



Published in final edited form as:

Pharmacogenomics J. 2019 October ; 19(5): 411–416. doi:10.1038/s41397-019-0093-1.

Concerns regarding use of Patient Reported Outcomes in Biomarker Studies of Chemotherapy-Induced Peripheral Neuropathy

Daniel L. Hertz^{1,2}

¹Department of Clinical Pharmacy, University of Michigan College of Pharmacy, Ann Arbor, MI, United States, 48109-1065

²University of Michigan Rogel Cancer Center, Ann Arbor, MI

Abstract

Clinician-rated toxicity data has been systematically collected within oncology clinical research using the National Cancer Institute's CTCAE scale, providing estimates of the occurrence and severity of toxicity from cancer treatment. CTCAE is being supplemented by collection of patient-reported outcome (PRO) toxicity within clinical research and clinical practice, where PRO has demonstrable benefits. There is general agreement that PRO data is more sensitive and reliable than CTCAE data, particularly for subjective adverse effects. Based on this premise, researchers have begun to use PRO toxicity data collected within prospective clinical trials as the primary endpoint to discover pharmacogenetic and other predictive biomarkers of treatment-related toxicity. This perspective raises caution about the superiority of PRO data to CTCAE data for biomarker research, particularly in regards to chemotherapy-induced peripheral neuropathy (PN). The reader is provided an introduction to PRO and their integration into clinical research and practice, comparisons of PN data collected by PRO and CTCAE, examples of attempts to use PRO PN data for biomarker discovery, and evidence suggesting that PRO may not be superior to CTCAE for PN biomarker studies. The perspective concludes with a proposed approach for empirically testing whether PRO or CTCAE data is the better option for use in PN biomarker research, which can serve as a model for similar comparisons within other treatment-related toxicities.

Pharmacogenetic analyses require accurate treatment outcomes data to maximize the likelihood of successfully discovering a genetic predictor. Integration of patient-reported outcome (PRO) toxicity data into oncology clinical research and practice, where it has documented benefit, provides an alternative to clinician-graded toxicity data; however, it is unclear whether the benefits of PRO extend to pharmacogenetic discovery.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding Author: Daniel L Hertz, 428 Church St., Room 3054 College of Pharmacy, Ann Arbor, MI 48109-1065, Office phone: (734) 763-0015, Fax: (734) 763-4480, DLHertz@med.umich.edu.

Conflict of Interest: The authors declare no potential conflicts of interest

Clinician-Graded and Patient-Reported Toxicity Data in Clinical Research and Practice

Since the mid-1980s the National Cancer Institute (NCI) has mandated that physician-reported toxicity data be collected during oncology clinical trials using the NCI Common Terminology Criteria for Adverse Events (NCI CTCAE) at a minimum. The current version, CTCAE V4.0, enables collection of more than 1,000 distinct AEs on a clinician-graded 0–5 scale. There is concern that reliance on CTCAE grading has led to incomplete recognition and documentation of treatment-related AEs, particularly for subjective toxicities such as fatigue, nausea, and peripheral neuropathy (PN)¹. Subjective toxicity information must be described by the patient to the clinician, who must recognize that it represents an AE, assess the severity grade based on the CTCAE, and document the toxicity in the study record. This complex process introduces several opportunities for information to be misinterpreted or lost, particularly in multi-center clinical trials in which toxicity assessment is conducted by many different clinicians with diverse approaches and varying levels of attentiveness to toxicity assessment and documentation (Figure 1).

Since the early 2000s, there has been a major push toward supplementing CTCAE with validated patient-reported outcome (PRO) information, which has several potential advantages. Incidence and severity of toxicity tend to be higher when using PRO than CTCAE, and matched comparisons generally report poor to modest agreement, which is lowest for subjective toxicities. PRO can also provide more detailed symptom information and use a larger dynamic range, potentially enabling earlier detection and identification of subtle changes in severity². In parallel with the integration of PRO into clinical research, real-time collection and sharing of PRO with clinicians during treatment has been reported to improve patient survival³. One of the main benefits of using PRO in clinical practice is that they reflect the effect the AE is having on the patient's quality of life, enabling clinicians and patients to make shared treatment decisions based on the individual patient's personal preferences and treatment goals.

