

Comparing and Validating Simple Measures of Patient-Reported Peripheral Neuropathy for Oncology Clinical Trials: NCCTG N0897 (Alliance) A Pooled Analysis of 2440 Patients

Heshan Liu¹, Angelina D. Tan¹, Rui Qin¹, Daniel J. Sargent¹, Axel Grothey², Jan C. Buckner², Paul L. Schaefer³, and Jeff A. Sloan^{1*}

1Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN 55905, USA

2Department of Medical Oncology, Mayo Clinic, Rochester, MN 55905, USA

3Toledo Community Hospital Oncology Program, Toledo, OH, USA

Received: December 16, 2014; Accepted: March 26, 2015; Published: April 10, 2015

*Corresponding author: Jeff A. Sloan, Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN 55905, USA, Tel: +507-261-4268; Fax: 507-266-2477; E-mail: jsloan@mayo.edu

Abstract

Introduction: Current standard evaluation of Peripheral Neuropathy (PN) is based on an investigator-reported classification system that is commonly unable to correctly reflect the subjective symptoms for patients. Thus more reliable methods to assess PN are needed. This study assessed alternative methods of assessing patient-reported PN in 5 North Central Cancer Treatment Group (NCCTG) clinical trials.

Method: Two single-item assessments relating to numbness and tingling were used to measure PN. Patients' Quality Of Life (QOL) was also assessed using the Uniscale, Symptom Distress Scale (SDS), Profile of Mood States (POMS), Brief Pain Inventory (BPI) and Subject Global Impression of Change (SGIC). Wilcoxon tests compared QOL scores between patients with PN (score > 50) vs. no PN (score ≤ 50). Changes from baseline in QOL were compared by Wilcoxon rank sum test with a 20-point change in PN defined as clinically meaningful. Both distribution-based and anchor-based approaches were used to derive estimates of Minimal Important Differences (MID). Standardized Response Means (SRM), Effect Sizes (ES) and Guyatt's responsiveness statistic were used to measure responsiveness.

Results: The proportion of patients reporting numbness (tingling) at baseline was 10.7% (10.0%) and 18.4% (17.8%) at last assessment. The correlation between numbness and tingling at baseline was 0.81, and at last assessment was 0.83. Patients with substantial PN reported an average of 10 points lower overall QOL, mood and worse symptom distress and 20 points lower in the BPI interference items. Patients having a ≤ 20 point worsening in PN score reported significantly worse in symptom distress and BPI worst pain, but not in POMS or overall QOL. The MID estimates were similar between numbness and tingling items but varied depending on the approach used. Responsiveness statistics indicated that the two PN assessments are sensitive and responsive instruments for cancer patients with PN.

Conclusions: The two PN items for numbness and tingling were redundant. Evidence of criterion validity and responsiveness indicates that these simple measures of PN can be used successfully in cancer clinical trials.

Keywords: Pooled Analysis; Quality of Life; Peripheral Neuropathy; Minimum Importance Difference; Oncology Trials

Introduction

Peripheral Neuropathy (PN) is a painful, common and potentially dose-limiting adverse effect of treatment with taxanes, vinca alkaloids, and platinum-based chemotherapeutic agents [1]. Neuropathic symptoms may compromise the quality of life, influencing physical, social and psychological functioning, which maybe particularly critical during cancer treatment and recovery among cancer patients [2]. Furthermore, neuropathy can become so severe as to limit the tolerable total dose of chemotherapy, potentially leading to disease progression [2,3]. The introduction of new anti-cancer therapies that cause PN has increased the need for methods to reliably assess PN in clinical and research settings.

To date, the commonly used physician-based instruments to assess cancer-induced PN include the National Cancer Institute - Common Terminology Criteria for Adverse Events (NCI-CTCAE), the World Health Organization (WHO) Common Toxicity Criteria for PN, and the Eastern Cooperative Oncology Group (ECOG) Grading Scale for CIPN [4-6]. These grading scales neither define grade uniformly nor define terms clearly. Hence each clinician undertakes the task of assessing peripheral neuropathy without clear points of reference and is hence forced to rely upon personal experience and specific knowledge with peripheral neuropathy which could be highly variable. This leads to subjective interpretation by the clinician and makes PN diagnosis and grading more controversial. The subjective nature of these ratings also results in substantial inter-observer and intra-observer variation in assessing PN. Moreover, current evidence suggests that physician-based assessments tend to under-report the incidence and severity of PN [7-9].

The lack of a universally recognized, standardized, valid and reliable patient-reported tool that quantifies PN symptoms makes

comparisons among published studies of PN difficult. Patient-reported instruments are often lengthy and do not provide extensive information about the intensity of this toxicity or the severity of resulting impairment. Therefore, the development of a new and reliable tool for **assessing** PN that allows early detection and captures conditions with the psychometric properties required for use in clinical and research setting would represent an important advance.

