HHS Public Access

Author manuscript

J Pain Symptom Manage. Author manuscript; available in PMC 2016 October 01.

Published in final edited form as:

J Pain Symptom Manage. 2015 October ; 50(4): 470–479.e9. doi:10.1016/j.jpainsymman.2015.04.016.

The Complementary Nature of Patient-Reported Outcomes (PROs) and Adverse Event Reporting in Cooperative Group Oncology Clinical Trials: A Pooled Analysis (NCCTG N0591)

Pamela J. Atherton, MS, **Deborah W. Watkins-Bruner, PhD, RN**, **Carolyn Gotay, PhD**, **Carol M. Moinpour, PhD**, **Daniel V. Satele, BS**, **Kathryn A. Winter, MS**, **Paul L. Schaefer, MD**, **Benjamin Movsas, MD**, and **Jeff A. Sloan, PhD**

Alliance Statistics and Data Center (P.J.A., D.V.S., J.A.S.), Mayo Clinic, Rochester, Minnesota; Nell Hodgson Woodruff School of Nursing (D.W.W.-B.), Emory University, Atlanta, Georgia; School of Population and Public Health (C.G.), University of British Columbia, Vancouver, British Columbia, Canada; Public Health Sciences Division (C.M.M.), Fred Hutchinson Cancer Research Center, Seattle, Washington; Statistical Department (K.A.W.), Radiation Therapy Oncology Group, Philadelphia, Pennsylvania; Toledo Community Hospital Oncology Program (P.L.S.), Toledo, Ohio; and Department of Radiation Oncology (B.M.), Henry Ford Medical Center, Detroit, Michigan, USA

Abstract

Context—Clinical trials utilize clinician-graded adverse events (AEs) and patient-reported outcomes (PROs) to describe symptoms.

Objectives—To examine the agreement between PROs and AEs in the clinical trial setting.

Methods—Patient-level data were pooled from seven North Central Cancer Treatment Group, two Southwest Oncology Group and three Radiation Therapy Oncology Group lung studies that included both PROs and AE data. Ten-point changes (on a 0–100 scale) in PRO scores were considered clinically significant differences (CSDs). PRO score changes were compared to AE grade (Gr) categories (2+ yes vs. no and 3+ yes vs. no) using Wilcoxon rank-sum or two-sample *t*-tests between Gr categories. Incidence rates and concordance of CSD in PRO scores and AE grade categories were compiled. Spearman correlations were computed between PRO scores and AE severity.

Results—PROs completed by patients (*N*=1013) were the Uniscale, Lung Cancer Symptom Scale (LCSS), Functional Assessment of Cancer Therapy-Lung (FACT-L), Symptom Distress Scale (SDS), and/or Functional Living Index-Cancer (FLIC). Significantly worse PRO score changes were found for the FACT-L in patients with Gr 2+ AEs. Worse scores were seen for the

Address correspondence to: Pamela J. Atherton, MS, Mayo Clinic, 200 1st Street SW, Harwick 836, Rochester, MN 55905, USA, atherton@mayo.edu.

Disclosures

The authors declare no conflicts of interest.

Uniscale for patients with grade 2+ AEs (*P*=0.07) and LCSS for patients with Gr 3+ AEs (*P*=0.09). Agreement between incidence of any Gr 2+(Gr 3+) AE and a CSD in PROs ranged from 27%–67% (36%–61%). Correlations between PRO scores and AE severity were low: −0.06 Uniscale, −0.03 LCSS, 0.10 FACT-L, −0.11 SDS and −0.51 FLIC.

Conclusion—These results support previous work and an a priori hypothesis that AEs and PROs measure differing aspects of the disease experience and are complementary.

Keywords

Patient reported-outcomes; adverse events; clinical trials

Introduction

Prior to 2005, there was little exploration of the relationships between patient-reported outcomes (PROs) and other data routinely collected as part of randomized clinical trials, such as the Common Toxicity Criteria (CTC) and the Common Terminology Criteria for Adverse Events (CTCAE) (1–3). As standard practice, adverse events (AEs) are collected in cases where the clinician actively asks a patient about a particular AE, the clinician makes an inference based on the patient/clinician interaction, the clinician observes the patient, or the patient independently volunteers information that an AE has occurred. The degree of distress that an AE imparts to a patient is not collected routinely (4). Cleeland et al. (5) suggest that the inclusion of PROs to capture patient-perceived AEs and their associated burden or distress in the clinical trial setting is becoming the norm.

A growing body of literature has documented that PRO scores and CTCs are correlated at only modest levels. For example, an analysis of three North Central Cancer Treatment Group (NCCTG) trials of symptom control regimens (total *N*=121) found a number of discrepancies between CTC ratings and PROs, e.g., 10% of patients with no CTC-reported diarrhea reported four or more diarrhea-related problems on the bowel function questionnaire, 4% reported rectal bleeding on the questionnaire without a corresponding CTC toxicity rating, and 14% of lung cancer patients (total *N*=106) reported fatigue with no CTC-recorded fatigue (6). Another NCCTG meta-analysis comparing Skindex-16 results to CTCAE grades determined that there were 855 instances where patients reported skin itching, burning/stinging, hurting or irritation when the physician recorded no AEs (7).

A Radiation Therapy Oncology Group (RTOG) trial of quality of life (QOL) during and after treatment for prostate cancer found that disagreement ranged from 13% to 45% at three months between patient self-reports of symptoms measured by the Functional Assessment Cancer Therapy (FACT) QOL scale and physician ratings on the RTOG acute toxicity rating scale of the same symptoms (8). Another RTOG trial examined sexual outcomes following radiotherapy ± androgen deprivation therapy in prostate cancer patients and showed physician and patient ratings of the patient's ability to have an erection differed up to 47% (9). This lack of agreement between patient-rated and physician-rated outcomes is consistent with other literature that has demonstrated a general lack of concordance between cancer patient and proxy ratings (10). More recently, Basch and colleagues developed the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events

(PRO-CTCAE) after demonstrating that clinicians are better at more objective physical assessments like rash or vomiting, and patients are better at assessing internally subjective assessments like hot flashes, nausea, pain and itching (1).