Comparison of CTCAE and PRO as an Endpoint for Biomarker Studies of Peripheral Neuropathy (PN)

PN data collected via CTCAE has substantial inter-observer variability and a known floor effect⁴. Several PRO tools have been developed to collect chemotherapy-induced PN including the Patient Neurotoxicity Questionnaire (PNQ), Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity (FACT/GOG-Ntx), and European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Chemotherapy Induced Peripheral Neuropathy (EORTC CIPN20)⁵. More than a dozen studies have directly compared one of these PRO instruments, scored according to published guidelines, with CTCAE (Table 1). Across these studies, the incidence and severity of PN is consistently higher when collected via PRO, and the correlations between CTCAE and PRO PN are poor to moderate (correlation coefficients: 0.2–0.7). The limited correlation is often used as evidence to support the conclusion that PRO PN data are more sensitive and reliable

than CTCAE data, however, this conclusion is difficult to empirically test without an objective benchmark to anchor comparisons.

It is critically important that pharmacogenomics studies use accurate treatment outcomes data. The first wave of genome-wide association studies in taxane-induced PN, conducted in clinical trials designed prior to the introduction of PRO, relied on CTCAE data as the endpoint of interest and frequently noted the limitations of CTCAE graded PN⁶⁻⁸. As PN collection within clinical research has transitioned from CTCAE to PRO, the biomarker research community has followed. Several recent PN biomarker studies have elected to use PRO data as the primary endpoint or conducted parallel analyses of PRO and CTCAE data (Table 2). Two of these studies were secondary analyses of PRO data collected within prospective interventional clinical trials, within which systematically collected CTCAE data was also available^{9, 10}, whereas in the remaining three cases investigators decided to collect PRO instead of CTCAE within prospective registries or case-control studies of PN¹¹⁻¹³. These studies also used a variety of strategies for translating PRO data into an outcome for analysis, ranging from using the continuous PRO score to arbitrarily or empirically selecting thresholds to dichotomize patients into PN “cases” and no-PN “controls,” as described in Table 2. The decision to use PRO instead of CTCAE PN data in these biomarker discovery studies was likely based on the assumption that the benefits of PRO for clinical research and practice extend to biomarker research. Indeed, the ability to collect PRO data more frequently, identification of a greater number of PN cases via PRO, enhanced sensitivity of PRO to detect subtle changes in severity, and improved specificity of PN symptom collection via PRO should substantially improve analytical power for PN biomarker discovery.

While the benefits of PRO for clinical research and practice are established, I recommend the pharmacogenomics community pause to consider whether these benefits extend to biomarker research, specifically regarding PN. Although PN is grouped with other subjective toxicities for which PRO are thought to be superior to CTCAE, there is evidence that CTCAE data for PN are more reliable than CTCAE data for other subjective toxicities (i.e., anxiety, pain). Comparisons of seven subjective symptoms assessed via CTCAE by multiple clinicians found that PN had the highest inter-observer correlation coefficient (ICC=0.71) and the lowest incidence of CTCAE assessed by two observers differing by more than 1 grade (<1%)¹. Furthermore, the floor effect noted in CTCAE PN data may not reflect a limitation of the scale; clinicians delay or discontinue treatment when they identify mild to moderate PN, therefore, it would be expected that grade 4 PN would rarely be encountered in CTCAE data.

More generally, the greater incidence of “severe” toxicity detected using PRO instruments may not actually indicate that CTCAE fails to detect clinically meaningful events, as has been generally assumed. Matched comparisons of CTCAE and PRO data for several toxicities, which did not include PN, found that despite detecting fewer toxicity events, CTCAE data was more strongly associated with future rehospitalization and survival than PRO data¹⁴. This suggests an alternative, but seldom considered, explanation. The discrepancy between PRO and CTCAE may reflect differences in the standard of “severe” between patients, who have likely never experienced the toxicity, and clinicians, who regularly interact with patients experiencing treatment-related toxicity and have seen cases

of truly severe toxicity. The FACT/GOG-NTx and EORTC CIPN20 ask patients to rate a series of PN-related symptoms on a numerical scale (0 or 1 thru 4), corresponding to textual descriptors of severity. There is likely substantial variability between patients in their definition of “A Little” and “Quite a Bit,” as most patients have no prior experience or knowledge of PN and wouldn’t know the upper limit of PN severity corresponding to a score of “Very Much.” This has been empirically demonstrated for several subjective symptoms, which again did not include PN, by modeling matched PRO and clinician-documented data to confirm that patients have lower and more variable thresholds for classifying symptoms as severe¹⁵. This finding is consistent with the previously described evidence that there is strong agreement between clinicians when grading PN (ICC=0.71). However, this estimate was based on a small group of clinicians at a single institution and agreement would likely decrease in a larger, more diverse group of clinicians, particularly if those clinicians were not carefully assessing and documenting PN.

