Title

Evaluating Treatment Tolerability Using the Toxicity Index With Patient-Reported Outcomes Data

Journal

Journal of Pain and Symptom Management, 63(2)

ISSN

0885-3924

Authors

Langlais, Blake
Mazza, Gina L
Thanarajasingam, Gita
et al.

Publication Date

2022-02-01

DOI

10.1016/j.jpainsymman.2021.07.031

Peer reviewed

HHS Public Access

Author manuscript

J Pain Symptom Manage. Author manuscript; available in PMC 2023 February 01.

Published in final edited form as:

J Pain Symptom Manage. 2022 February ; 63(2): 311–320. doi:10.1016/j.jpainsymman.2021.07.031.

Evaluating treatment tolerability using the toxicity index with patient-reported outcomes data

Blake Langlais, MS[1], **Gina Mazza, PhD**[1], **Gita Thanarajasingam, MD**[2], **Lauren J Rogak, MA**[3], **Brenda Ginos, MS**[1], **Narre Heon, MPA**[3], **Howard I Scher, MD**[3], **Gisela Schwab, MD**[4], **Patricia Ganz, MD**[5], **Ethan Basch, MD**[3],[6], **Amylou C Dueck, PhD**[1],*

[1]Mayo Clinic, Department of Quantitative Health Sciences, Division of Biostatistics and Clinical Trials, Phoenix, Arizona, USA

[2]Mayo Clinic, Department of Hematology, Mayo Clinic, Rochester, Minnesota, USA

[3]Memorial Sloan Kettering Cancer Center, New York, New York, USA

[4]Exelixis, Inc., Alameda, California, USA

[5]Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, California, USA

[6]UNC Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, USA

Abstract

Context—Summarizing longitudinal symptomatic adverse events during clinical trials is necessary for understanding treatment tolerability. The Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) provides insight for capturing treatment tolerability within trials. Tolerability summary measures, such as the maximum score, are often used to communicate the potential negative symptoms both in the medical literature and directly to patients. Commonly, the proportions of present and severe symptomatic adverse events are used and reported between treatment arms among adverse event types. The toxicity index is also a summary measure previously applied to clinician-reported CTCAE data.

Objectives—Apply the toxicity index to PRO-CTCAE data from the COMET-2 trial alongside the maximum score, then present and discuss considerations for using the toxicity index as a summary measure for communicating tolerability to patients and clinicians.

Methods—Proportions of maximum PRO-CTCAE severity levels and median toxicity index were computed by arm using all trial data and adjusting for baseline symptoms.

*Corresponding Author Amylou C Dueck (dueck.amylou@mayo.edu), Address: Mayo Clinic, 13400 E. Shea Blvd., Scottsdale, AZ 85259.

Results—Group-wise statistical differences were similar whether using severity level proportions or the toxicity index. The impact of adjusting for baseline symptoms was equivalently seen when comparing arms using severity rates or the toxicity index.

Conclusion—The toxicity index is a useful method when ranking patients from those with the least to most symptomatic adverse event burden. This study showed the toxicity index can be applied to PRO-CTCAE data. Though as a tolerability summary measure, further study is needed to provide a clear clinical or patient-facing interpretation of the toxicity index.

Keywords

PRO-CTCAE; tolerability; toxicity index; patient-reported outcomes; symptomatic adverse event; cancer clinical trials

Introduction

Cancer clinical trials have utilized the National Cancer Institute's (NCI) Common Terminology Criteria for Adverse Events (CTCAE) for decades to facilitate a standardized process for clinicians to observe and rate therapeutic toxicities, or side effects, impacting patient health. In this setting, toxicity generally refers the level of damage an experimental treatment can have on the body's organs or entire system. Though the CTCAE is critical for tracking toxicity, clinicians may miss up to half of symptomatic burden related to treatment side effects compared to routine patient self-reporting [1, 2]. As such, a Patient-Reported Outcomes (PRO) version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) was developed. The PRO-CTCAE is a self-administered library of 124 items evaluating the frequency, severity, interference, or presence of 78 symptomatic adverse events. More broadly, patient reported outcomes (PROs) are reports directly made by the patient about their health status without inference from an observing clinician [3]. In cases where PROs encompass symptoms related to treatment side effects such as the PRO-CTCAE, these PROs provide valuable information about toxicity. In a trial, information about toxicity is one of the contributing factors in understanding the extent to which overt adverse events affect a patient's willingness and ability to continue the treatment regimen [4, 5]. This is referred to as treatment tolerability. Implementing a trial-specific subset of the PRO-CTCAE library allows clinicians and investigators a more comprehensive understanding of tolerability from a patient's point of view, specifically the negative impacts of treatment so it can be weighed against a demonstrated efficacy.