The question remains as to what distinctive information is provided by PROs relative to toxicity data. Morton et al. (11) presented results from the NCCTG/Intergroup protocol 9741, a study in non-small cell lung cancer, which indicated patients reported peripheral neuropathy using PROs two to three months earlier than providers using CTCs. Huschka et al. (4) also demonstrated in a series of NCCTG lung cancer clinical trials that PROs were able to detect clinically meaningful AEs on an array of symptoms (nausea, vomiting, etc.) earlier and with greater frequency than the CTC (4).

The intent of this study is to examine the degree of redundancy – or lack thereof – between PRO and toxicity information. This intergroup collaborative protocol (NCCTG N0591) was designed to combine data from three cooperative groups: NCCTG, RTOG and the Southwest Oncology Group (SWOG). The N0591 protocol was approved by the Institutional Review Board of the NCCTG and was followed by the NCCTG Data Safety Monitoring Board. This patient-level pooled analysis reflecting experiences across three cooperative groups, which have considerable variability in terms of systems and procedures, provided an opportunity to achieve sufficient sample size with varied patient populations and experimental settings to test this research question. Specifically, the question addressed was: How well do patient-reported symptoms correspond to reports of the same symptoms as rated by the CTC?

Methods

Lung trials in which both AE criteria and PROs were utilized to measure AEs and toxicity were identified from the NCCTG, RTOG and SWOG (Table 1). There were one pilot, one phase I/II, six phase II, and four phase III trials. AEs were recorded using the CTC v2.0, the RTOG Cooperative Group CTC (12) or the SWOG Toxicity Criteria (13). All patients provided informed consent upon individual study enrollment.

Measures

Five PRO assessment tools in these trials were utilized for analysis.

1. Overall QOL was measured using the Spitzer Uniscale or a single question within a multiple item scale. The visual analogue version of the Spitzer Uniscale has been modified in recent studies to a numeric scale with values of 0 to 10, without loss of validity or reliability (14).
2. The Functional Assessment of Cancer Therapy-Lung Questionnaire (FACT-L) is composed of a 27 question FACT-General component assessing four dimensions of QOL: physical, social and family, emotional, and functional well-being plus a nine-question lung component evaluating nine specific lung cancer-related additional concerns. Each question is evaluated on a 0 (not at all) to 4 (very much) scale. Each subscale score and the total score are computed by summing the responses.

Reliability, validity, and factor structure of the scale have been documented for cancer patients (15–18).

3. The Functional Living Index-Cancer (FLIC) instrument has been shown to be valid and reliable for use in cancer (19, 20). It contains 22 items that use visual analogue scales to assess the effect of the symptoms of cancer and its treatment on functional ability in all areas of life. The FLIC assesses body care, household maintenance, physical exercise, recreation, spiritual activities, and social activities. Evidence supports its high internal consistency, reproducibility in stable groups, and predictive validity for survival in patients with metastatic breast cancer (21).

4. The Lung Cancer Symptom Scale (LCSS) was designed as a site-specific measure containing nine items evaluated on a visual analogue scale that can be categorized by two subscales: symptom burden and QOL. Reliability and validity have been documented (22–25).

5. The Symptom Distress Scale (26) is a valid and reliable 13-item cancer-specific instrument intended for assessing the degree of distress associated with cancer symptomatology (27) and has been shown to be prognostic for survival (28).

Complete forms are included in the Appendix (available at jpsmjourna.com). Patients completed assessments at baseline prior to study treatment and at least one time post-baseline. If a patient did not complete an overall QOL assessment, but completed a FACT-L assessment, the question from the FACT-L “I’m content with my overall quality of life right now” was used as a surrogate for the overall QOL. All assessments were scored according to the appropriate scoring algorithm described by the questionnaire developers. For ease of comparability across measures, all scores were converted to a 0–100 point scale where 100 indicated the best QOL (14). Changes from baseline were calculated and decreases in scores (or worsening) of at least 10 points were categorized as clinically significant differences (CSDs) (29–33). AE information was gathered for 10 clinician- reported toxicities that could be mapped to patient-reported symptoms: anorexia, confusion, constipation, diarrhea, dyspnea, fatigue, nausea, pain-arthralgia, pain-headache, and pain. These symptoms were chosen as they are among the most prevalent symptoms experienced by cancer patients (34). Patients were assigned binomial outcomes for CTC grade 2+ and grade 3+ toxicity incidence.

Statistical Analysis

Summary statistics were calculated to describe the patient population PRO scores and toxicity grades. Associations between changes in PRO scores and toxicity incidence were compared using two sample *t*-tests and Wilcoxon rank-sum tests as appropriate. The procedures had greater than 90% power to detect a 10-point difference between group PRO averages. Cross tabulation was performed to determine percent agreement between toxicity incidence and CSD incidence. Correlation statistics were calculated for PRO scores and maximum AE grades. The criteria published by Cohen were used for interpreting the size of a correlation; specifically, correlations from 0.10 to 0.29 were considered low, correlations from 0.30 to 0.49 were considered moderate, and correlations greater than 0.5 were considered high (35).

