Supportive Care in Cancer (2019) 27:2599–2608
https://doi.org/10.1007/s00520-018-4553-y

ORIGINAL ARTICLE

Rasch model-based testing of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire–Chemotherapy-Induced Peripheral Neuropathy (QLQ-CIPN20) using Alliance for Clinical Trials in Oncology (Alliance) A151408 study data

Ellen M. Lavoie Smith[1] · Noah Zanville[2] · Grace Kanzawa-Lee[1] · Clare Donohoe[1] · Celia Bridges[1] · Charles Loprinzi[3] · Jennifer Le-Rademacher[4] · James J. Yang[1]

Received: 14 August 2018 / Accepted: 13 November 2018 / Published online: 20 November 2018

Abstract

Purpose To test the psychometric properties of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire–Chemotherapy-Induced Peripheral Neuropathy (QLQ-CIPN20) using Rasch-based methods.

Methods A secondary data analysis was performed using pooled QLQ-CIPN20 data from patients ($N = 1008$) who had participated in any of four multi-site chemotherapy-induced peripheral neuropathy (CIPN) treatment and prevention trials. QLQ-CIPN20 responses were evaluated using a polytomous Rasch partial credit model. Data were assessed for person-item fit using the chi-square statistic, item scaling based on response proportions, threshold ordering using item characteristic curves and logit threshold locations, differential item response (DIF) (i.e., response bias) using likelihood ratio tests, and unidimensionality using cluster analysis.

Results A statistically significant chi-square test indicated poor fit of the observed to the expected responses. More than 70% of the respondents reported a complete absence of six symptoms, reflecting significant floor effects and poor item scaling. Disordered/non-ordinal or narrow response thresholds were found for 11 of the 20 items. Item responses were significantly different by gender ($p < 0.0001$) and chemotherapy type ($p < 0.0001$). Cluster analysis findings suggest that the QLQ-CIPN20 is a unidimensional scale due to the absence of item clusters.

Conclusions Rasch model testing revealed psychometric weaknesses that could be addressed by revising the QLQ-CIPN20's problematic items and response options. Alternatively, perhaps the new gold standard CIPN measurement approach in future intervention trials should involve use of only the best items, which would also allow comparisons across previous trials that utilized the QLQ-CIPN20.

Keywords QLQ-CIPN20 · Rasch analysis · Chemotherapy-induced peripheral neuropathy (CIPN) · Psychometric testing

ClinicalTrials.gov identifiers: NCT00516503 (N06CA); NCT00860041 (N08C1); NCT02311907 (N08CA); NCT01099449 (N08CB)

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00520-018-4553-y) contains supplementary material, which is available to authorized users.

✉ Ellen M. Lavoie Smith
ellenls@med.umich.edu

[1] University of Michigan School of Nursing, 400 North Ingalls, Ann Arbor, MI 48109, USA

[2] Indiana University School of Nursing, Indianapolis, IN, USA

[3] Mayo Clinic, Rochester, MN, USA

[4] Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN, USA

Introduction

Chemotherapy-induced peripheral neuropathy (CIPN) is a dose-limiting toxicity associated with a number of commonly used antineoplastic therapies [1]. CIPN symptoms—typically tingling, numbness, pain, weakness, and a loss of reflexes in the extremities—can affect as many as 60% of patients receiving neurotoxic cancer therapy by their third month of treatment [2]. In addition to being distressing, symptoms can become debilitating, forcing providers to alter or stop cancer therapy [3]. Studies also show that CIPN is associated with falls [4, 5], disturbed sleep [6, 7], difficulty performing work-related tasks [8], increased healthcare utilization [9], and greater use of medication (including opioids) to manage symptoms [9, 10]. Methods for preventing CIPN have yet to be identified [1], increasing the need for trials evaluating potential interventions.

Currently, several measures, but no gold-standard approach, are used to evaluate CIPN in intervention trials [11, 12], including grading scales such as the National Cancer Institute Common Toxicity Criteria Adverse Events (CTCAE) [11], and the different versions of the Total Neuropathy Score (TNS), which combine questions about CIPN symptoms with simplified physical assessment procedures [13, 14]. Several *patient-reported outcome* (PRO) measures that evaluate patients' perceptions of their CIPN have also been developed [15]. Examples of CIPN PRO measures include the Patient Neurotoxicity Questionnaire (PNQ) [16], the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity (FACT-GOG/Ntx) scale [17, 18], and the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire–Chemotherapy-Induced Peripheral Neuropathy (QLQ-CIPN20) [19]. Because the QLQ-CIPN20 has been widely used and extensively tested [20–28], it has potential to move forward as the gold standard PRO measure.

An important issue in how PRO measures like the QLQ-CIPN20 have been tested is our reliance on *classical test theory* (CTT). CTT assumes that the distance between items on a scale, like inches on a ruler, is equal [29, 30]. While CTT-based procedures are well-suited for validating measures with interval or ratio-level items, CTT is poorly suited for testing the QLQ-CIPN20's ordinal-level item responses (i.e., 1 = not at all, 2 = a little bit, 3 = quite a bit, 4 = very much). Since the distance between its response levels cannot be assumed to be equal, this can lead to floor and ceiling effects that undermine sensitivity. Another assumption of CTT is that when items are summed to create a total score, all items have equal importance/weight [29] and, therefore, can be combined into different subscales. In the case of the QLQ-CIPN20, items have been summed to obtain three subscale scores (sensory, motor, and autonomic). However, neither this subscale structure nor the relative weight of the items has yet been established empirically; further, recent evidence suggests that the QLQ-CIPN20 may perform better as a unidimensional scale [23, 25].

One way to overcome the limitations associated with CTT-based methods for testing ordinal-level measures like the QLQ-CIPN20 is to use models based on *Item Response Theory* (IRT) instead. Unlike CTT-models, IRT-models use probability to determine the likelihood that patients with symptoms at a certain level of severity (e.g., moderately severe CIPN) will select a given response on a corresponding scale (e.g., 2 = moderate symptoms) [31]. The Rasch model, a specific type of IRT model, is based on the assumption that a scale is unidimensional (i.e., items do not cluster into subscales), and compares individual item response patterns from the entire sample to the person location estimates (symptom severity) and item difficulty (based on the proportion of patients who select a particular response option) [31]. Person estimates (severity) and item response patterns (difficulty) are then reflected in a logit-based interval scale [31]. Like other IRT models, Rasch models assume that patients with more severe symptoms will have a higher probability of selecting responses that reflect worse symptom severity [31]. This relationship between symptom severity and the probability that participants will select a specific response can be illustrated using an item characteristic curve (Fig. 1) [29, 31]. When using QLQ-CIPN20 items to measure CIPN severity, each curve in Fig. 1 represents the probability (y-axis) that the patient will rate CIPN items as 1 (black curve = not at all), 2 (green curve = a little), 3 (red curve = quite a bit), or 4 (purple curve = very much) based on their CIPN severity. The x-axis (estimate person location logits) reflects low to high CIPN severity—the latent trait. Points where the curves overlap indicate equal probability of selecting either response. These overlapping points, the logit threshold, can be used to determine if the data are consistent with an ordinal level of measurement. Logit threshold location values should be ordered from low to increasingly higher values. When the data are not ordered, this suggests that respondents are unable to distinguish the difference between the first- and second-highest response options and, perhaps, a more granular response criterion (e.g., 0–10) would facilitate better distinction between subtle differences.