One of the primary advantages to using PRO data for clinical practice is the integration of a patient’s perception about the effect of toxicity on their life. In clinical practice, whether a patient is experiencing an objectively “severe” adverse effect is less relevant than whether the patient considers this adverse effect intolerable. Patients may have limited basis to accurately judge PN severity, but they are the only relevant assessor of whether the toxicity is tolerable, given their lifestyle and treatment priorities. There is undoubtedly a subset of patients for whom fine motor skills and tactile sensitivity are critical. For example a blind patient who reads Braille or a patient who loves to sew would consider an objectively minor amount of PN intolerable, and likely rate it as “Very Much” on a PRO. While this information is extremely useful for making treatment decisions for this patient in clinical care, classifying this patient as a case of “severe” PN in a pharmacogenetic analysis would be inappropriate and decrease the likelihood of identifying a genetic PN biomarker (Figure 2).

Empirical Comparison of CTCAE and PRO as Endpoints for Biomarker Studies

Biomarkers that are strongly predictive of treatment outcomes should be identifiable regardless of limitations of the endpoint selected. However, most clinical outcomes, including PN, seem to be multi-factorial and require highly reliable outcomes data to discover and replicate individual predictors that have relatively small effects on overall risk. In these cases it is critical that the outcomes data used in pharmacogenomics studies are as accurate as possible. We recently reported the results of a prospective observational clinical trial that characterized the relationship between systemic paclitaxel concentrations and PN. To our surprise, we did not detect an association with PRO PN, collected via EORTC CIPN-20, but detected a strong association with PN-induced treatment disruption (dose decreases, delays, or discontinuations due to PN)¹². Treatment disruptions are assumed to reflect CTCAE grade, which we unfortunately did not collect within this cohort study due to our previously held confidence in the superiority of PRO data.

This finding, and the above described data suggesting greater overall concordance between clinicians than patients in assessing toxicity severity, should motivate the biomarker research community to more carefully consider which toxicity data source should be used as the primary endpoint for biomarker studies of PN. There are some best practices for collecting PN data that should be implemented regardless of the data source, such as collecting as many PN data types as feasible and assessing PN prior to and at the end of treatment, or at the time of a PN-induced treatment disruption. PN biomarker discovery analyses should *a priori* select a primary endpoint, account for cumulative dosing, and avoid unnecessarily collapsing continuous or ordinal scales into “cases” and “controls.” PN biomarker discovery studies could also consider including an objective measure of PN, such as thermal discrimination, vibration perception, or nerve conduction, which have been included in some comparisons of PRO and CTCAE, as indicated in Table 1. These quantitative measurements should further improve analytical power to discover PN predictors, however, it is unclear whether these highly controlled, experimental assessments reflect clinically relevant PN. An important next step in this field is to determine what magnitude of change in CTCAE, PRO, or an objective PN measure is “clinically-relevant,” since a biomarker must predict a clinically relevant endpoint to be useful in clinical practice. There is currently no established approach to translate changes in PRO to changes in CTCAE or to a clinically-relevant worsening of PN, necessitating investigators to select arbitrary thresholds in confirmatory studies. To that end, it could be valuable to the field to use the clinically-meaningful endpoint of PN-induced treatment disruption as a benchmark to estimate a clinically relevant magnitude of change in PRO PN.