Results

Data from 1013 patients were compiled from the individual trials (seven NCCTG, two RTOG and three SWOG) (Table 1) (36–46). Patients completed one or more of the Uniscale (*N*=770), LCSS (*N*=132), FACT-L (*N*=347), SDS (*N*=53), and FLIC (*N*=16). Baseline characteristics are reported in Table 2. The majority of the patients were white (88.4%) and male (62.6%). The mean age was 65 years, 78.2% of the patients were currently receiving chemotherapy and 11% were currently receiving radiation therapy. Frequencies of recorded AEs are shown in Tables 3a and 3b. Table 3a reports the maximum grade of each AE per patient during the study and Table 3b reports the grades of all AE incidences during the study. The most prevalent toxicity for patients with grade 2+ AEs was fatigue (30%) followed by nausea (29%) and dyspnea (24%). Confusion was not present in most patients.

Associations of the change from baseline in each PRO score per person to AE grade 2+ and 3+ incidence is reported in Tables 4a and 4b, respectively. The mean decline in FACT-L total score was significantly larger (i.e., patient had worsening QOL) for patients experiencing any grade 2+ AE ($P<0.01$). Uniscale score decline was larger for those experiencing any grade 2+ AE (non-statistically significant $P=0.07$). Patients experiencing any grade 3+ AE experienced worse LCSS total scores (non-statistically significant $P=0.09$).

Patient AE incidence and PRO scores were compared using the CSD criteria for the 10 most prevalent symptoms (Table 5). Incidence rates of a CSD in PRO score were compared between patients having a grade 2+ AE and between patients having a grade 3+ AE. For example, clinicians recorded grade 2+ anorexia in 13% of the patients who completed the FACT-L “losing weight” question, but 43% of patients were categorized as having a CSD on the “losing weight” question score. There were 20 patients who had both a grade 2+ anorexia and a CSD in “losing weight” score leading to a 56% agreement in the two measures. The agreement percent includes both situations where the clinician rating and patient score agree (the grade 2+ AE with a CSD in PRO and no grade 2+ AE with no CSD in PRO). Similarly there was a 56% agreement in grade 3+ anorexia and “losing weight” score.

In addition to the FACT-L “losing weight” question, anorexia was measured by three other questions from three different assessments: the FACT-L item “good appetite”, the LCSS item “appetite”, and the SDS appetite question. Data for the FACT-L “good appetite” were similar to the “losing weight” item (data not shown). The highest agreement in patient-reported and clinician-defined AEs was 73% for grade 2+ and SDS “appetite”. Dyspnea was specifically measured by four questions from two assessments: the FACT-L items for “shortness of breath” (data not shown as it is similar to “tightness in chest”), “tightness in chest”, and “ease of breathing” and the LCSS item “shortness of breath”. Data for FACT-L “shortness of breath” were similar to ‘tightness in chest” so data are not shown. Agreement rates ranged from 55% (grade 2+ and CSD in FACT-L “ease of breathing” and LCSS “shortness of breath”) to 64% (grade 3+ and CSD in FACT-L “tightness in chest”).

Fatigue was measured with three questions from three assessments: the FACT-L item “energy”, LCSS item “fatigue”, and SDS item “fatigue”. Agreement rates ranged from 47%

(grade 3+ and CSD in FACT-L “energy”) to 61% (grade 2+ and CSD in LCSS “fatigue”). Nausea also was measured with three questions: the FACT-L “nausea” and the SDS “nausea incidence” and “nausea severity”. Agreement rates ranged from 50% (grade 2+ and CSD in SDS “nausea severity”) to 73% (grade 3+ and CSD in SDS “nausea incidence”).

Confusion was measured using the FACT-L “clear thinking” question and the SDS “concentration” question. The agreement rate for the FACT-L “clear thinking” was 57% for both the grade 2+ and 3+. Similarly, the agreement rate for the SDS item “concentration” was 59% for both grade categories. Constipation and diarrhea were both assessed by the SDS “bowel” question. The agreement rates for constipation were 77% for grade 2+ and 68% for grade 3+. The agreement rates for diarrhea were 59% for grade 2+ and 63% for grade 3+.

Pain was collected via the CTCAE using the symptoms: arthralgia, headache and pain. The FACT-L item “physical functioning” and SDS item “pain severity” assessed pain. Agreement rates for arthralgia were 41% for grade 2+ and 40% for grade 3+ in the FACT-L and 67% for grade 2+ and 57% for grade 3+ in the SDS. Agreement rates for headache were 40% for grade 2+ and 40% for grade 3+ in the FACT-L and 62% for grade 2+ and 57% for grade 3+ in the SDS. Agreement rates for pain were 44% for grade 2+ and 42% for grade 3+ in the FACT-L and 52% for grade 2+ and 52% for grade 3+ in the SDS.

All AEs were compared to the Uniscale responses in the same fashion. The agreement in AE grade 2+ and QOL ranged from 45% to 52%. The agreement in AE grade 3+ and QOL ranged from 44% to 46%.

Correlations between PRO scores and maximum AE severity per week on study were uniformly low. Pearson correlation coefficients of −0.06 were observed for AE severity with the Uniscale, −0.03 for AE severity with the LCSS, 0.10 for AE severity with the FACT-L, −0.11 for AE severity with the SDS and −0.51 for AE severity with the FLIC.

Discussion

Combining data from three cooperative groups identified that significantly greater decreases in PRO scores were related to AE grade incidences for individuals who experienced grade 2+ or grade 3+ AEs. In addition, similarly low correlations between PROs and AE grades were discovered, as has been found in past research (4,7). The low correlation may be a reflection of the minimal AE reporting required of the clinician when utilizing the CTC version 2.0 where clinicians are instructed not to report disease symptoms (47). But, these results support prior studies’ conclusions that there is an advantage of including the patient’s perspective on AE assessments, and this study’s findings are consistent with Basch et al. (1), demonstrating that the inclusion of both PROs and clinician-reported assessments are complementary and can more fully document the burden of toxicities and symptoms (1).