Rasch models have several advantages for evaluating CIPN PRO measures over CTT-based techniques. First, provided that a test's assumptions are met, Rasch techniques have the ability to convert ordinal responses to more precise interval-level responses [31], making it possible to use statistical techniques generally reserved for interval-level data [32]. Second, Rasch models can be used to identify redundant items and restrictive item-response criteria that can negatively compromise the sensitivity and validity of the measures. Rasch techniques can also be used to evaluate item bias (differential item functioning [DIF]): variations in responses based on a participant's unique characteristics such as age, gender,

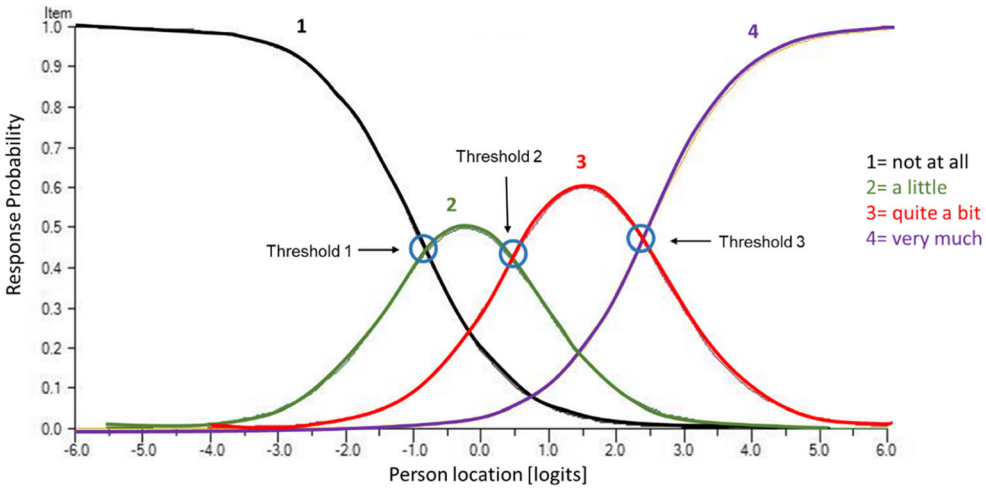

Fig. 1 Exemplar ordered item characteristic curves

diagnosis, or cultural background [31]. Minimizing item bias is particularly important for CIPN PRO measures so that these tools will provide valid and reliable data on CIPN severity in diverse populations [30].

Therefore, the purpose of this secondary data analysis was to perform additional testing on the EORTC QLQ-CIPN20 using Rasch analysis. The goal was to gain additional insight regarding the CIPN20's psychometric properties in an effort to improve its ability to detect and quantify CIPN in individuals with diverse demographic characteristics.

Methods

Sample

QLQ-CIPN20 data were obtained from a large pooled sample ($N = 1008$) of patients who had participated in any of four multi-site CIPN treatment and prevention trials (N06CA, N08C1, N08CA, and N08CB) conducted by the North Central Cancer Treatment Group (NCCTG) [10, 33–35], which is now part of the Alliance for Clinical Trials in Oncology, a large National Cancer Institute (NCI)-funded multisite cancer research network. Institutional review boards at participating sites approved all of the original studies, and written informed consent was obtained from all participants.

One randomized double-blind, placebo-controlled trial (RCT) (N06CA) tested the efficacy for CIPN of a topical compounded cream containing baclofen, amitriptyline, and ketamine (BAK) in patients with CIPN due to a variety of neurotoxic chemotherapy drugs [35]. In N08C1, the QLQ-CIPN20 was used to describe CIPN patterns over time in patients who had received paclitaxel [10, 33]. Two of the four studies were RCTs designed to test drugs to prevent CIPN caused by paclitaxel/carboplatin (N08CA) [36] and oxaliplatin (N08CB) [34]. The eligibility criteria for these four studies have been previously described [10, 33–36]. Of note, two of the intervention studies were negative, and one (N06CA) showed a trend favoring topical BAK cream for sensory neuropathy and statistically significant improvements in motor neuropathy.

The QLQ-CIPN20

The QLQ-CIPN-20 [19] is a 20-item self-report questionnaire originally designed to supplement the EORTC Quality of Life Questionnaire. Based on the initially hypothesized factor structure, it contains 9 items assessing sensory neuropathy, 8 assessing motor function, and 3 assessing autonomic neuropathy (i.e., dizziness, erectile dysfunction, and blurred vision). Items are scored 1–4 with 1 representing "not at all" and 4 "very much." Item scores are summed to obtain subscale scores. However, recent evidence suggests that all items can be summed to obtain a total CIPN score [23]. Scores are then linearly converted to a 0–100 scale [37].

Statistical analyses

Descriptive statistics (i.e., mean, standard deviation [SD], range, and frequency) were calculated for demographics (i.e., age and gender), chemotherapy variables (i.e., parent protocol, chemotherapy treatment [if known]), and QLQ-CIPN20 scores. The QLQ-CIPN20 data used in the analyses reflect the point of greatest CIPN severity for the patients in each study.

The QLQ-CIPN20 responses were evaluated using a polytomous Rasch model using eRm package [38] because the item responses ranged from 1 = not at all to 4 = very much. The partial credit model (PCM) was used because the distances between overlapping points on the response curves varied across the items. Chi-square statistics were used to assess the difference between the observed and expected responses across individual item responses. A nonsignificant chi-square indicates a good fit [31].

To determine whether the items were appropriately targeted (item targeting) to encompass the full range of CIPN severity, a zero logit was selected to represent average severity, with a negative logit value indicating mild or below average neuropathy and a positive logit indicating above average/more severe CIPN [31]. This approach is used to determine the presence of floor or ceiling effects based on the proportion of low–high logit values.