There have been more recent advances in the measurement of PN, most notably by the European Organization for the Treatment and Research in Cancer (EORTC) which produced a 20-item measure for patient-reported peripheral neuropathy (ref) and the Functional Assessment of Cancer Therapy Scale/Gynecologic Oncology Group-Neurotoxicity (FACT/GOG-Ntx) questionnaire [10,11]. Other studies have looked at simple visual analogue scales and the Neuropathy Symptom Score and Neuropathy Symptoms and Change Score used in other studies [12,13]. Challenges remain however, in that the relative psychometric qualities of the various assessments has yet to be established in detecting PN, and there is no evidence comparing the more involved assessments to simple single-item measures.

The primary aim of this study was to compare and validate two single-item measures of PN using data from a series of North Central Cancer Treatment Group (NCCTG) clinical trials. A secondary aim was to produce an estimate of the Minimal Important Difference (MID) for the PN measures and compare the responsiveness to clinical change of the PN measures in cancer patients.

Methods

Data were pooled from five NCCTG clinical trials involving a total of 2440 patients. Details of the individual trials and their publications are in Table 1 [14-16]. Each participant signed an IRB-approved, protocol-specific informed consent in accordance with federal and institutional guidelines. All trials used two single-item questions relating to numbness and tingling in fingers and toes to measure PN, where a score of 0 represented no numbness or tingling and a score of 10 represented the worst imaginable numbness or tingling. These items were developed within the NCCTG specifically for the assessment of oxaliplatin-related toxicity [17]. Patients also completed other instruments to assess health-related quality of life (QOL). Overall QOL was measured using the Uniscale, symptoms were measured using the Symptom Distress Scale (SDS), mood was measured using the Profile of Mood States (POMS) Short Form, pain was measured using the Brief Pain Inventory (BPI) and physical status was measured using the Subject Global Impression of Change (SGIC). All assessments were provided to the patient in professionally prepared booklets by study coordinators at each site and were measured at baseline and at least one other time point during the trial. The psychometric properties of study instruments are described below.

The studies were analyzed separately and together in a pooled dataset. Results in terms of the responsiveness of the measures, and other psychometric properties were consistent

across studies (data not shown). Pooling was justified as the results comparing intra-patient differences and correlation among the various PN assessments were our goals.

The Uniscale is a single-item measure of global QOL using a numerical rating scale ranging from 0 to 10 with well-established validity data in cancer populations [18-20].

The SDS is a valid and reliable 13-item cancer-specific instrument intended for assessing the degree of distress associated with the following 11 cancer-related symptoms as perceived by the patient: nausea, appetite, insomnia, pain, fatigue, bowel pattern, concentration, appearance, outlook, breathing, and cough. Each symptom is rated on a 1-5 Likert-type scale [21,22].

The POMS is a 37 item scale utilized to assess patients' overall mood and specific mood items of fatigue-inertia, vigor-activity, tension-anxiety, depression-dejection, anger-hostility and confusion-bewilderment. It is valid for use, and is discriminative, in the evaluation of cancer- and pain-associated mood disturbance or psychological distress [23].

The BPI is a pain assessment tool that consists of 15 items to locate pain, determine pain severity, determine how the pain interferes with daily activities and assess the extent of pain relief received from analgesics [24,25]. All items except those concerning pain location and medication are measured using numeric analogue scales ranging from 0 to 10. The BPI has been widely used and has been validated for use in cancer populations [26].

The SGIC is a 7-point item in which patients rate the change in the overall status since the beginning of the study (ranging from much improved, moderately improved, minimally improved, no change, minimally worse, moderately worse, too much worse). The item has been found to be effectively discriminate treatment effects in neuropathic pain trials [27].

The QOL assessments were scored according to each tool's established scoring algorithm. The scores were coded so that a low score was representative of poor patient condition, as necessary, and scores converted into 0-100 point scale with 100 being the best possible score for ease of interpretability and comparability between instruments with differing ranges [28-30]. Each of the two single-item PN scores is on a 0 to 10 ordinal scale and thus was also converted to a 0-100 scale. Patients were assigned a dichotomous category for substantial PN at baseline using the scoring cut-off for Clinically Deficient PN (CDPN) based on the numbness and tingling scores (CDPN of ≤ 50 vs. non-CDPN of > 50). This approach was consistent with prior work that determined a score of 50 or less on the 100-point scale was indicative of a deficit that required clinical intervention or at least further clinical investigation and assessment [30-32]. Changes in scores were calculated for PN and QOL assessments using baseline and the last assessment. Patients were further categorized as becoming ≥ 20 points worse vs. < 20 points worse in PN rating as measured by the worst change from baseline in numbness or tingling [29]. The Wilcoxon rank-sum test was used to compare QOL scores between assigned categorical groups [33].

Associations between PN and QOL scores were examined using Spearman correlation coefficients and the two PN assessments were compared using a Bland-Altman analysis [34,35]. Weighted kappa was used to measure agreement between the two PN assessments [36].