Results for agreement between PRO and clinician ratings were relatively consistent across all four PRO assessments (Uniscale, Fact-L, LCSS, and SDS). Agreement was typically between 40–60%, with a few results in the 70% range. Similarly, agreement across the 10 symptoms was consistent. Hence, for a broad range of situations, we can expect the general

tenet to hold that PROs add substantial information to clinician ratings. This is somewhat surprising because the four tools and 10 symptoms represent different psychometric approaches and different clinical issues. However, there is evidence in the literature supporting this as described in the introduction (e.g., most recently Basch in 2013). While there may be specific situations where a PRO or a clinician rating has more relevance for clinical care, the evidence both in this study and others indicates that in the majority of situations (if not all), both PRO and clinician ratings should be included.

This study had some limitations. Most notably, there were missing data. Not all studies had AE evaluations at the same time as PRO assessments, and not all studies utilized the same PRO assessments nor evaluated them at the same time points. The SWOG studies, in particular, had only maximum toxicity grade reported, not cycle based. These limitations resulted in subsets of appropriate data being utilized for various analyses, rather than using the entire study population. We did have a heterogeneous population regarding treatment and treatment length, which impedes our ability to specifically describe a subpopulation.

As reported by Huschka (4), more research is needed regarding real-time feedback of clinically meaningful changes in patient-reported symptoms and a delineation of the clinical pathways that need to be followed to address these symptoms. Recently studies have been, or are being, conducted utilizing real-time collection of QOL data (48–52), providing data indicating benefit of electronic symptom monitoring by both the patient and clinician. Technology also has provided for the validation of the PRO-CTCAE (53). The creation of the PRO-CTCAE reflects both the need for incorporating the patient perspective and the means to do so. Our data provide examples of how PRO data add value to physician CTCAEs yet are complementary; for example, the increased worsening indicated in PRO scores could indicate the need for symptom interventions at an earlier time point. Outstanding questions remain, however, such as identifying optimal and efficient ways to efficiently synthesize information from patients and clinicians for real-time reporting and feedback.

This multi-institutional patient-level pooled analysis indicated a low correlation among PRO scores and related AEs. These results support our previous work and a priori hypothesis that physician-reported AEs and PROs do not measure all of the same aspects of the disease experience and are complementary, both providing information that is useful for cancer patient management.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This trial was conducted by the Radiation Therapy Oncology Group (RTOG), and was supported by RTOG grant U10 CA21661 and CCOP grant U10 CA37422 from the National Cancer Institute. This manuscript's contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

References

1. Basch E, Reeve B, Cleeland C, et al. Development of the patient-reported version of the common terminology criteria for adverse events (PRO-CTCAE). Value Health. 2010; 13:A274–A275.
2. Bruner, DW.; Hanisch, LJ.; Reeve, BB., et al. Perceived barriers to implementing the patient-reported outcomes-common toxicity criteria adverse event (PRO-CTCAE) system in cancer clinical trials. Qual Life Res; International Society of Quality of Life Research 2010 Conference Abstracts; 2010. p. 1006
3. Basch E. Toward patient-centered drug development in oncology. N Engl J Med. 2013; 369:397–400. [PubMed: 23822654]
4. Huschka MM, Mandrekar SJ, Schaefer PL, Jett JR, Sloan JA. A pooled analysis of quality of life measures and adverse events data in north central cancer treatment group lung cancer clinical trials. Cancer. 2007; 109:787–795. [PubMed: 17211864]
5. Cleeland CS, Sloan JA, Cella D, et al. CPRO (Assessing the Symptoms of Cancer Using Patient-Reported Outcomes) Multisymptom Task Force. Recommendations for including multiple symptoms as endpoints in cancer clinical trials: a report from the ASCPRO (Assessing the Symptoms of Cancer Using Patient-Reported Outcomes) Multisymptom Task Force. Cancer. 2013; 119:411–420. [PubMed: 22930243]
6. Varricchio CG, Sloan JA. The need for and characteristics of randomized, Phase III trials to evaluate symptom management in patients with cancer (editorial). J Natl Cancer Inst. 2002; 94:1184–1185. [PubMed: 12189214]
7. Atherton PJ, Burger KN, Loprinzi CL, et al. Using the Skindex-16 and CTCAE to assess rash symptoms: results of a pooled-analysis (N0993). Support Care Cancer. 2012; 20:1729–1935. [PubMed: 21922203]
8. Bruner DW, Scott C, Lawton C, et al. RTOG's first quality of life study - RTOG 9020: a phase III trial of external beam radiation therapy with etanidazole for locally advanced prostate cancer. Int J Radiat Oncol Biol Phys. 1995; 33:901–906. [PubMed: 7591900]
9. Bruner, DW.; Scott, CB.; McGowan, D., et al. The RTOG Modified Sexual Adjustment Questionnaire: psychometric testing in the prostate cancer population. Int J Radiat Oncol Biol Phys; Presented at the American Society for Therapeutic Radiology and Oncology (ASTRO), Phoenix, AZ, October 1998; 1998. p. 202
10. Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. J Clin Epidemiol. 2002; 55:1130–1143. [PubMed: 12507678]
11. Morton RF, Sloan JA, Grothey A, et al. A comparison of simple single-item measures and the common toxicity criteria in detecting the onset of oxaliplatin- induced peripheral neuropathy in patients with colorectal cancer (Abstract 8087). J Clin Oncol. 2005; 23(16S part I):750s.
12. RTOG Cooperative Group Common Toxicity Criteria. Available at : http://www.rtog.org/ResearchAssociates/AdverseEventReporting/CooperativeGroupCommonTo xicityCriteria.aspx
13. Green S, Weiss G. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. Invest New Drugs. 1992; 10:239–253. [PubMed: 1487397]
14. Sloan JA, Dueck A. Issues for statisticians in conducting analyses and translating results for quality of life end points in clinical trials. J Biopharm Stat. 2004; 14:73–96. [PubMed: 15027501]
15. Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy (FACT) scale: development and validation of the general measure. J Clin Oncol. 1993; 11:570–579. [PubMed: 8445433]
16. Webster K, Odom L, Peterman A, Lent L, Cella D. The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: validation of version 4 of the core questionnaire. Qual Life Res. 1999; 8:604.
17. Winstead-Fry P, Schultz A. Psychometric assessment of the Functional Assessment of Cancer Therapy-General (FACT-G) scale in a rural sample. Cancer. 1997; 79:2446–2452. [PubMed: 9191537]
18. Overcash J, Extermann M, Parr J, Perry J, Balducci L. Validity and reliability of the FACT-G scale for use in the older person with cancer. J Clin Oncol. 2001; 24:591–596.