Threshold ordering was evaluated using item characteristic curves and logit threshold locations, which indicate whether item scores reflect an ordered/ordinal response [31]. The point where two adjacent response category curves intersect (logit threshold locations), which indicates that either response is equally probable, was assessed to determine if patients were able to distinguish between the various response options.

Differential item functioning (DIF) was assessed as a measure of item bias using the likelihood ratio test. DIF testing indicates whether item responses differ within sample subgroups (e.g., male/female, neuropathy subtypes) [39]. Although age is another demographic variable commonly used to assess *DIF*, age was not used as a *DIF* variable in this study: because peripheral nerve function normally deteriorates with age, differing item responses between younger and older participants would be expected and not indicative of item bias [27].

K-means clustering method was used to explore various empirical partitions of the observed data. Since the Rasch model is based on the assumption that the summed item score reflects a unidimensional construct, items should not form significant clusters [31].

Results

Demographics

The demographic characteristics of the pooled sample (N = 1008) are reported in Table 1. Age, gender, and chemotherapy type were the demographic variables common to all four Alliance studies. Participants' mean age was 58.8 years (SD = 11.2; range 23.0–89.0) and most were female (68.8%). Most participants received oxaliplatin (34.7%), 31.5% received paclitaxel plus carboplatin, and 13.5% received paclitaxel alone; the remaining 20.2% (BAK participants) received various neurotoxic drugs/combinations. The sample included patients with myeloma, lymphoma, and colon, breast, lung, and ovarian cancer; however, frequency data are not available for every study. Similarly, race frequencies are available for only 10% of the sample, and thus are not reported.

Item and person fit

The chi-square test tests the assumption that the probability of selecting a specific item response (e.g., 1 = not at all) will be closely aligned with symptom severity. A statistically significant chi-square test for four items (#16, 17, 18, and 20) ($p < 0.0001$) indicated that item responses for questions assessing dizziness, blurred version, hearing, and erectile dysfunction were not aligned with symptom severity.

Table 1 Demographic characteristics (N = 1008)

Characteristic	Descriptive statistics
Age	
Mean (SD, range)	58.8 (11.2; 23.0–89.0)
Gender	*n* (%)
Male	315 (31.3)
Female	693 (68.8)
Protocol	
N06CA	204 (20.2)
N08C1	277 (27.5)
N08CA	177 (17.6)
N08CB	350 (34.7)
Chemotherapy treatment	
BAK study - multiple chemotherapy types	204 (20.2)
Oxaliplatin	350 (34.7)
Paclitaxel alone	136 (13.5)
Paclitaxel + carboplatin	318 (31.5)

SD standard deviation

Scale targeting

Tingling and numbness in the fingers or hands (mean = 2.3, 2.2; SD = 1.0) and toes or feet (mean = 2.3, SD = 1.1) were the most severe symptoms. Difficulty using pedals was reported infrequently, with just 11.5% of participants reporting a mean score of greater than 1.1 (SD = 0.4). All items encompassed the entire score range (1–4). Response distributions are also illustrated in Fig. 2. The high proportion of low severity responses suggests that the 1–4 response range does not adequately capture subtle nuances in symptom severity. Specifically, a high proportion (> 70%) of patients denied having burning pain in fingers or hands (item #5), hand cramps (item #7), difficulty distinguishing between hot and cold water (item #10), difficulty walking due to foot drop (item #14), hearing loss (item #18), or difficulty using the pedals (item #19).

Threshold ordering

Other problematic items were identified based on tests assessing threshold ordering. Disordered or narrow thresholds are illustrated via (1) item characteristic curves (Fig. 3), (2) and logit threshold locations (Table 2). Eleven of 20 CIPN20 items have either disordered responses or small differences between response options: the points between adjacent response curves (Fig. 3) where either response is equally