Both distribution-based and anchor-based approaches were used to derive estimates of Minimal Important Differences (MID), which has been defined as the smallest change in a patient-reported outcome that is perceived by patients as beneficial or that would result in a change in treatment [37,38]. The anchor-based approach relied upon SGIC related to QOL, physical condition and emotional status. The distribution-based approach in this analysis applied the 1/2 standard deviation method was applied to determine the MID [39]. Only the two trials, N00C3 and N01C3 that collected SGIC at week 8 were used to determine the MID estimates.

Standardized Response Means (SRM), Effect Sizes (ES) and Guyatt's responsiveness statistic were used to measure responsiveness. The SRM was calculated as the mean change in scores divided by the standard deviation of the *change scores* [40]. The ES was calculated as the mean change in scores divided by the standard deviation of the baseline scores [41,42]. Guyatt's responsiveness statistic was calculated as the mean change in scores divided by the standard deviation of change of patients in the placebo group [43,44]. For these indices, small effects were considered higher than 0.20 but less than 0.50; moderate effects, higher than 0.50 but less than 0.80; and large effects, higher than 0.80 [45-47]. Data collection and statistical analyses were conducted by the Alliance Statistics and Data Center. Data was frozen by 12/11/2014, and statistical analyses were performed using SAS version 9.2.

Results

Patient characteristics are reported in Table 2. The median age was 61 years (range, 19-88 years), 55% patients were men, and 88% patients were white. The proportion of patients reporting numbness (tingling) at baseline was 10.7% (10.0%) and 18.4% (17.8%) at last assessment. The proportion of patients reporting PN on either item was 11.4% at baseline and 20.3% at last assessment.

Patients with substantial PN at baseline (CDPN \leq 50) reported an average of 10 points lower overall QOL, SDS and POMS and 20 points lower in the BPI interference items (Table 3). Mean changes in these QOL assessments from baseline to last observation reflected greater reduction in patients having at least a 20 point decrease in PN scores over time compared to those who did not. This was true for the SDS (-1.5 vs. 1.0, respectively, $P < 0.001$), and the BPI. The mean difference in decline in most BPI items between the two groups was greater than 10 points, suggesting a reasonable responsiveness of these QOL components to PN (Table 4).

The correlations between the two PN questions at baseline and last assessment were 0.81 (95% CI: 0.78-0.84) and 0.83 (95% CI: 0.81-0.86), respectively. The redundancy indices in canonical correlation analysis at baseline and last assessment were 0.85 and 0.79, respectively (both above the typical cutoff of 0.75 to indicate redundancy). The average difference generated by subtracting tingling from numbness at baseline was -0.55 (SD = 9.48, $p=0.01$) and the average difference generated by subtracting tingling from numbness at last assessment was -0.25 (SD = 12.52, $p=0.39$). A strong agreement between these two PN items was observed with a weighted kappa of 0.85 (95% CI: 0.83-0.88). The Bland-Altman plot in Figure 1 displays the

Table 1: List of Trials that Were Included in Neuropathy Pooled Analysis.

Protocol	Description	Accrual	Study Open	Study Closed	QOL Tools	References
N00C3	The Efficacy of Gabapentin in the Management of Chemotherapy-Induced Peripheral Neuropathy: A Phase III Randomized, Double-Blind, Placebo-Controlled, Crossover Trial	115	02/08/2002	12/05/2003	UNISCALE, SDS, SGIC, POMS, BPI	RD Rao et al. [14]
N01C3	The Efficacy of Lamotrigine in the Management of Chemotherapy-Induced Peripheral Neuropathy: A Phase III Randomized, Double-Blind, Placebo-Controlled Trial	131	02/13/2004	03/11/2005	UNISCALE, SDS, SGIC, POMS, BPI	
N05C3	The Use of Vitamin E for Prevention of Chemotherapy Induced Peripheral Neuropathy: A Phase III Double-Blind Placebo Controlled Study	207	12/01/2006	12/14/2007	Only Two single-item assessment relating to numbness and tingling	
N9741	A Randomized Phase III Trial of Combinations of Oxaliplatin (OXAL), 5-Fluorouracil (5-FU), and Irinotecan (CPT-11) as Initial Treatment of Patients With Advanced Adenocarcinoma of the Colon and Rectum	1751	10/27/1998	07/19/2002	UNISCALE, SDS	Goldberg and DJ Sargent [15]
N9841	A Randomized Phase III Equivalence Trial of Irinotecan (CPT-11) Versus Oxaliplatin (OXAL)/5-Fluorouracil (5-FU)/Leucovorin (CF) in Patients With Advanced Colorectal Carcinoma Previously Treated With 5-FU	507	11/05/1999	12/17/2003	UNISCALE, SDS	Kim [16]

Table 2: Patient Characteristics.