19. Schipper H, Clinch J, McMurray A, Levitt M. Measuring the quality of life of cancer patients: The Functional Living Index-Cancer: development and validation. J Clin Oncol. 1984; 2:472–483. [PubMed: 6374052]
20. Clinch, JJ. The Functional Living Index-Cancer: ten years later. In: Spilker, B., editor. Quality of life and pharmacoeconomics in clinical trials. 2nd ed.. Philadelphia, PA: Lippincott-Raven Publishers; 1996. p. 215-225.
21. Seidman AD, Portenoy R, Yao TJ, et al. Quality of life in phase II trials: a study of methodology and predictive value in patients with advanced breast cancer treated with paclitaxel plus granulocyte colony-stimulating factor. J Natl Cancer Inst. 1995; 87:1316–1322. [PubMed: 7544834]
22. Hollen, PJ.; Gralla, RJ.; Kris, MG. An overview of the Lung Cancer Symptom Scale. In: Gralla, RJ.; Moinpour, C., editors. Assessing quality of life in patients with lung cancer: A guide for clinicians. New York: NCM Publishers; 1995. p. 57-63.
23. Hollen PJ, Gralla RJ, Kris MG, Potanovich LM. Quality of life assessment in individuals with lung cancer: testing the Lung Cancer Symptom Scale (LCSS). Eur J Cancer. 1993; 29A:S51–S58. [PubMed: 8381294]
24. Hollen PJ, Gralla RJ, Kris MG, Cox C. Quality of life during clinical trials: conceptual model for the Lung Cancer Symptom Scale (LCSS). Support Care Cancer. 1994; 2:213–222. [PubMed: 8087439]
25. Hollen PJ, Gralla RJ, Kris MG, et al. Measurement of quality of life in patients with lung cancer in multicenter trials of new therapies: psychometric assessment of the Lung Cancer Symptom Scale. Cancer. 1994; 73:2087–2098. [PubMed: 8156514]
26. McCorkle R, Young K. Development of a symptom distress scale. Cancer Nurs. 1978; 1:373–378. [PubMed: 250445]
27. McCorkle R, Benoliel JQ. Symptom distress, current concerns, and mood disturbances after diagnosis of life threatening disease. Soc Sci Med. 1983; 17:431–438. [PubMed: 6867788]
28. Degner LF, Sloan JA. Symptom distress in newly diagnosed ambulatory cancer patients and as a predictor of survival in lung cancer. J Pain Symptom Manage. 1995; 10:1–8. [PubMed: 7714343]
29. Sloan JA, Cella D, Hays R. Clinical significance of patient-reported questionnaire data: another step toward consensus. J Clin Epidemiol. 2005; 58:1217–1219. [PubMed: 16291464]
30. Sloan JA, Symonds T, Vargas-Chanes D, Fridley B. Practical guidelines for assessing the clinical significance of health-related quality of life changes within clinical trials. Drug Information Journal. 2003; 37:23–31.
31. Cella DF. Quality of life outcomes: measurement and validation. Oncology (Huntingt). 1996; 10:233–246. [PubMed: 8953606]
32. 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; 16:139–144. [PubMed: 9440735]
33. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Qual Life Res. 1996; 5:555–567. [PubMed: 8993101]
34. Reilly CM, Bruner DW, Mitchell SA, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. Support Care Cancer. 2013; 21:1525–1550. [PubMed: 23314601]
35. Cohen, J. Statistical power analysis for the behavioral sciences. 2nd ed.. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
36. Colon-Otero G, Niedringhaus RD, Hillman SL, et al. A phase II trial of edatrexate, vinblastine, adriamycin, cisplatin and filgrastim (evac/g-csf) in patients with advanced non-small cell carcinoma of the lungs: an NCCTG trial. Am J Clin Oncol. 2001; 24:551–555. [PubMed: 11801752]
37. Schild SE, Bonner JA, Hillman SL, et al. Results of a pilot study of high-dose thoracic radiation therapy with concurrent cisplatin and toposide in limited-stage small cell lung cancer. J Clin Oncol. 2007; 25:3124–3129. [PubMed: 17634491]
38. Johnson EA, Marks RS, Mandrekar SJ, et al. Phase III randomized, double-blind study of maintenance CAI or placebo in patients with advanced non-small cell lung cancer after completion of initial therapy (NCCTG 97-24-51). Lung Cancer. 2008; 60:200–207. [PubMed: 18045731]