CTCAE and PRO PN data each have benefits and limitations when used as the endpoint in biomarker discovery, and, in my opinion, it is unclear which is superior for this purpose. It may be possible to empirically test whether CTCAE or PRO are the optimal endpoint for biomarker discovery studies of PN, and other subjective AEs. In our paclitaxel pharmacokinetics project we conducted parallel analyses of pharmacokinetics as a predictor of two endpoints; EORTC CIPN-20, a PRO, and PN-induced treatment disruption, a surrogate of CTCAE. Our results, in a relatively small pilot study, indicate that PK is more strongly predictive of CTCAE than PRO. Using a conceptually similar approach of conducting parallel analyses of multiple endpoints but employing a single validated predictor variable could be a reasonable approach to determining which endpoint should be used in future analyses to discover additional predictor variables. For example, for paclitaxel-induced PN, perhaps cumulative dose, which is an established predictor of PN severity during paclitaxel treatment, could be used as an indicator variable in a secondary analysis of a large prospective clinical trial to empirically determine whether PRO or CTCAE data should be used as the endpoint in future pharmacogenetic discovery efforts.

In conclusion, although PRO have demonstrable benefits in clinical research and practice, it has not been adequately established that these benefits extend to biomarker research. There is suggestive evidence that the increased frequency and severity of toxicity detected via PRO reflects differences in patient’s assessments of severity, or perhaps tolerability, making its use as an endpoint for pharmacogenomics research problematic. While researchers continue to develop improved approaches to collect toxicity data, I recommend the biomarker research community pause to consider whether PRO or CTCAE data should be used as the

primary endpoint within pharmacogenomics studies of PN and more generally in biomarker studies of subjective treatment-related toxicities.

Acknowledgements:

I would like to acknowledge Lynn Henry, Karen Farris, Steve Erickson, Ellen Lavoie Smith, and Jane Perlmutter for their critical review and feedback.

References:

1. Atkinson TM, Li Y, Coffey CW, Sit L, Shaw M, Lavene D, et al. Reliability of adverse symptom event reporting by clinicians. *Qual Life Res* 2012; 21(7): 1159–1164. [PubMed: 21984468]
2. Di Maio M, Basch E, Bryce J, Perrone F. Patient-reported outcomes in the evaluation of toxicity of anticancer treatments. *Nature reviews Clinical oncology* 2016; 13(5): 319–325.
3. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. *JAMA* 2017; 318(2): 197–198. [PubMed: 28586821]
4. Postma TJ, Heimans JJ, Muller MJ, Ossenkoppele GJ, Vermorken JB, Aaronson NK. Pitfalls in grading severity of chemotherapy-induced peripheral neuropathy. *Ann Oncol* 1998; 9(7): 739–744. [PubMed: 9739440]
5. Curcio KR. Instruments for Assessing Chemotherapy-Induced Peripheral Neuropathy: A Review of the Literature. *Clin J Oncol Nurs* 2016; 20(2): 144–151. [PubMed: 26991707]
6. Hertz DL, Owzar K, Lessans S, Wing C, Jiang C, Kelly WK, et al. Pharmacogenetic Discovery in CALGB (Alliance) 90401 and Mechanistic Validation of a VAC14 Polymorphism that Increases Risk of Docetaxel-Induced Neuropathy. *Clin Cancer Res* 2016; 22(19): 4890–4900. [PubMed: 27143689]
7. Baldwin RM, Owzar K, Zembutsu H, Chhibber A, Kubo M, Jiang C, et al. A genome-wide association study identifies novel loci for paclitaxel-induced sensory peripheral neuropathy in CALGB 40101. *Clin Cancer Res* 2012; 18(18): 5099–5109. [PubMed: 22843789]
8. Schneider BP, Li L, Radovich M, Shen F, Miller KD, Flockhart DA, et al. Genome-Wide Association Studies for Taxane-Induced Peripheral Neuropathy in ECOG-5103 and ECOG-1199. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2015; 21(22): 5082–5091. [PubMed: 26138065]
9. Beutler AS, Kulkarni AA, Kanwar R, Klein CJ, Therneau TM, Qin R, et al. Sequencing of Charcot-Marie-Tooth disease genes in a toxic polyneuropathy. *Annals of Neurology* 2014; 76(5): 727–737. [PubMed: 25164601]
10. Park SB, Kwok JB, Asher R, Lee CK, Beale P, Selle F, et al. Clinical and genetic predictors of paclitaxel neurotoxicity based on patient- versus clinician-reported incidence and severity of neurotoxicity in the ICON7 trial. *Ann Oncol* 2017; 28(11): 2733–2740. [PubMed: 29117336]
11. Chen EI, Crew KD, Trivedi M, Awad D, Maurer M, Kalinsky K, et al. Identifying Predictors of Taxane-Induced Peripheral Neuropathy Using Mass Spectrometry-Based Proteomics Technology. *PloS one* 2015; 10(12): e0145816. [PubMed: 26710119]
12. Hertz DL, Kidwell KM, Vangipuram K, Li F, Pai MP, Burness M, et al. Paclitaxel Plasma Concentration After the First Infusion Predicts Treatment-Limiting Peripheral Neuropathy. *Clin Cancer Res* 2018.
13. Dolan ME, El Charif O, Wheeler HE, Gamazon ER, Ardeshir-Rouhani-Fard S, Monahan P, et al. Clinical and Genome-Wide Analysis of Cisplatin-Induced Peripheral Neuropathy in Survivors of Adult-Onset Cancer. *Clin Cancer Res* 2017; 23(19): 5757–5768. [PubMed: 28611204]
14. Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst* 2009; 101(23): 1624–1632. [PubMed: 19920223]
15. Atkinson TM, Rogak LJ, Heon N, Ryan SJ, Shaw M, Stark LP, et al. Exploring differences in adverse symptom event grading thresholds between clinicians and patients in the clinical trial setting. *J Cancer Res Clin Oncol* 2017; 143(4): 735–743. [PubMed: 28093637]