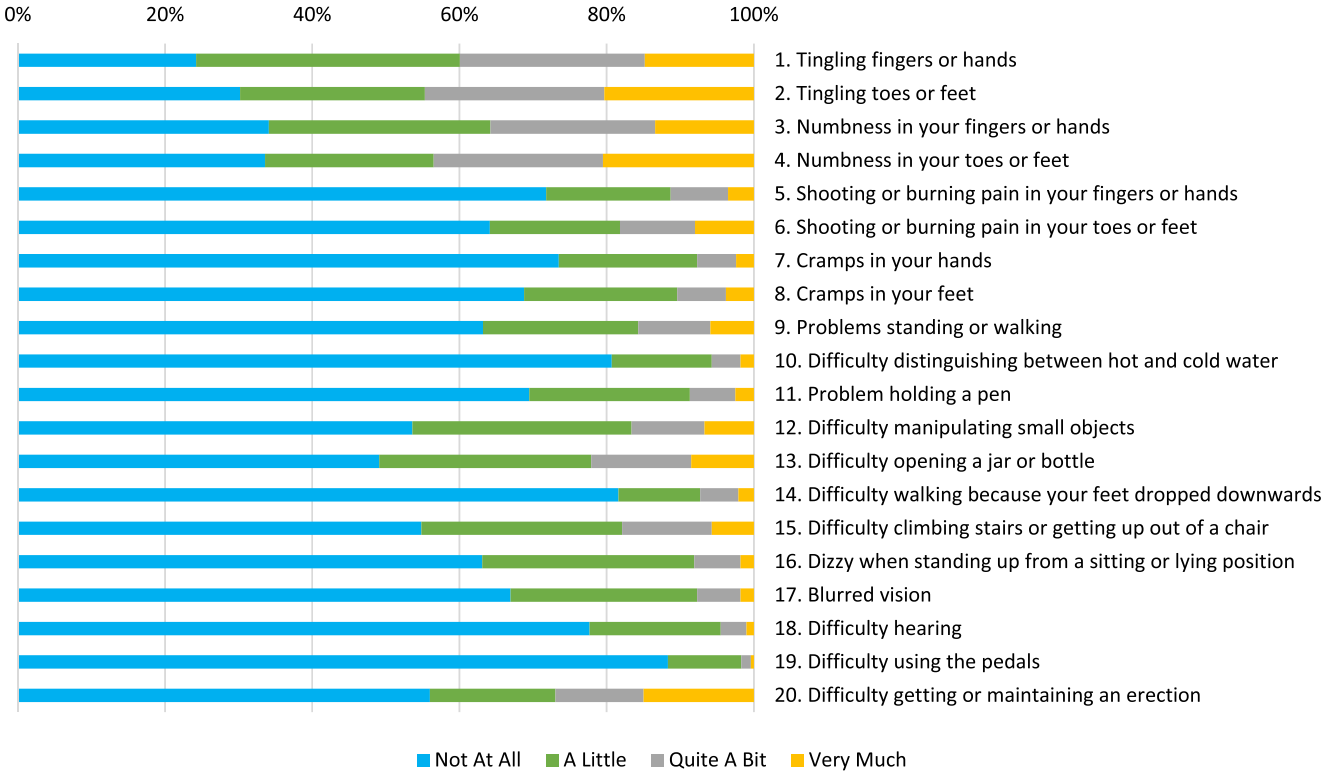

Fig. 2 Distribution of responses for each CIPN20 item

probable are not ordered/ordinal (items 12, 14, 19, and 20) or are very close together (items 5–11). Disordered thresholds are also evidenced in Table 2 in which the threshold logit (point of curve overlap) should progress from lower to higher logit threshold values. For example, the logit thresholds for item CIPN 1 (tingling in fingers and hands) show ordered responses because threshold 1 (−2.5776)—the overlap point between experiencing no neuropathy and a little neuropathy—is lower than threshold 2 logit (−0.8158). Similarly, the threshold 2 logit is lower than logit 3 (0.1073). If an item is functioning as expected, progressively ordered logits will reflect survey respondents' ability to distinguish differences in CIPN symptoms using the 1–4 response options. In contrast, the item-response curves and logit thresholds for item CIPN 12 (difficulty manipulating small objects) are disordered (threshold logit 2 is larger than logit 3). Another obvious example of disordered responses can be seen when examining the item-response curves and threshold logit values for item 20 (erectile dysfunction); the threshold logit 2 is lower than logit 1. If the response items truly reflect ordinal level data, logit 2 should be higher than logit 1, and there should be a greater distance between the various response options. Instead, the logits overlap and are indistinguishable from each other (see supplemental Person Item Map Figure). Further, small standard deviations (Table 2) suggest that survey respondents were less able to indicate differences in symptom severity.

Differential item functioning (DIF)

For the DIF analysis, item responses were significantly different when comparing responses of males and females ($p < 0.0001$), and when comparing patients in four distinct chemotherapy groups: those who received (1) oxaliplatin, (2) paclitaxel alone, (3) paclitaxel plus carboplatin, and (4) a variety of different chemotherapy agents (N06CA) ($p < 0.0001$) (supplemental Tables 1 and 2).

Unidimensionality

Consistent with published findings [23, 25], the cluster analysis findings illustrated in Fig. 4 suggest the lack of a clear factor structure. Items do not cluster together (items with the same color form a cluster) based on sensory (i.e., items 1–6, 9, 10, and 18), motor (i.e., items 7, 8, 11–15, and 19), or autonomic subscales (i.e., items 16, 17, and 20), as previously defined by Postma et al. [19].

Discussion

This paper presents the results of a Rasch model-based analysis that expands our current knowledge regarding the QLQ-CIPN20's psychometric properties. Use of the Rasch model to test and modify ordinal-level PRO measures such as the QLQ-

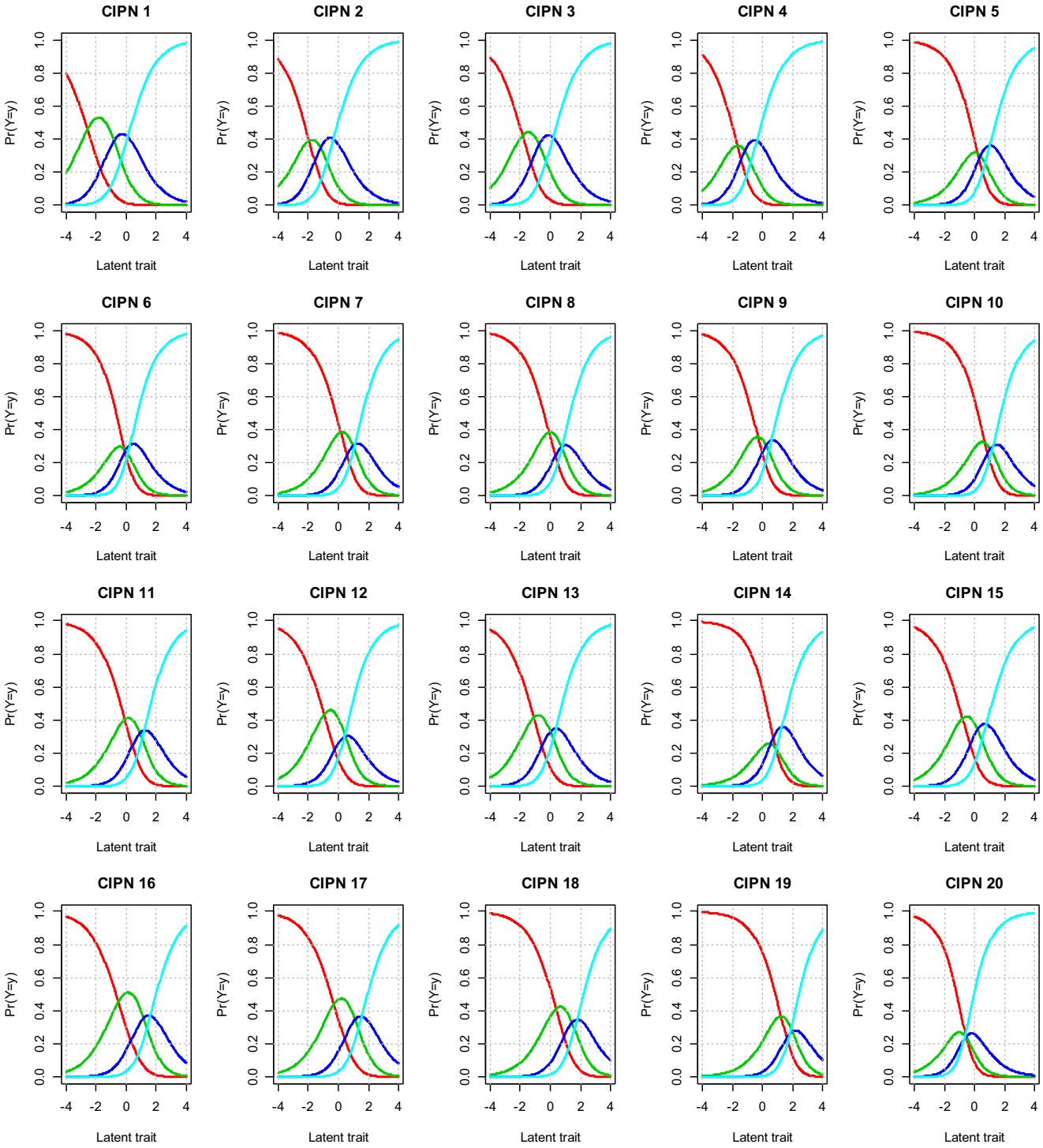

Fig. 3 Threshold order item characteristic curves (partial credit model)

CIPN20 can result in conversion of the instrument into an interval measure that more accurately captures variations in CIPN severity.