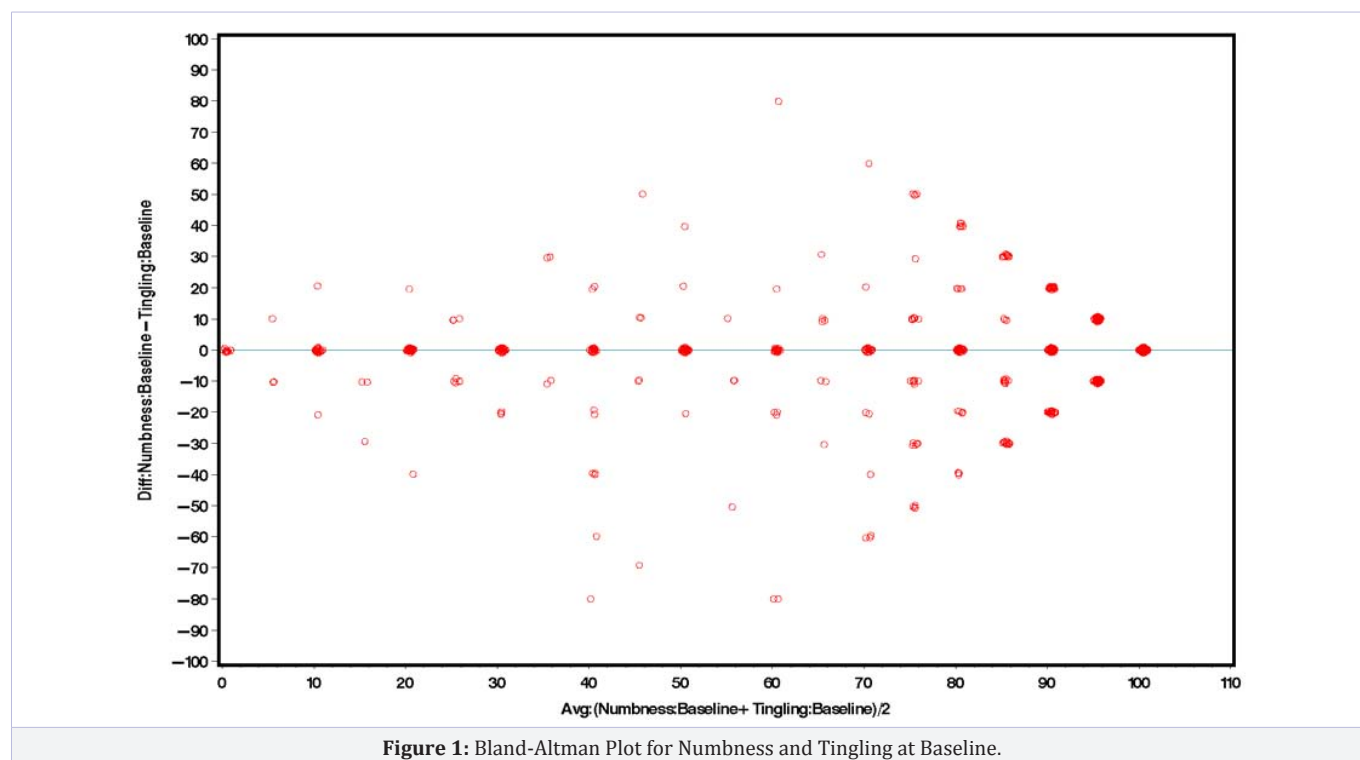
		Total	(N = 2440)
Age			
	N	2435	
	Mean (SD)	59.9 (11.77)	
	Median	61.0	
	Q1, Q3	52.0, 68.0	
	Range	(19.0-88.0)	
Gender			
	female	1103 (45.2%)	
	male	1337 (54.8%)	
Race			
	White	2137 (87.6%)	
	Black or African American	167 (6.8%)	
	Native Hawaiian or Other Pacific Islander	6 (0.2%)	
	Asian	28 (1.1%)	
	American Indian or Alaska Native	13 (0.5%)	
	Not reported: patient refused or not available	84 (3.4%)	
	Unknown: Patient unsure	5 (0.2%)	
Disease#Status			
	Missing	423 (%)	
	Measurable	1668 (82.7%)	
	Evaluable	349 (17.3%)	
Performance#Score			
	Missing	931 (%)	
	0	838 (55.5%)	
	1	600 (39.8%)	
	2	71 (4.7%)	
ECOG#PS			
	Missing	423 (%)	
	0-1	1930 (95.7%)	
	2	87 (4.3%)	

Table 3: Mean Baseline QOL Scores by Baseline Peripheral Neuropathy.

QOL	No PN (N = 1769)	PN (N = 227)	Total (N = 1996)	p value
Uniscale Overall QOL	77.5	65.4	76.0 (19.66)	<0.0001
SDS Total	82.6	73.7	81.5 (12.97)	<0.0001
POMS Total	77.2	69.4	70.8 (13.17)	0.0011
BPI Worst Pain	66.3	51.6	54.2 (27.92)	0.0009
BPI Least Pain	86.0	73.6	75.8 (21.10)	0.0006
BPI Average Pain	72.8	60.1	62.3 (24.69)	0.0024
BPI Pain Now	79.2	65.5	67.9 (26.32)	0.0061
BPI Activity Interference	79.7	60.5	63.8 (30.94)	0.0005
BPI Mood Interference	85.8	65.9	69.4 (28.77)	0.0002
BPI Walking Interference	79.5	51.6	56.5 (33.04)	<0.0001
BPI Work Interference	73.0	51.1	55.0 (32.21)	0.0001
BPI Relations Interference	92.6	77.2	79.9 (26.17)	0.0012
BPI Sleep Interference	79.0	61.7	64.8 (32.39)	0.0049
BPI Enjoyment Interference	79.8	54.8	59.2 (33.06)	<0.0001
BPI Pain Interference	81.0	60.6	64.3 (25.93)	<0.0001

Table 4: Mean QOL Change from Baseline Score by 20 Points Worsening PN.