39. Garces YI, Okuno SH, Schild SE, et al. Phase I NCCTG Trial-N9923 of escalating doses of twice daily thoracic radiation therapy with amifostine and with alternating chemotherapy in limited stage small cell lung cancer. Int J Radiat Oncol Biol Phys. 2007; 67:995–1001. [PubMed: 17336213]
40. Okuno SH, Delaune R, Sloan JA, et al. A phase II study of gemcitabine and epirubicin for the treatment of pleural mesothelioma: a North Central Cancer Treatment study, N0021. Cancer. 2008; 112:1772–1779. [PubMed: 18224661]
41. Kanard A, Jatoi A, Castillo RA, et al. Oral vinorelbine for the treatment of metastatic non-small cell lung cancer in elderly patients: a phase II trial of efficacy, toxicity, and patients' preference for oral therapy. Lung Cancer. 2004; 43:345–353. [PubMed: 15165094]
42. Johnstone DW, Byhardt RW, Ettinger D, et al. Radiation Therapy Oncology Group. Phase III study comparing chemotherapy and radiotherapy with preoperative chemotherapy and surgical resection in patients with non-small-cell lung cancer with spread to mediastinal lymph nodes (N2); final report of RTOG 89-01. Int J Radiat Oncol Biol Phys. 2002; 54:365–369. [PubMed: 12243809]
43. Hartsell WF, Scott CB, Dundas GS, et al. Can serum markers be used to predict acute and late toxicity in patients with lung cancer? Analysis of RTOG 91-03. Am J Clin Oncol. 2007; 30:368–376. [PubMed: 17762437]
44. Hesketh PJ, Chansky K, Lau DH, et al. Sequential vinorelbine and docetaxel in advanced non-small cell lung cancer patients age 70 and older and/or with a performance status of 2: a phase II trial of the Southwest Oncology Group (S0027). J Thorac Oncol. 2006; 1:537–544. [PubMed: 17409914]
45. Kelly K, Crowley J, Bunn PA Jr, et al. Randomized Phase III trial of paclitaxel plus carboplatin versus vinorelbine plus cisplatin in the treatment of patients with advanced non-small-cell lung cancer: a Southwest Oncology Group Trial. J Clin Oncol. 2001; 19:3210–3218. [PubMed: 11432888]
46. Moinpour CM, Lyons B, Grevstad P, et al. Quality of life in advanced non-small-cell lung cancer: results of a Southwest Oncology Group randomized trial. Qual Life Res. 2002; 11:115–126. [PubMed: 12018735]
47. National Cancer Institute. Cancer therapy evaluation program. Common Toxicit6y Criteria manual. Common Toxicity Criteria. 1999 Jun 1. version 2.0.
48. Halyard MY, Tan A, Callister MD, et al. Assessing the clinical significance of real-time quality of life (QOL) data in cancer patients treated with radiation therapy. [abstract]. J Clin Oncol. 2010; 28(Supplement):9107.
49. Lee H, Friedman ME, Cukor P, Ahern D. Interactive voice response system (IVRS) in health care services. Nurs Outlook. 2003; 51:277–283. [PubMed: 14688763]
50. Movsas B, Hunt D, Watkins-Bruner D, et al. Can electronic web-based technology improve quality of life data collection? Analysis of RTOG 0828. Pract Radiat Oncol. 2014; 4:187–191. [PubMed: 24766686]
51. Gershon RC, Rothrock NE, Hanrahan RT, et al. The development of a clinical outcomes survey research application: assessment center. Qual Life Res. 2010; 19:677–685. [PubMed: 20306332]
52. Ridgeway JL, Beebe TJ, Chute CG, et al. Putting patient-reported information at the heart of cancer care: a patient reported outcomes quality of life (PROQOL) instrument. PLOS Medicine. in press.
53. Dueck AC, Mendoza TR, Mitchell SA, et al. Validity and reliabilityof the patient-reported outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). [abstract]. J Clin Oncol. 2012; 30(Supplement):9047.

Table 1

Participating Trials

Refer-ence	Sample Size	QOL Scale	QOL Schedule	AE Criteria	AE Schedule
34	30	UNISCALE	BL, Q3 Weeks	CTC 2.0	Q3 Weeks
35	82	UNISCALE	BL, 3& 5 months post reg; 3 mo, 1 & 2 year post tx	CTC 2.0	Q28 days while on tx, then 2, 3, 6, 9, 12, 16, 20, 24, 30, 36, 48, and 60 months post tx
36	135	UNISCALE FACT-L	BL, at time of tumor measurement (anywhere from 28 days to 8 weeks)	CTC 2.0	Q28 days
*	106	UNISCALE LCCS	BL, Q28 days	CTC 2.0	Q28 days
37	11	UNISCALE	BL, cycle 2 (day 42), after chemo (day 132); 3 mo, 1 & 2 year post tx	CTC 2.0	Q21 days for 2 cycles, then Q48 days for one cycle, then Q21 days for 2 more cycles, then Q16 days for 1 cycle, then Q3 months for 1 year, Q4 months for 1 year, and Q6 months for 1 year
38	64	UNISCALE SDS	BL, prior to cycle 3 (day 64); 3 mo. and 1 year post tx	CTC 2.0	Q21 days
39	59	UNISCALE LCSS	BL and at 8 weeks after treatment initiation	CTC 2.0	Q28 days until end of treatment, then Q3 months for 1 year
40	56	FACT-L, FLIC	BL, end of RT (week 7–8); q 3 mo for 1 year; then annually	CTC 2.0	Weekly during RT, then 3, 6, 9, and 12 months after RT, then annually
41	134	BDI, TDI	BL; q 3 mo. For 1 yr post tx.	Coopera tive Group CTC	Weekly during RT (through day 90), then 3, 6, 9, and 12 months, then annually
*	8	LCSS	BL, weeks 8 & 1;, at 3, 6, 9, 12, 18 & 24 mos. Post tx	Coopera tive Group CTC	Q7 days for 8 weeks, then at 12 weeks, then at 3, 6, 9, 12, 18, and 24 months after RT
43	29	FACT-L TOI	BL, start of cycle 2-6, and at week 22	CTC 2.0	Q21 days for 3 cycles, then Q28 days for 3 cycles, then week 22, then 3, 6, 9, 12, 18, 24, 36,42, and 48 months after treatment