While multiple papers support the QLQ-CIPN20's reliability, validity, sensitivity, and responsiveness based on CTT methods [20–28], to our knowledge, this is the first paper to suggest that the QLQ-CIPN20 items and response options are not ideal, based on Rasch model testing. Although the findings are consistent with other published reports suggesting the QLQ-CIPN20 is a unidimensional scale [23, 25], disordered/nonordinal or narrow response thresholds affecting 11 of the 20 items suggest several potential problems with the CIPN20 items.

Table 2 Logit threshold locations

Item #	Item description	Mean Logit	Threshold 1 Logit	Threshold 1 *SD*	Threshold 2 Logit	Threshold 2 *SD*	Threshold 3 Logit	Threshold 3 *SD*
1	Did you have tingling fingers or hands?	−1.0954	−2.5776	0.0992	−0.8158	0.0901	0.1073	0.1111
2	Did you have tingling toes or feet?	−1.1496	−1.9445	0.0987	−1.1840	0.0974	−0.3204	0.1023
3	Did you have numbness in your fingers or hands?	−0.8065	−1.8479	0.0919	−0.7461	0.0956	0.1746	0.1170
4	Did you have numbness in your toes or feet?	−1.0795	−1.6786	0.0982	−1.1802	0.1010	−0.3798	0.1037
5	Did you have shooting or burning pain in your fingers or hands?	0.5364	0.1969	0.0926	0.3925	0.1432	1.0197	0.2133
6	Did you have shooting or burning pain in your toes or feet?	0.0320	−0.1241	0.0923	−0.0311	0.1300	0.2511	0.1578
7	Did you have cramps in your hands?	0.7503	0.1655	0.0891	0.9725	0.1619	1.1129	0.2584
8	Did you have cramps in your feet?	0.4816	−0.0886	0.0873	0.7449	0.1475	0.7884	0.2153
9	Did you have problems standing or walking because of difficulty feeling the ground under your feet?	0.1730	−0.3028	0.0875	0.2332	0.1281	0.5887	0.1740
10	Did you have difficulty distinguishing between hot and cold water?	0.9866	0.6917	0.0994	1.0408	0.1870	1.2273	0.2993
11	Did you have a problem holding a pen, which made writing difficult?	0.6678	−0.0938	0.0856	0.9069	0.1503	1.1903	0.2486
12	Did you have difficulty manipulating small objects with your fingers (for example, fastening small buttons)?	−0.0247	*−0.9441*	0.0818	*0.4688*	0.1217	*0.4010*	0.1671
13	Did you have difficulty opening a jar or bottle because of weakness in your hands?	−0.2457	−1.1065	0.0843	0.0040	0.1104	0.3655	0.1466
14	Did you have difficulty walking because your feet dropped downwards?	0.9105	*0.8965*	0.1074	*0.5398*	0.1752	*1.2953*	0.2720
15	Did you have difficulty climbing stairs or getting up out of a chair because of weakness in your legs?	0.0610	−0.8014	0.0840	0.1941	0.1163	0.7903	0.1713
16	Were you dizzy when standing up from a sitting or lying position?	0.7437	−0.5241	0.0807	1.1478	0.1466	1.6075	0.2879
17	Did you have blurred vision?	0.7933	−0.2940	0.0832	1.1153	0.1524	1.5585	0.2903
18	Did you have difficulty hearing?	1.1923	0.3683	0.0918	1.4380	0.1928	1.7706	0.3827
19	Did you have difficulty using the pedals?	1.6796	*1.2395*	0.1214	*1.9769*	0.3246	*1.8223*	0.6566
20	Did you have difficulty getting or maintaining an erection?	−0.6094	*−0.5449*	0.1778	*−0.5809*	0.2338	*−0.7023*	0.2420

Italicized numbers indicate non-ordinal disordered threshold locations. *SD* standard deviation

Questions assessing autonomic neuropathy and hearing were not aligned with symptom severity. Other problematic items were those assessing patients' difficulty manipulating small objects, walking due to foot drop, and using pedals. Participants had difficulty distinguishing between the four response items, perhaps because they did not understand the intended meaning of the items, which reflects suboptimal content validity. This hypothesis is supported by our previous work testing the content validity of an abbreviated QLQ-CIPN version that excluded the suboptimal autonomic subscale and hearing loss items [26]. We reported that patients did not fully understand the item about foot drop, leading to the recommendation that this item be revised to ask about "trouble flexing the ankle because of weakness" instead of "difficulty walking because your feet dropped downwards."

Another potential reason for disordered thresholds may be that respondents did not answer the question honestly so as to avoid a negative consequence (i.e., social desirability). In the current

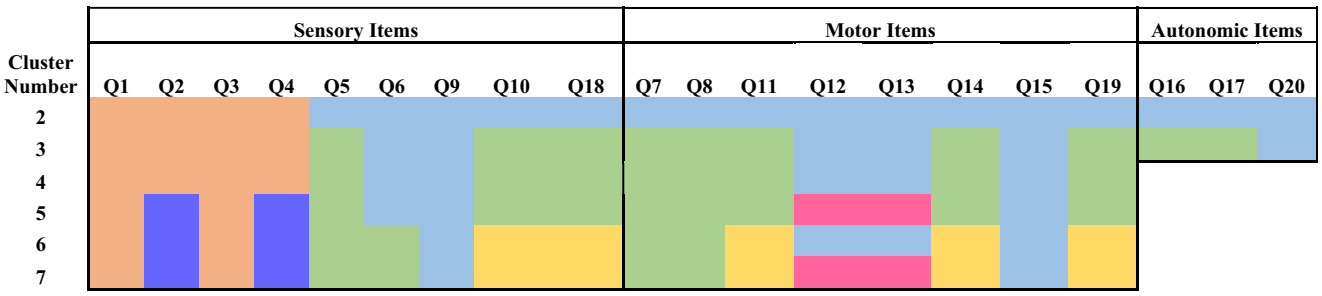

Fig. 4 Cluster analysis

research, responses to item 19 about using pedals were disordered, perhaps partly due to social desirability. Again, this idea is supported by our previous research results: interviewees stated that patients were not likely to answer the question about "using pedals" honestly due to concerns about losing driving privileges. Although we did not evaluate the content validity of the erection item in our prior research, it is likely that some patients would not answer this question honestly either.