QOL	Without 20 Points Worsening (N = 980)	With ≥ 20 Points Worsening (N = 611)	Total (N = 1591)	p value
Uniscale Overall QOL	-4.9	-4.8	-4.9 (23.20)	0.8624
SDS Total	1.0	-1.5	0.0 (12.66)	0.0002
POMS Total	1.4	-0.8	1.0 (10.54)	0.6397
BPI Worst Pain	9.4	-7.3	6.4 (29.18)	0.0193
BPI Least Pain	2.4	-5.0	1.0 (23.31)	0.0996
BPI Average Pain	7.6	-2.9	5.6 (26.82)	0.0869
BPI Pain Now	5.8	-6.1	3.5 (28.83)	0.1904
BPI Activity Interference	9.1	-4.1	6.6 (32.95)	0.0920
BPI Mood Interference	7.6	0.4	6.2 (27.37)	0.3031
BPI Walking Interference	11.1	-0.7	8.9 (33.83)	0.1812
BPI Work Interference	14.8	-0.4	12.0 (34.21)	0.0248
BPI Relations Interference	5.5	0.7	4.6 (25.67)	0.7614
BPI Sleep Interference	11.6	0.7	9.5 (33.02)	0.1834
BPI Enjoyment Interference	11.9	2.1	10.1 (32.55)	0.3590
BPI Pain Interference	10.2	-1.2	8.0 (25.11)	0.1018



differences between the numbness and tingling items versus the average of the two items. The picture indicates that differences are scattered equally on either side of zero and that there is no pattern in the differences relative to the degree of numbness or tingling reported ($r = 0.08$). The arrow head shape on the extreme right and left sides of the figure only indicates the lack of extreme differences. Collectively this indicates that these two neuropathy

measures were highly correlated to the point of redundancy. Table 5 presents correlations between QOL assessments and both numbness and tingling at baseline. Table 6 presents the correlations between changes in QOL and change in numbness and tingling. Both PN symptoms were weakly correlated with overall QOL, SDS and POMS. Change in numbness and tingling were also weakly correlated with change in overall QOL, SDS

and POMS, indicating that patient-reported PN symptoms are not strongly associated with overall QOL, SDS and POMS, and are therefore measuring distinctly different constructs. Both PN symptoms were found to be moderately correlated with pain and physical function. Baseline and changes in numbness and tingling were found to be moderately correlated with changes in pain and physical function, suggesting that the changes in PN are associated with changes in pain and physical activity functions.

One important feature of the PN items was their sensitivity to treatment. Increase and decrease of the PN items were related to the patient rated improvement in physical status. The MID estimate using the overall QOL, physical condition and emotional state of SCIG were 24, 21, 25 points for numbness (respectively); and 19, 18, 17 points for tingling (respectively). The distribution based MID estimate was 11 points for numbness and 12 for tingling. The correlation between the patient anchors and the neuropathy items are low (< 0.4), indicating that the anchors may not be the valid ones. Patients who experienced substantial PN reported significant lower QOL, mood and worse symptom distress: the average difference is around 10, and the average difference in the BPI interferences are even higher (around 20), indicating a substantial association of neuropathy to BPI interference.

The overall ES, SRM and Guyatt's statistic for numbness after cycle 1 are 0.61, 0.47 and 0.77 (respectively); and 0.55, 0.44 and 0.73 for tingling (respectively), indicates that the two PN symptoms are sensitive and responsive QOL instruments for cancer patients with PN.

Discussion

The primary finding of importance in this comparative study of alternative measures of neuropathy is that the two neuropathy measures measuring numbness and tingling separately were redundant. Hence, in designing future clinical trials to assess PN, these data support the hypothesis that a simple single-item

drawn from the numbness or tingling or combining numbness and tingling into one item would be sufficient to produce valid and reliable data, although generalizability is restricted to the sort of patients studied and the type of clinical trials involved

The major advantage of single-item PN assessments is their simplicity, in terms of administration, scoring, and interpretation as has been demonstrated repeatedly [48]. The disadvantage of the two assessments is that they do not describe of all symptoms of the PN such as neuropathy pain. As a screening tool in clinical practice or stratification variable in clinical trials, the single-item assessment can be the trigger that launches a subsequent, more comprehensive investigation into uncovering the specific PN deficit and/or initiating appropriate clinical interventions [49].

A secondary finding more relevant to methodologists was that anchor-based MID methods for determining clinical significance produced larger estimates than the distribution-based methods. This is an important point because the relative merit of anchor-based and distribution-based approaches to clinical significance assessment is still being debated [50,51]. The findings are consistent with previous work indicating the choice of anchor can substantially impact MID estimates [52,53].