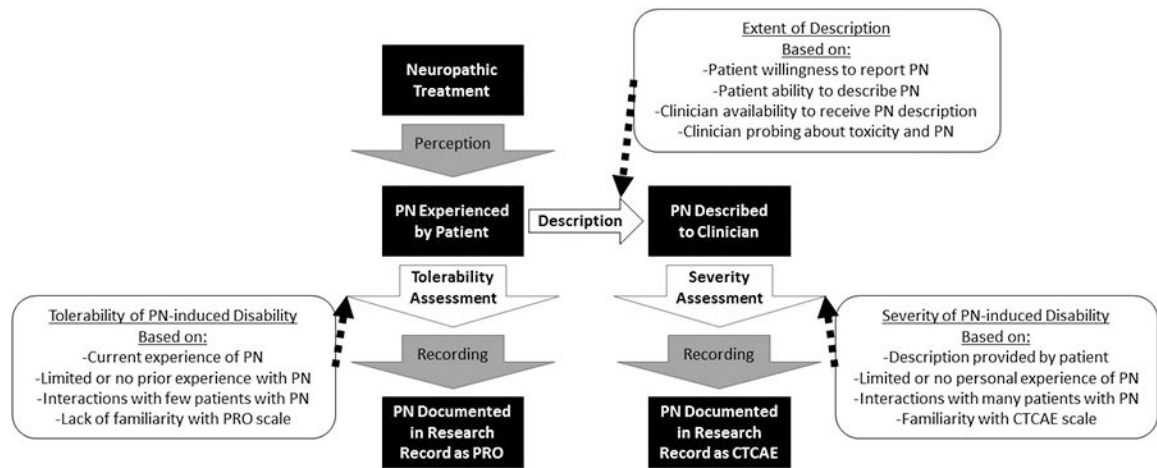


Figure 1: Flow of Information from a Treated Patient to PRO or CTCAE PN Data.

The flow of information from patient to the research record as either PRO or CTCAE includes several steps during which information can be lost or misinterpreted. Generation of CTCAE data requires a description by the patient to the clinician, which represents an additional opportunity for information loss. PRO and CTCAE each require an assessment, either the patient (PRO) assesses the tolerability of PN or the clinician (CTCAE) assesses the severity of PN; it is possible that clinicians have a more consistent scale for assessment than do patients.