Another possible explanation for the disordered or narrow logit thresholds and suboptimal item scaling could be unclear survey instructions and the number of response options. Again, based on our previous research, patients stated that they were sometimes confused about whether to respond based on the severity or on the frequency of the problem [26]. Other respondents indicated that they would have preferred having more response options; a 0–10 response range might facilitate patients' selection of a response that better reflects their symptom experience. In contrast, some narrow logit thresholds could be corrected by dichotomizing the response options, such as for the erection question, because it may be sufficiently informative merely to determine if the patient is experiencing the problem via a yes/no response.

Consistent with published literature, the DIF analyses revealed that item responses varied by gender and chemotherapy type [27]. The DIF findings regarding chemotherapy type validate that CIPN clinical manifestations vary by causative neurotoxic agent, and variation by gender is most likely related to the different types of chemotherapy received by women and men (e.g., women with breast cancer receive paclitaxel ± carboplatin). Thus, differences in responses are expected and may be unrelated to item response bias.

We were unable to find other comparable studies evaluating the QLQ-CIPN20's psychometric properties using Rasch-modeling. However, two PRO measures (Rasch-built Overall Disability Scale, CIPN-Rasch-built Overall Disability Scale) and four provider-completed scales (Total Neuropathy Score, Neuropathy Impairment Scale, Medical Research Council Scale, Inflammatory Neuropathy Cause and Treatment Scale) have been tested using Rasch techniques [40–45]. Van Nes et al. (2011) used Rasch modeling to test a patient-reported disability scale, the Rasch-built Overall Disability Scale (R-ODS), when used by individuals with immune-mediated peripheral neuropathies [40]. The authors provide evidence supporting that the R-ODS is a valid measure of neuropathy-associated functional deficits because items not meeting Rasch model expectations (e.g., unidimensionality, threshold ordering, DIF) were removed. The same Rasch methods were used by Binda et al. (2013) to retest and refine the R-ODS for use in a different population—those with CIPN [41]. When CIPN-R-ODS scores were compared to the traditional NCI-CTC grading scale scores, a measure commonly used in therapeutic oncology clinical trials, the sensory NCI-CTC was inferior in discriminating between various levels of neuropathy severity. Although these varied measures had been previously tested and shown to possess adequate reliability and validity, Rasch analyses uncovered additional psychometric limitations that CTT-based parametric testing had not previously revealed. This highlights the importance of evaluating ordinal-level neuropathy measures using Rasch-based analytic methods. Thus, we emphasize that our findings should not lead to the conclusion that the QLQ-CIPN20 is an inferior measure, especially when compared to other extensively validated PROs such as the FACT GOG/NTX that have not undergone Rasch-model testing. The Rasch-model-based testing approach used in the current study might reveal similar limitations if used to test other CIPN PRO measures.

A limitation of this work is that we were unable to evaluate DIF due to education, race, or ethnicity because these data are not available. Since the QLQ-CIPN20 continues to be used widely across the United States and Europe, it will be important to assess item bias based on cultural and regional differences in item interpretation. Until such analyses can be performed, CIPN intervention study results obtained in the United States using the QLQ-CIPN20 may not be directly comparable to similar studies conducted in other countries. Further, objective CIPN measures such as physical examination or nerve conduction studies were not obtained to validate the presence of CIPN. However, objective CIPN measures do not typically correlate with subjective symptoms [15].

In conclusion, Rasch-model-based testing newly revealed psychometric weaknesses of the QLQ-CIPN20 that can be addressed by eliminating or revising problematic items, and by rescaling the 1–4 item response scale. Alternatively, the new gold standard CIPN measurement approach in future intervention trials could involve use of only the best QLQ-CIPN20 items, which would also allow comparisons across previous trials that utilized the QLQ-CIPN20.

Funding information Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under the Award Number UG1CA189823 (Alliance for Clinical Trials in Oncology NCORP Grant), U10CA180790, U10CA180795, and R03 CA186183-02. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

References

1. Hershman DL, Lacchetti C, Dworkin RH, Lavoie Smith EM, Bleeker J, Cavaletti G, Chauhan C, Gavin P, Lavino A, Lustberg MB, Paice J, Schneider B, Smith ML, Smith T, Terstriep S, Wagner-Johnston N, Bak K, Loprinzi CL (2014) Prevention and management of chemotherapy-induced peripheral neuropathy in survivors of adult cancers: American