Reference	Sample Size	QOL Scale	QOL Schedule	AE Criteria	AE Schedule
43, 44	222	HRQOL	BL, then at weeks 13 and 25.	SWOG toxicity criteria	Q4 Weeks during treatment, then Q3 months thereafter

* There is no published manuscript

Table 2

Patient Characteristics

	RTOG (N=151)	SWOG (N=322)	NCCTG (N=540)	Total (N=1013)
Age				
Mean (SD)	62.7 (9.23)	65.6 (10.70)	65.3 (10.26)	65.0 (10.29)
Median	63.0	67.0	66.0	66.0
Sex				
Female	54 (35.8%)	119 (37%)	206 (38.1%)	379 (37.4%)
Male	97 (64.2%)	203 (63%)	334 (61.9%)	634 (62.6%)
Race				
Missing/Unknown/Other	4 (2.6%)	2 (0.6%)	19 (3.5%)	25 (2.5%)
Asian	4 (2.6%)	3 (0.9%)	1 (0.2%)	8 (0.8%)
Black	24 (15.9%)	46 (14.3%)	8 (1.5%)	78 (7.8%)
Hispanic	4 (2.6%)	6 (1.9%)	9 (1.7%)	19 (1.9%)
Native American	0 (0%)	0 (0%)	5 (0.9%)	5 (0.5%)
White	115 (76.2%)	265 (82.3%)	498 (92.2%)	878 (86.7%)
RX				
Currently Receiving Chemotherapy	22 (14.6%)	322 (100%)	449 (83.1%)	793 (78.3%)
Undergone Surgery	18 (11.9%)	0 (0%)	0 (0%)	18 (1.8%)
Currently on RT	111 (73.5%)	0 (0%)	0 (0%)	111 (11%)
Placebo	0 (0%)	0 (0%)	91 (16.9%)	91 (9%)
Follow-up Status				
Alive	24 (15.9%)	40 (12.4%)	36 (6.7%)	100 (9.9%)
Dead	127 (84.1%)	282 (87.6%)	504 (93.3%)	913 (90.1%)
Baseline Uniscale				
N	4	301	465	770
Mean (SD)	32.5 (23.6)	51.7 (35.0)	72.6 (22.2)	64.2 (29.8)
Baseline FACT-L Total Score				
N		191	156	347
Mean (SD)		65.3 (13.5)	78.7 (11.8)	71.3 (14.4)
Baseline FLIC Total Score				
N	16			16
Mean (SD)	70.5 (14.8)			70.5 (14.8)
Baseline LCSS Score				
N	4		128	132
Mean (SD)	93.3 (2.4)		68.2 (25.0)	69.0 (25.0)
Baseline SDS Average				

	RTOG (N=151)	SWOG (N=322)	NCCTG (N=540)	Total (N=1013)
N			53	53
Mean (SD)			78.4 (12.5)	78.4 (12.5)

Table 3

a: Maximum Grade of Each AE Per Patient (selected AEs only)								
	GRADE						# of patients (%)	
Toxicity	1	2	3	4	5	Total	Grade 2+	Grade 3+
Anorexia	166	70	31	3	0	270	104 (11.2)	34 (3.7)
Confusion	10	1	7	1	0	19	9 (1.0)	8 (0.9)
Constipation	112	76	21	0	0	209	97 (10.4)	21 (2.3)
Diarrhea	99	40	25	5	0	169	70 (7.5)	30 (3.2)
Dyspnea	6	128	67	25	3	229	223 (24.0)	95 (10.2)
Fatigue	188	183	88	10	0	469	281 (30.2)	98 (10.5)
Nausea	233	170	94	3	0	500	267 (28.7)	97 (10.4)
Pain-Arthralgia	29	16	4	0	0	49	20 (2.2)	4 (0.4)
Pain-Headache	27	10	4	0	0	41	14 (1.5)	4 (0.4)
Pain	62	55	35	5	0	157	95 (10.2)	40 (4.3)
Total	932	749	376	52	3	2112	572 (61.5)	302 (32.5)

b: All Incidences of AEs Occurring During the Trial (selected AEs only)						
	GRADE					
Toxicity	1	2	3	4	5	Total
Anorexia	305	99	32	4	0	440
Confusion	11	1	7	1	0	20
Constipation	175	106	22	0	0	303
Diarrhea	155	59	28	5	0	247
Dyspnea	18	331	90	28	3	470
Fatigue	641	325	99	10	0	1075
Nausea	567	228	115	3	0	913
Pain-Arthralgia	44	22	4	0	0	70
Pain-Headache	34	13	4	0	0	51
Pain	97	84	39	5	0	225
Total	2047	1268	440	56	3	3814

J Pain Symptom Manage. Author manuscript; available in PMC 2016 October 01.

Table 4

a: Associations of PRO outcomes in patients with and without Grade 2+ Toxicity						
Assessment	**Grade 2+ Tox?**	**N**	**Mean**	**SD**	**Median**	**P-Value**
Uniscale	NO	168	−13.5	33.6	−10.9	0.07[1]
	YES	386	−20.0	33.1	−18.3	
FACT-L Total	NO	101	−4.5	11.4	−3.7	0.001[2]
	YES	133	−10.0	13.4	−10.1	
FLIC Total	NO	10	−13.9	12.0	−12.9	0.56[2]
	YES	18	−17.6	17.8	−15.5	
LCSS Total	NO	22	−1.2	11.9	−0.3	0.14[2]
	YES	89	−6.4	15.2	−5.9	
SDS Average	NO	1	16.0	-	16.0	0.09[2]
	YES	21	−7.7	12.8	−6.0	

b: Associations of PRO outcomes in patients with and without Grade 3+ Toxicity						
Assessment	**Grade 3+ Tox?**	**N**	**Mean**	**SD**	**Median**	**P-Value**
Uniscale	NO	359	−16.9	33.6	−16.8	0.60[1]
	YES	195	−20.1	32.8	−14.8	
FACT-L Total	NO	189	−6.9	11.9	−7.6	0.14[2]
	YES	45	−10.7	15.8	−12.1	
FLIC Total	NO	19	−14.8	16.2	−10.7	0.49[2]
	YES	9	−19.3	15.5	−20.4	
LCSS Total	NO	63	−3.2	13.2	−2.7	0.08[2]
	YES	48	−8.1	16.2	−8.7	
SDS Average	NO	12	−4.8	10.3	−5.5	0.50[2]
	YES	10	−8.8	16.8	−5.0	