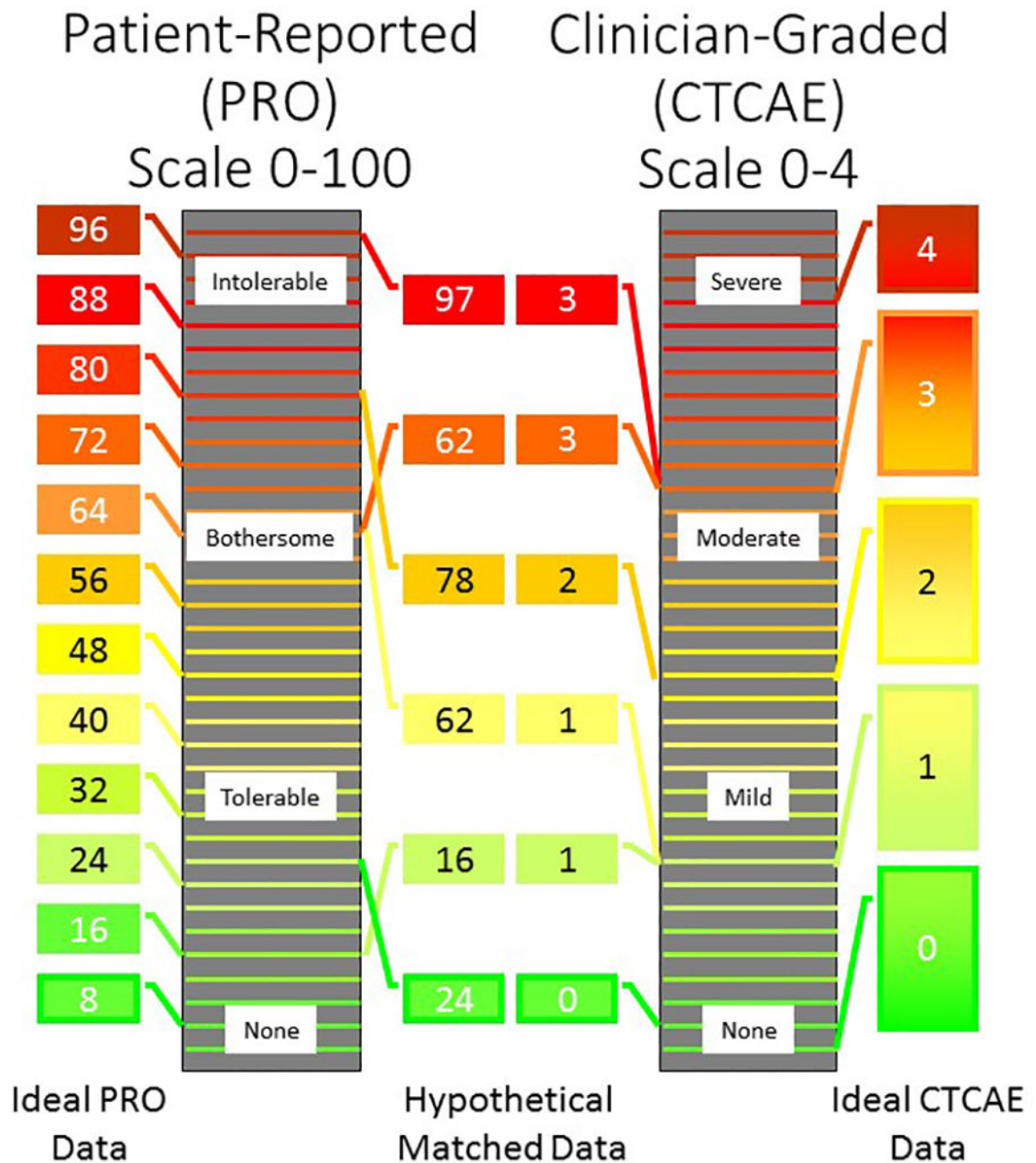


Figure 2: Assessment of PN Using PRO and CTCAE.

Schematic representation of the PRO (left) and CTCAE (right) grading scales for peripheral neuropathy (PN). The height of the box represents the extent of PN and the number inside the box is the tolerability assigned by the patient (left) or severity assigned by the clinician (right). Ideally, all patients (far left) and clinicians (far right) would have consistent scales, in which case PRO would likely be a superior endpoint for biomarker research due to its increased sensitivity. However, there is evidence that thresholds of severity are more consistent for clinicians (inner right) than patients (inner left)¹⁵. For example, two patients rated their PN a 62 on a scale of 0–100, but the actual extent of PN was very different

between these patients. This increased inter-rater variability could be a critical drawback for using PRO as an endpoint in PN biomarker studies.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