Society of Clinical Oncology clinical practice guideline. J Clin Oncol 32(18):1941–1967
2. Seretny M, Currie GL, Sena ES, Ramnarine S, Grant R, MacLeod MR, Colvin LA, Fallon M (2014) Incidence, prevalence, and predictors of chemotherapy-induced peripheral neuropathy: a systematic review and meta-analysis. Pain 155:2461–2470
3. Speck RM, Sammel MD, Farrar JT, Hennessy S, Mao JJ, Stineman MG, DeMichele A (2013) Impact of chemotherapy-induced peripheral neuropathy on treatment delivery in nonmetastatic breast cancer. J Oncol Pract 9:e234–e240
4. Tofthagen C, Overcash J, Kip K (2012) Falls in persons with chemotherapy-induced peripheral neuropathy. Support Care Cancer 20:583–589
5. Bao T, Basal C, Seluzicki C, Li SQ, Seidman AD, Mao JJ (2016) Long-term chemotherapy-induced peripheral neuropathy among breast cancer survivors: prevalence, risk factors, and fall risk. Breast Cancer Res Treat 159:327–333
6. Tofthagen C, Donovan KA, Morgan MA, Shibata D, Yeh Y (2013) Oxaliplatin-induced peripheral neuropathy's effects on health-related quality of life of colorectal cancer survivors. Support Care Cancer 21:3307–3313
7. Desaulniers GA (2011) Chemotherapy induced peripheral neuropathy and subjective sleep quality in non-small cell lung cancer. Oncol Nurs Forum 38:A56
8. Zanville NR, Nudelman KNH, Smith DJ, Von AD, McDonald BC, Champion VL, Saykin AJ (2016) Evaluating the impact of chemotherapy-induced peripheral neuropathy symptoms (CIPN-sx) on perceived ability to work in breast cancer survivors during the first year post-treatment. Support Care Cancer 24:4779–4789
9. Pike CT, Birnbaum HG, Muehlenbein CE, Pohl GM, Natale RB (2012) Healthcare costs and workloss burden of patients with chemotherapy-associated peripheral neuropathy in breast, ovarian, head and neck, and nonsmall cell lung cancer. Chemother Res Pract 2012:1–10
10. Reeves BN, Dakhil SR, Sloan JA, Wolf SL, Burger KN, Kamal A, Le-Lindqwister NA, Soori GS, Jaslowski AJ, Kelaghan J, Novotny PJ, Lachance DH, Loprinzi CL (2012) Further data supporting that paclitaxel-associated acute pain syndrome is associated with development of peripheral neuropathy: North Central Cancer Treatment Group trial N08C1. Cancer 118:5171–5178
11. Griffith K, Merkies ISJ, Hill E, Cornblath D (2010) Measures of chemotherapy-induced peripheral neuropathy: a systematic review of psychometric properties. J Peripher Nerv Syst 15:314–325
12. Cavaletti G, Frigeni B, Lanzani F, Mattavelli L, Susani E, Alberti P, Cortinovis D, Bidoli P (2010) Chemotherapy-induced peripheral neurotoxicity assessment: a critical revision of the currently available tools. Eur J Cancer 46:479–494
13. Smith EM, Cohen JA, Pett MA, Beck SL (2010) The reliability and validity of a modified Total Neuropathy Score-reduced and neuropathic pain severity items when used to measure chemotherapy-induced peripheral neuropathy in patients receiving taxanes and platinums. Cancer Nurs 33:173–183
14. Smith EL, Beck SL, Cohen J (2008) The Total Neuropathy Score (TNS): a tool for measuring chemotherapy-induced peripheral neuropathy. Oncol Nurs Forum 35:96–102
15. Alberti P, Rossi E, Cornblath DR, Merkies IS, Postma TJ, Frigeni B, Bruna J, Velasco R, Argyriou AA, Kalofonos HP, Psimaras D, Ricard D, Pace A, Galie E, Briani C, Dalla Torre C, Faber CG, Lalisang RI, Boogerd W, Brandsma D, Koeppen S, Hense J, Storey D, Kerrigan S, Schenone A, Fabbri S, Valsecchi MG, Cavaletti G, CI-PeriNomS Group (2014) Physician-assessed and patient-reported outcome measures in chemotherapy-induced sensory peripheral neurotoxicity: two sides of the same coin. Ann Oncol 25:257–264
16. Shimozuma K, Ohashi Y, Takeuchi A, Aranishi T, Morita S, Kuroi K, Ohsumi S, Makino H, Mukai H, Katsumata N, Sunada Y, Watanabe T, Hausheer FH (2009) Feasibility and validity of the Patient Neurotoxicity Questionnaire during taxane chemotherapy in a phase III randomized trial in patients with breast cancer: N-SAS BC 02. Support Care Cancer 17:1483–1491
17. Huang HQ, Brady MF, Cella D, Fleming G (2007) Validation and reduction of FACT/GOG-Ntx subscale for platinum/paclitaxel-induced neurologic symptoms: a gynecologic oncology group study. Int J Gynecol Cancer 17:387–393
18. Cella D, Peterman A, Hudgens S, Webster K, Socinski MA (2003) Measuring the side effects of taxane therapy in oncology: the Functional Assessment of Cancer Therapy-taxane (FACT-taxane). Cancer 98:822–831
19. Postma TJ, Aaronson NK, Heimans JJ, Muller MJ, Hildebrand JG, Delattre JY, Hoang-Xuan K, Lanteri-Minet M, Grant R, Huddart R, Moynihan C, Maher J, Lucey R, EORTC Quality of Life G (2005) The development of an EORTC quality of life questionnaire to assess chemotherapy-induced peripheral neuropathy: the QLQ-CIPN20. Eur J Cancer 41:1135–1139
20. Kim HY, Kang JH, Youn HJ, So HS, Song CE, Chae SY, Jung SH, Kim SR, Kim JY (2014) Reliability and validity of the Korean version of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire to assess chemotherapy-induced peripheral neuropathy. J Korean Acad Nurs 44:735–742
21. Cavaletti G, Cornblath DR, Merkies IS, Postma TJ, Rossi E, Frigeni B, Alberti P, Bruna J, Velasco R, Argyriou AA, Kalofonos HP, Psimaras D, Ricard D, Pace A, Galie E, Briani C, Dalla Torre C, Faber CG, Lalisang RI, Boogerd W, Brandsma D, Koeppen S, Hense J, Storey D, Kerrigan S, Schenone A, Fabbri S, Valsecchi MG, CI-PeriNomS Group (2013) The chemotherapy-induced peripheral neuropathy outcome measures standardization study: from consensus to the first validity and reliability findings. Ann Oncol 24:454–462
22. Lavoie Smith EM, Barton DL, Qin R, Steen PD, Aaronson NK, Loprinzi CL (2013) 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 22:2787–2799
23. Kieffer JM, Postma TJ, van de Poll-Franse L, Mols F, Heimans JJ, Cavaletti G, Aaronson NK (2017) Evaluation of the psychometric properties of the EORTC chemotherapy-induced peripheral neuropathy questionnaire (QLQ-CIPN20). Qual Life Res 26:2999–3010
24. Smith EML, Knoerl R, Yang JJ, Kanzawa-Lee G, Lee D, Bridges CM (2018) In search of a gold standard patient-reported outcome measure for use in chemotherapy-induced peripheral neuropathy clinical trials. Cancer Control 25:1–10
25. Smith EML, Banerjee T, Yang JJ, Bridges C, Alberti P, Sloan J, Loprinzi C (2018) Psychometric testing of the European Organisation for Research and Treatment of Cancer (EORTC) QLQ-CIPN20 using pooled Chemotherapy-Induced Peripheral Neuropathy Outcome Measures Standardization (CI-PeriNomS) and Alliance for Clinical Trials in Oncology (Alliance) A151408 study data. Cancer Nurs:1. https://doi.org/10.1097/NCC.0000000000000596
26. Smith EML, Haupt R, Kelly J, Lee D, Kanzawa-Lee G, Knoerl R, Bridges C, Alberti P, Prasertsri N, Donohoe C (2017) The content validity of a chemotherapy-induced peripheral neuropathy patient-reported outcome measure. Oncol Nurs Forum 44:580–588
27. Mols F, van de Poll-Franse LV, Vreugdenhil G, Beijers AJ, Kieffer JM, Aaronson NK, Husson O (2016) Reference data of the European Organisation for Research and Treatment of Cancer (EORTC) QLQ-CIPN20 questionnaire in the general Dutch population. Eur J Cancer 69:28–38
28. Le-Rademacher J, Kanwar R, Seisler D, Pachman DR, Qin R, Abyzov A, Ruddy KJ, Banck MS, Lavoie Smith EM, Dorsey SG, Aaronson NK, Sloan J, Loprinzi CL, Beutler AS (2017) Patient-reported (EORTC QLQ-CIPN20) versus physician-reported