There are limitations to this evaluation of PN measures. The five trials in this pooled analysis included two CIPN treatment trials, one prevention trial testing interventions to reduce PN, and two chemotherapy treatment trials where CIPN was expected to be a prominent toxicity. The influence of heterogeneity in the cancer patients and its relation to patient outcome may have influenced individual levels of CIPN but as the construct under study was the inpatient differences among the various measures of CIPN, each patient served as their own control. Furthermore, results remained consistent when each trial was independently analyzed. Replication of the psychometric properties will be studied further in ongoing and planned clinical trials. Another limitation arises because the responsiveness of the QOL scale depends on

Table 5: Correlation between Baseline Numbness, Tingling and Baseline QOL items.

QOL	Numbness		Tingling	
	Correlation(r)	95% CI	Correlation(r)	95% CI
Uniscale Overall QOL	0.23	0.18-0.28	0.21	0.16-0.25
SDS Total	0.25	0.21-0.29	0.21	0.16-0.25
POMS Total	0.16	0.03-0.29	0.23	0.10-0.36
BPI Worst Pain	0.32	0.19-0.44	0.30	0.18-0.43
BPI Least Pain	0.25	0.11-0.38	0.22	0.08-0.36
BPI Average Pain	0.27	0.14-0.41	0.26	0.13-0.39
BPI Pain Now	0.26	0.13-0.39	0.21	0.08-0.34
BPI Activity Interference	0.35	0.22-0.47	0.30	0.17-0.43
BPI Mood Interference	0.34	0.22-0.46	0.29	0.16-0.42
BPI Walking Interference	0.44	0.33-0.56	0.37	0.25-0.49
BPI Work Interference	0.41	0.30-0.53	0.38	0.26-0.50
BPI Relations Interference	0.26	0.14-0.38	0.23	0.10-0.36
BPI Sleep Interference	0.28	0.16-0.41	0.32	0.20-0.45
BPI Enjoyment Interference	0.41	0.29-0.52	0.38	0.26-0.50
BPI Pain Interference	0.43	0.32-0.54	0.39	0.27-0.52

Table 6: Correlation between Change from Baseline in Numbness, Tingling and Change from Baseline in QOL items.

QOL change from baseline	Numbness change from baseline		Tingling changes from baseline	
	Correlation(r)	95% CI	Correlation(r)	95% CI
SDS Total	0.12	0.07-0.18	0.13	0.08-0.18
POMS Total	0.11	-0.06-0.28	0.25	0.09-0.41
BPI Worst Pain	0.46	0.32-0.60	0.43	0.28-0.57
BPI Least Pain	0.37	0.22-0.51	0.30	0.13-0.46
BPI Average Pain	0.41	0.26-0.57	0.39	0.24-0.55
BPI Pain Now	0.35	0.20-0.49	0.28	0.12-0.43
BPI Activity Interference	0.31	0.17-0.45	0.30	0.15-0.45
BPI Mood Interference	0.24	0.08-0.40	0.18	0.01-0.34
BPI Walking Interference	0.40	0.26-0.55	0.42	0.28-0.56
BPI Work Interference	0.42	0.29-0.55	0.43	0.30-0.57
BPI Relations Interference	0.15	-0.01-0.31	0.14	-0.03-0.29
BPI Sleep Interference	0.26	0.09-0.42	0.28	0.13-0.44
BPI Enjoyment Interference	0.31	0.16-0.46	0.33	0.18-0.47
BPI Pain Interference	0.40	0.26-0.54	0.38	0.24-0.53

the follow-up time point and patients' initial health status. Three trials have different study-specific QOL collection schedules, so two trials only were used for deriving responsiveness statistics. Further, the mean PN assessments at baseline are high, indicating good health status in regard to PN. As such, it is difficult for the CIPN measure to detect further improvement [54,55]. This might have led to an under-estimating of the responsiveness statistics in the study. Finally, we did not compare the CIPN measures with objective measures of nerve conduction, nor did we undertake discriminant validity in this paper, as these aspects have been reported elsewhere for patient-reported CIPN measures [56].

Ultimately, the key finding of this pooled analysis is that the simple measures of CIPN demonstrate acceptable psychometrics including criterion validity and responsiveness indicating that these simple measures of PN can be used successfully in cancer clinical trials. While there are other measures of CIPN available, they are all longer and more complicated, representing additional burden both to the patients and clinical trial system. We are carrying out further research focusing on direct comparisons of alternative measures of CIPN to learn about the relative merits of the various assessment approaches.