[1]Wilcoxon Rank-Sum

[2]Two-Sample T-test

Table 5

Incidence of Severe AE and Clinically Significant Differences (CSD) in the Related PRO Items

Toxicity	**Assessment**		
Anorexia	FACT-L Losing weight	LCSS Appetite	SDS Appetite
Number evaluable*	324	106	22
AE Grade 2+	13%	8%	41%
AE Grade 3+	3%	1%	5%
CSD in QOL	43%	43%	50%
AE 2+ and CSD in QOL	6%	5%	32%
% Agreement AE 2+ and QOL	56%	59%	73%
AE 3+ and CSD in QOL	1%	0%	5%
% Agreement AE 3+ and QOL	56%	57%	55%
Dyspnea	FACT-L Tightness in chest	FACT-L Ease of Breathing	LCSS Shortness of breath
Number evaluable*	322	321	105
AE Grade 2+	31%	31%	40%
AE Grade 3+	12%	13%	17%
CSD in QOL	35%	43%	45%
AE 2+ and CSD in QOL	14%	15%	20%
% Agreement AE 2+ and QOL	62%	55%	55%
AE 3+ and CSD in QOL	6%	6%	11%
% Agreement AE 3+ and QOL	64%	56%	59%
Fatigue	FACT-L Have energy	LCSS Fatigue	SDS Fatigue
Number evaluable*	334	105	22
AE Grade 2+	37%	58%	46%
AE Grade 3+	11%	19%	18%
CSD in QOL	57%	57%	46%
AE 2+ and CSD in QOL	23%	38%	23%
% Agreement AE 2+ and QOL	53%	61%	55%
AE 3+ and CSD in QOL	8%	14%	9%
% Agreement AE 3+ and QOL	47%	52%	55%
Nausea	FACT-L Nausea	SDS Nausea incidence	SDS Nausea severity
Number evaluable*	333	22	14
AE Grade 2+	25%	41%	43%
AE Grade 3+	5%	9%	14%
CSD in QOL	47%	27%	21%
AE 2+ and CSD in QOL	16%	14%	7%

Toxicity	**Assessment**		
% Agreement AE 2+ and QOL	60%	59%	50%
AE 3+ and CSD in QOL	3%	5%	0%
% Agreement AE 3+ and QOL	54%	73%	64%
Confusion	FACT-L Clear thinking	SDS Concentration	
Number evaluable*	328	22	
AE Grade 2+	1%	5%	
AE Grade 3+	1%	5%	
CSD in QOL	44%	36%	
AE 2+ and CSD in QOL	1%	0%	
% Agreement AE 2+ and QOL	57%	59%	
AE 3+ and CSD in QOL	1%	0%	
% Agreement AE 3+ and QOL	57%	59%	
Constipation	SDS Bowel		
Number evaluable*	22		
AE Grade 2+	23%		
AE Grade 3+	5%		
CSD in QOL	36%		
AE 2+ and CSD in QOL	18%		
% Agreement AE 2+ and QOL	77%		
AE 3+ and CSD in QOL	5%		
% Agreement AE 3+ and QOL	68%		
Diarrhea	SDS Bowel		
Number evaluable*	22		
AE Grade 2+	5%		
AE Grade 3+	0%		
CSD in QOL	36%		
AE 2+ and CSD in QOL	0%		
% Agreement AE 2+ and QOL	59%		
AE 3+ and CSD in QOL	0%		
% Agreement AE 3+ and QOL	63%		
Arthralgia	FACT-L Physical Functioning	SDS Pain Severity	
Number evaluable*	376	21	
AE Grade 2+	2%	10%	
AE Grade 3+	0%	0	
CSD in QOL	60%	43%	
AE 2+ and CSD in QOL	2%	10%	

Toxicity	**Assessment**		
% Agreement AE 2+ and QOL	41%	67%	
AE 3+ and CSD in QOL	0%	0	
% Agreement AE 3+ and QOL	40%	57%	
Headache	FACT-L Physical Functioning	SDS Pain Severity	
Number evaluable*	376	21	
AE Grade 2+	2%	5%	
AE Grade 3+	0%	0	
CSD in QOL	60%	43%	
AE 2+ and CSD in QOL	1%	5%	
% Agreement AE 2+ and QOL	40%	62%	
AE 3+ and CSD in QOL	0	0	
% Agreement AE 3+ and QOL	40%	57%	
Pain	FACT-L Physical Functioning	SDS Pain Severity	
Number evaluable*	376	21	
AE Grade 2+	13%	5%	
AE Grade 3+	5%	5%	
CSD in QOL	60%	43%	
AE 2+ and CSD in QOL	9%	0	
% Agreement AE 2+ and QOL	44%	52%	
AE 3+ and CSD in QOL	3%	0	
% Agreement AE 3+ and QOL	42%	52%	

* Represents the number of patients that had an adverse event (grade specified) **and** completed a QOL assessment at baseline and at least once post-baseline. **Does not include** patients with verified baseline AE of grade 2+.