Comparisons of Patient and Clinician Reported Peripheral Neuropathy

PRO and Subscale	n	Collection Time Point	Objective Assessments ^b	Findings	Journal Reference
FACT/GOG-Ntx	99	BL (controls) or Post-Tx (cases)	Yes	Ntx significantly worse in cases than controls (p<0.001)	Calhoun, Welshman et al. 2003
	27	BL, each cycle, 3M Post-Tx	Yes	Ntx and CTCAE Spearman's $\rho=-0.59$ (p=0.001)	Moore, Donnelly et al. 2003
	134	Each cycle		Ntx sensory ROC curve AUC=0.90	Huang, Brady et al. 2007
	300	BL, C3, C5, C7, 7M and 12M		Ntx and CTCAE sensory Spearman's $\rho=0.45$	Shimozuma, Ohashi et al. 2009
	422	BL, within 4W of Tx end, 6M Post-Tx		Ntx and CTCAE sensory Pearson r=0.69	Cella, Huang et al. 2010
EORTC QLQ-CIPN20	29	Each cycle	Yes	Ntx sensory worse in patients with sensory CTCAE 1 (p<0.001)	Griffith, Couture et al. 2014
	230	BL		CIPN20 sensory and CTCAE sensory Pearson r=0.20 (p 0.01)	Lavoie Smith, Barton et al. 2013
	281	Post-Tx	Yes	CIPN20 and CTCAE sensory Spearman's $\rho=0.47$ (p 0.007)	Cavaletti, Cornblath et al. 2013
	281	Post-Tx	Yes	CIPN20 sensory and CTCAE sensory linear correlation $\beta=13.3$ (p<0.001)	Alberti, Rossi et al. 2014
	414	Each cycle		CIPN20 sensory and CTCAE linear association (p<0.0001)	Le-Rademacher, Kanwar et al. 2017
PNQ	300	BL, C3, C5, C7, 7M and 12M		PNQ sensory and CTCAE sensory Spearman's $\rho=0.44$	Shimozuma, Ohashi et al. 2009
	35	BL, 8W, 16W		PNQ sensory and CTCAE sensory Spearman's $\rho=0.58$	Kuroi, Shimozuma et al. 2009
	20	Post-Tx	Yes	PNQ and CTCAE sensory statistical comparison NR	Bennett, Park et al. 2012

^a: Used NCI-Sanofi grading scale, which is similar to NCI CTCAE^b: Objective assessments included: Thermal discrimination, Vibration perception, Nerve conduction, Quantitative sensory testing, Mechanical stimulation with a monofilament, Grip strength, Pinprick sensibility, Deep tendon reflexes

Acronyms: NR: Not reported, ROC: receiver operating curve, BL: baseline, Post-Tx: after treatment ended, C: cycle, W: week, M: month

Table 2: Selected Biomarker Studies that have used Patient-Reported Outcomes as the Peripheral Neuropathy Endpoint

PRO Instrument	PRO Scale	Outcome used in analysis	Parallel analysis of CTCAE	Biomarker	n	Findings	Ref
FACT-NTx	NTx score	Case defined as 20% worsening of cumulative score from baseline	No	Proteomics	17	Hypothesis-generating association with a protein signature	11
EORTC QLQ-CIPN20	CIPN8*	Cumulative score at each week of treatment	No, but used PN-induced treatment disruption	Paclitaxel PK and clinical variables	60	PRO associated with cumulative dose, age, and alcohol intake. Treatment disruption associated with cumulative dose and paclitaxel pharmacokinetics	12
	CIPN8*	Ordinal groups empirically derived from cumulative score	No	GWAS and clinical variables	680	PRO associated with age, smoking, and drinking. Hypothesis-generating associations with genetics	13
	CIPN20	Cases and controls defined by selecting phenotypic extremes from distribution of increase in cumulative score throughout treatment	No	Candidate gene sequencing	119	PRO-derived cases carried more deleterious variants in genes associated with hereditary neuropathy	9
EORTC QLQ-OV28	Two neuropathy questions	Cases defined as response of "Quite a Bit" or "Very Much" on either question	Yes	Candidate genotypes and clinical variables	454	PRO associated with age, residual disease. CTCAE associated with age, bevacizumab, and bowel resection. Hypothesis-generating associations with genetics	10

*: CIPN9 Sensory subscale without question 18 (difficulty hearing)

Abbreviations: PK: pharmacokinetics, GWAS: Genome-wide association study