Funding

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers U10CA180821 and U10CA180882 to the Alliance for Clinical Trials in Oncology. This work was also supported in part by Public Health Service grants CA-25224, CA-37404, CA-35431, CA-35415, CA-35103, and CA-35269. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Lee JJ, Swain SM. Peripheral neuropathy induced by microtubule-stabilizing agents. *J Clin Oncol.* 2006; 24(10): 1633-42.
2. Dunlap B, Paice JA. Chemotherapy-induced peripheral neuropathy: a need for standardized in measurement. *J Support Oncol.* 2006; 4(8): 398-9.
3. Wilkes G. Peripheral neuropathy related to chemotherapy. *Semin Oncol Nurs.* 2007; 23(3): 162-73.
4. Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0, 2009.
5. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer.* 1981; 47(1): 207-14.
6. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern cooperative oncology group. *Am J Clin Oncol.* 1982; 5(6): 649-55.
7. Stephens RJ, Hopwood P, Girling DJ, Machin D. Randomized trials with quality of life endpoints: are doctors' ratings of patients' physical symptoms interchangeable with patients' self-ratings? *Qual Life Res.* 1997; 6(3): 225-36.
8. Shimozuma K, Ohashi Y, Takeuchi A, Aranishi T, Morita S, Kuroi K, et al. Validation of the Patient Neurotoxicity Questionnaire (PNQ) during taxane chemotherapy in a phase III randomized trial of breast cancer: N-SAS BC 02. *Support Care Cancer.* 2009; 17(12): 1483-91. doi: 10.1007/s00520-009-0613-7.
9. Sloan JA, Berk L, Roscoe J, Fisch MJ, Shaw EG, Wyatt G, et al. Integrating patient-reported outcomes into cancer symptom management clinical trials supported by the National Cancer Institute-sponsored clinical trials networks. *J Clin Oncol.* 2007; 25(32): 5070-7.
10. Lavoie Smith EM, Barton DL, Qin R, Steen PD, Aaronson NK, Loprinzi CL. Assessing patient-reported peripheral neuropathy: the reliability and validity of the European Organization for Research and Treatment of Cancer QLQ-CIPN20 Questionnaire. *Qual Life Res.* 2013; 22(10): 2787-99. doi: 10.1007/s11136-013-0379-8.

11. Richardson PG, Briemberg H, Jagannath S, Wen PY, Barlogie B, Berenson J, et al. Frequency, Characteristics, and Reversibility of Peripheral Neuropathy During Treatment of Advanced Multiple Myeloma With Bortezomib. *J Clin Oncol*. 2006; 24(19): 3113-20.
12. Dyck PJ, Davies JL, Litchy WJ, O'Brien PC. Longitudinal assessment of diabetic polyneuropathy using a composite score in the Rochester Diabetic Neuropathy Study cohort. *Neurology*. 1997; 49(1): 229-39.
13. Apfel SC, Schwartz S, Adornato BT, Freeman R, Biton V, Rendell M, et al. Efficacy and safety of recombinant human nerve growth factor in patients with diabetic polyneuropathy: A randomized controlled trial—rhNGF Clinical Investigator Group. *JAMA*. 2000; 284(17): 2215-21.
14. Rao RD, Michalak JC, Sloan JA, Loprinzi CL, Soori GS, Nikcevic DA, et al. Efficacy of gabapentin in the management of chemotherapy-induced peripheral neuropathy A phase 3 randomized, double-blind, placebo-controlled, crossover trial (N00C3). *Cancer*. 2007; 110(9): 2110-8.
15. Goldberg RM, Sargent DJ, Morton RF, Fuchs C, Ramanathan RK, Williamson SK, et al. N9741: FOLFOX (oxaliplatin(Oxal)/ 5-fluorouracil (5-FU)/ leucovorin (LV) or reduced dose R-IFL (CPT-11 + 5-FU/LV) in advanced colorectal cancer (CRC): Final efficacy data from an intergroup study. *Journal of Clinical Oncology*. 2004; 22(14S): 3621.
16. Kim GP, Sargent DJ, Mahoney MR, Rowland KM Jr, Philip PA, Mitchell E, et al. Phase III Noninferiority Trial Comparing Irinotecan With Oxaliplatin, Fluorouracil, and Leucovorin in Patients With Advanced Colorectal Carcinoma Previously Treated With Fluorouracil: N984. *J Clin Oncol*. 2009; 27(17): 2848-54. doi: 10.1200/JCO.2008.20.4552.
17. Goldberg RM, Sargent DJ, Morton RF, Fuchs CS, Ramanathan RK, Williamson SK, et al. A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J Clin Oncol*. 2004; 22(1): 23-30.
18. Hürny C, Bernhard J, Coates A, Peterson HF, Castiglione-Gertsch M, Gelber RD, et al. Responsiveness of a single-item indicator versus a multi-item scale: assessment of emotional well-being in an international adjuvant breast cancer trial. *Med Care*. 1996; 34(3): 234-48.
19. Spitzer WO, Dobson AJ, Hall J, Chesterman E, Levi J, Shepherd R, et al. Measuring the quality of life of cancer patients. *J Chronic Dis*. 1981; 34(12): 585-97.
20. Sloan JA, Loprinzi CL, Kuross SA, Miser AW, O'Fallon JR, Mahoney MR, et al. Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. *J Clin Oncol*. 1998; 16(11): 3662-73.
21. McCorkle R, Young K. Development of a symptom distress scale. *Cancer Nurs*. 1978; 1(5): 373-8.
22. McCorkle R, Benoliel JQ. Symptom distress, current concerns, and mood disturbances after diagnosis of life threatening disease. *Soc Sci Med*. 1983; 17(7): 431-8.
23. Baker F, Denniston M, Zabora J, Polland A, Dudley WN. A POMS short form for cancer patients: psychometric and structural evaluation. *Psychooncology*. 2002; 11(4): 273-81.
24. Daut RL, Cleeland CS. The prevalence and severity of pain in cancer. *Cancer*. 1982; 50(9): 1913-8.
25. Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. *Pain*. 1983; 17(2): 197-210.
26. Cleeland CS, Ryan KM. Pain Assessment: Global Use of the Brief Pain Inventory. *Ann Acad Med Singapore*. 1994; 23(2): 129-38.
27. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001; 94(2): 149-58.
28. Sloan A, O'Fallon JR, Suman VJ. Incorporating quality of life measurement into oncology clinical trials. *Proc Am Stat Assoc*. 1998; 282-287.
29. Sloan J, Symonds T, Vargas-Chanes D, Fridley B. Practical guidelines for assessing the clinical significance of health-related quality of life changes within clinical trials. *Therapeutic Innovation & Regulatory Science*. 2003; 37(1):23-31. doi: 10.1177/009286150303700105.
30. 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(1): 73-96.
31. Temel JS, Pirl WF, Recklitis CJ, Cashavelly B, Lynch TJ. Feasibility and validity of a one-item fatigue screen in a thoracic oncology clinic. *J Thorac Oncol*. 2006; 1(5): 454-9.
32. Butt Z, Wagner LI, Beaumont JL, Paice JA, Peterman AH, Shevrin D, et al. Use of a single-item screening tool to detect clinically significant fatigue, pain, distress, and anorexia in ambulatory cancer practice. *J Pain Symptom Manage*. 2008; 35(1): 20-30.
33. Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*. 1947; 18 (1): 50-60.
34. Spearman C. The proof and measurement of association between two things. *Am J Psychol*. 1987; 100(3-4): 441-71.
35. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician*. 1983; 32: 307-17. doi:10.2307/2987937.
36. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968; 70(4): 213-20.
37. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987; 40(2): 171-8.
38. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clinic Proc*. 2002; 77(4): 371-83.
39. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Med Care*. 2003; 41(5): 582-92.
40. Katz JN, Larson MG, Philips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care*. 1992; 30(10): 917-25.
41. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. 1989; 27(3, suppl): S178-89.
42. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. *Control Clin Trial*. 1991; 12(4, suppl): 142S-158S.
43. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol*. 1992; 45(12): 1341-5.