(CTCAE) quantification of oxaliplatin- and paclitaxel/carboplatin-induced peripheral neuropathy in NCCTG/Alliance clinical trials. Support Care Cancer 25:3537–3544
29. Cappelleri JC, Jason Lundy J, Hays RD (2014) Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. Clin Ther 36:648–662
30. Merkies ISJ, Lauria G, Faber CG (2012) Outcome measures in peripheral neuropathies: requirements through statements. Curr Opin Neurol 25:556–563
31. Tennant A, Conaghan PG (2007) The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 57: 1358–1362
32. Pallant JF, Tennant A (2007) An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 46:1–18
33. Loprinzi CL, Reeves BN, Dakhil SR, Sloan JA, Wolf SL, Burger KN, Kamal A, Le-Lindqwister NA, Soori GS, Jaslowski AJ, Novotny PJ, Lachance DH (2011) Natural history of paclitaxel-associated acute pain syndrome: prospective cohort study NCCTG N08C1. J Clin Oncol 29:1472–1478
34. Loprinzi CL, Qin R, Dakhil SR, Fehrenbacher L, Flynn KA, Atherton P, Seisler D, Qamar R, Lewis GC, Grothey A (2013) Phase III randomized, placebo-controlled, double-blind study of intravenous calcium and magnesium to prevent oxaliplatin-induced sensory neurotoxicity (N08CB/Alliance). J Clin Oncol 32:997–1005
35. Barton DL, Wos EJ, Qin R, Mattar BI, Green NB, Lanier KS, Bearden JD 3rd, Kugler JW, Hoff KL, Reddy PS, Rowland KM, Jr RM, Christensen B, Loprinzi CL (2011) A double-blind, placebo-controlled trial of a topical treatment for chemotherapy-induced peripheral neuropathy: NCCTG trial N06CA. Support Care Cancer 19:833–841
36. Leal AD, Qin R, Atherton PJ, Haluska P, Behrens RJ, Tiber CH, Watanaboonyakhet P, Weiss M, Adams PT, Dockter TJ, Loprinzi CL, for the Alliance for Clinical Trials in Oncology (2014) North Central Cancer Treatment Group/Alliance trial N08CA-the use of glutathione for prevention of paclitaxel/carboplatin-induced peripheral neuropathy: a phase 3 randomized, double-blind, placebo-controlled study. Cancer 120:1890–1897
37. Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group (2001) The EORTC QLQ-C30 Scoring Manual, 3rd edn. European Organisation for Research and Treatment of Cancer, Brussels
38. Mair P, Hatzinger R (2007) Extended Rasch modeling: the eRm package for the application of IRT models in R. J Stat Softw 20: 1–20
39. Mair P, Hatzinger R (2007) CML based estimation of extended Rasch models with the eRm package in R. Psychol Sci 49:26–43
40. van Nes SI, Vanhoutte EK, van Doorn PA, Hermans M, Bakkers M, Kuitwaard K, Faber CG, Merkies ISJ (2011) Rasch-built Overall Disability Scale (R-ODS) for immune-mediated peripheral neuropathies. Neurology 76:337–345
41. Binda D, Vanhoutte EK, Cavaletti G, Cornblath DR, Postma TJ, Frigeni B, Alberti P, Bruna J, Velasco R, Argyriou AA, Kalofonos HP, Psimaras D, Ricard D, Pace A, Galie E, Briani C, Dalla Torre C, Lalisang RI, Boogerd W, Brandsma D, Koeppen S, Hense J, Storey D, Kerrigan S, Schenone A, Fabbri S, Rossi E, Valsecchi MG, Faber CG, Merkies ISJ, CI-PeriNomS study group, Galimberti S, Lanzani F, Mattavelli L, Piatti ML, Bidoli P, Cazzaniga M, Cortinovis D, Lucchetta M, Campagnolo M, Bakkers M, Brouwer B, Boogerd W, Grant R, Reni L, Piras B, Pessino A, Padua L, Granata G, Leandri M, Ghignotti I, Plasmati R, Pastorelli F, Heimans JJ, Eurelings M, Meijer RJ, Grisold W, Lindeck Pozza E, Mazzeo A, Toscano A, Russo M, Tomasello C, Altavilla G, Penas Prado M, Dominguez Gonzalez C, Dorsey SG (2013) Rasch-built Overall Disability Scale for patients with chemotherapy-induced peripheral neuropathy (CIPN-R-ODS). Eur J Cancer 49:2910–2918
42. Binda D, Cavaletti G, Cornblath DR, Merkies ISJ, CI-PeriNomS study group (2015) Rasch-transformed Total Neuropathy Score clinical version (RT-TNSc(©)) in patients with chemotherapy-induced peripheral neuropathy. J Peripher Nerv Syst 20:328–332
43. Draak THP, Vanhoutte EK, van Nes SI, Gorson KC, Van der Pol W, Notermans NC, Nobile-Orazio E, Lewis RA, Leger J, Van den Bergh PY, Lauria G, Bril V, Katzberg H, Lunn MPT, Pouget J, van der Kooi AJ, Hahn AF, van den Berg LH, van Doorn PA, Cornblath DR, Faber CG, ISJ M, PeriNomS Study Group (2015) Comparing the NIS vs MRC and INCAT sensory scale through Rasch analyses. J Peripher Nerv Syst 20:277–288
44. Sadjadi R, Reilly MM, Shy ME, Pareyson D, Laura M, Murphy S, Feely SME, Grider T, Bacon C, Piscosquito G, Calabrese D, Burns TM (2014) Psychometrics evaluation of Charcot-Marie-Tooth Neuropathy Score (CMTNSv2) second version, using Rasch analysis. J Peripher Nerv Syst 19:192–196
45. Vanhoutte EK, Faber CG, van Nes SI, Cats EA, Van der Pol W, Gorson KC, van Doorn PA, Cornblath DR, van den Berg LH, Merkies ISJ, PeriNomS Study Group (2015) Rasch-built Overall Disability Scale for multifocal motor neuropathy (MMN-RODS(©)). J Peripher Nerv Syst 20:296–305