Health Sciences Research and The William J. von Liebig Transplant Center, Mayo Clinic; 2007.
27. Gönen KK, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika.* 2005;92(4):965–970.
28. Bjordal K, de Graeff A, Fayers PM, et al. A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (The EORTC QLQ-H&N35). *Eur J Cancer.* 2000;36(14):1796–1807.
29. Osoba D, Aaronson NK, Zee B, et al. Modification of the EORTC QLQ-C30 (version 2.0) based upon content validity and reliability testing in large samples of patients with cancer. *Qual Life Res.* 1997;6(2):103–108.
30. Huschka M, Burger K. Does QOL provide the same information as toxicity data? *Curr Prob Cancer.* 2006;30(6):244–254.
31. Gotay CC. Patient symptoms and clinician toxicity ratings: both have a role in cancer care. *J Natl Cancer Inst.* 2009;101(23):1602–1603.
32. Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients versus clinicians: relationship with clinical outcomes. *J Natl Cancer Inst.* 2009;101(23):1624–1632.
33. Mauer M, Taphoorn MJB, Bottomley A, et al. Prognostic value of health-related quality-of-life data in predicting survival in patients with anaplastic oligodendrogliomas, from a phase III EORTC Brain Cancer Group Study. *J Clin Oncol.* 2007;25(36):5731–5737.
34. Efficace E, Bottomley A, Coens C, et al. Does a patient's self-reported health-related quality of life predict survival beyond key biomedical data in advanced colorectal cancer? *Eur J Cancer.* 2006;42(1):42–49.
35. Ryberg M, Nielsen D, Østerlind K, et al. Prognostic factors and long-term survival in 585 patients with metastatic breast cancer treated with epirubicin-based chemotherapy. *Ann Oncol.* 2001;12(1):81–87.
36. Bottomley A, Coens C, Efficace F, et al. Symptoms and patient-reported outcomes in malignant pleural mestothelioma? A prognostic factor analysis of EORTC-NCIC 08983: randomized phase III study of cisplatin with or without raltitrexed in patients with malignant pleural mesothelioma. *J Clin Oncol.* 2007;25(36):5770–5776.
37. National Cancer Institute. *Cancer Therapy Evaluation Program. Common Toxicity Criteria Manual.* Bethesda, MD; 1999.

38. Trask PC, Hsu MA, McQuellon R. Other paradigms: health-related quality of life as a measure in cancer treatment: its importance and relevance. *Cancer J.* 2009;15(5):435–440.

39. National Cancer Institute. *U.S. Department of Health and Human Services.* https://wiki.nci.nih.gov/x/CKul. Accessed November 30, 2010.

Affiliations of authors: Quality of Life Department, EORTC Headquarters, Brussels, Belgium (CQ, JM, FM, CCo, AB); School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada (CCG); British Columbia Cancer Research Centre, Vancouver, BC, Canada (CCG); Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC (BBR); Clinic for Child and Adolescent Psychiatry and Psychotherapy, University of Magdeburg, Magdeburg, Germany (HF); Department of Obstetrics and Gynecology, Medical University Graz, Graz, Austria (EG); Psycho-oncology Co-operative Research Group, School of Psychology, University of Sydney, NSW, Australia (MK); Quality of Life Consulting, West Vancouver, BC, Canada (DO); Department of Symptom Research, University of Texas M.D. Anderson Cancer Center, Houston, TX (CCl); The Princess Margaret Hospital, University of Toronto, Toronto, ON, Canada (JR); Medical Ethics, Department of Philosophy, University of Regensburg, Regensburg, Germany (JS-VK); Haaglanden Medical Center, The Hague, the Netherlands (MJBT); VU University Medical Center, Amsterdam, the Netherlands (MJBT); Clinic of Tumor Biology, University of Freiburg, Freiburg, Germany (JW).