44. Ruof J, Sangha O, Stucki G. Comparative responsiveness of 3 functional indices in ankylosing spondylitis. *J Rheumatol*. 26(9): 1959-63.
45. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther*. 1996; 18(5): 979-92.
46. Liang MH. Evaluating measurement responsiveness. *J Rheumatol*. 1995; 22(6): 1191-2.
47. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine (Phila Pa 1976)*. 2000; 25(24): 3192-9.
48. Sloan JA, Aaronson N, Cappelleri JC, Fairclough DL, Varricchio C. Clinical Significance Consensus Meeting Group. Assessing the clinical significance of single items relative to summated scores. *Mayo Clin Proc*. 2002; 77(5): 479-87.
49. Frost MH, Bonomi AE, Cappelleri JC, Schunemann HJ, Moynihan TJ, Aaronson NK. Clinical Significance Consensus Meeting Group. Applying quality-of-life data formally and systematically into clinical practice. *Mayo Clin Proc*. 2007; 82(10): 1214-28.
50. Cella D, Eton DT, Lai J-S, Peterman A, Merkel DE. Combining anchor and distribution based methods to derive clinically meaningful differences on the Functional Assessment of Cancer Therapy (FACT) Anemia and Fatigue Scales. *J Pain Symptom Manage*. 2002; 24(6): 547-61.
51. Norman GR, Sridhar FG, Guyatt GH, Walter SD. Relation of distribution and anchor-based approaches in interpretation of changes in health-related quality of life. *Med Care*. 2001; 39(10): 1039-1047.
52. Yost K, Sorenson M, Hahn E, Glendenning A, Gnanasakthy A, Cella D. Using multiple anchor-and distribution based estimates to evaluate clinically meaningful change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) instrument. *Value Health*. 2005; 8(2): 117-27.
53. Revicki DA, Erickson PA, Sloan JA, Dueck A, Guess H, Santanello NC. Interpreting and reporting results based on patient-reported outcomes. *Value Health*. 2007; 10 Suppl 2: S116-24.
54. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Med Care*. 2000; 38(9, suppl): I184-I190.
55. Lydick E. Approches to the interpretation of quality-of-life scales. *Med Care*. 2000; 38(9 Suppl): I1180-3.
56. Bakitas MA. Background noise: the experience of chemotherapy-induced peripheral neuropathy. *Nurs Res*. 2007; 56(5): 323-31.