Continued use of patient-centered health outcome measures motivated the NCI Cancer Moonshot Initiative to accelerate improvements in toxicity and tolerability reporting and analysis methods. Traditional methods of reporting clinical trial tolerability consist of aggregating a patient's overall adverse event experience into a single numeric value. Typically, these tolerability summary measures reflect the single most severe symptomatic adverse event during a trial period for each patient. Conveying tolerability by treatment arm can then be achieved by reporting the proportion of patients, whom at worst, responded with present or severe symptomatic adverse events (e.g., "45% of patients experienced grade 3 or higher pain severity while on Treatment X"). These unambiguous dichotomizations of severity levels provide clinicians an easily interpretable metric for communicating with

patients in a way they are likely to understand. This may enable patients to set appropriate expectations, empower them to take part in the treatment plan, and anticipate potential symptom management.

There is a tradeoff in using a simple proportion for the purposes of interpretability. From an analytical point of view, this single worst adverse event summary measure can fall short in reflecting the fluctuation of symptoms during treatment. This can be most evident when statistically discriminating tolerability between arms. Temporal profiles of symptomatic adverse event burden are often indistinguishable among acute, chronic, cumulative, cyclic, or late incipient treatment toxicity [5]. Incorporating these longitudinal profiles can be critical for fully characterizing tolerability as both isolated-severe and persistent-moderate symptoms have been shown to correlate with decrements in quality of life [6].

There are various graphical techniques and statistical strategies proposed in the literature that attempt to capture broader aspects of adverse event data beyond the maximum grade [7]. Graphical approaches enable a deep dive into longitudinal profiles via visual inspection within individual adverse event categories. A more holistic interpretation of tolerability can be accessible this way; however, it is not suitable for a succinct reporting of all adverse events. Although necessary, even the more complex statistical strategies can be constrained by the difficult or narrow interpretation of the results. A summary measure that aims to overcome these challenges is the toxicity index. The toxicity index was designed to incorporate the most severe grade and the frequency of all lower grade adverse events, resulting in a single summary measure for each patient [8]. Recent applications of the toxicity index using CTCAE data show potential gains in statistical power when using a probability index modeling approach [9, 10]. To date, the toxicity index has not been applied to PRO-CTCAE data. Given the high rate of symptoms reported by patients at baseline, methods for accounting for pre-existing symptoms are likely needed when applying the toxicity index to PRO-CTCAE data [11].

In this study, we present an application of the toxicity index to PRO-CTCAE tolerability data alongside a typical application of the maximum score. We apply a standard baseline adjustment approach to account for pre-existing symptoms to both summary measures side-by-side with the unadjusted results. Finally, we discuss considerations for reporting and interpreting the toxicity index as a summary measure when applied to PRO-CTCAE tolerability data.

Methods

Study data

The toxicity index summary measure was investigated using PRO-CTCAE data from the COMET-2 trial -- a phase 3, 1:1 randomized, double-blind, placebo-controlled trial with a primary pain endpoint comparing cabozantinib and mitoxantrone-prednisone among men with previously treated symptomatic castration-resistant prostate cancer. In one arm, cabozantinib was administered as the experimental treatment with mitoxantrone-matched placebo infusion, plus prednisone-matched placebo. In the other arm, mitoxantrone was

administered, plus prednisone and cabozantinib-matched placebo. Details on clinical findings and trial design are reported elsewhere [12].

PRO-CTCAE items were assessed at baseline, 1 and 2 months, and every two months thereafter over the study period. PRO-CTCAE items included constipation, decreased appetite, diarrhea, fatigue, insomnia, nausea, numbness or tingling in the hands or feet, pain, shortness of breath, and vomiting. Respective symptom item frequency, severity, and/or interference attributes were evaluated as specified per NCI PRO-CTCAE Item Library Version 1.0. PRO-CTCAE composite grades were computed from the individual items scores [13]. The PRO-CTCAE composite grades create a single grade per PRO-CTCAE symptom item group on a scale akin to other common adverse event tools like CTCAE or MedDRA.

Evaluating tolerability

The toxicity index is a summary measure aimed at ranking patients within a clinical trial by their respective adverse event experience over the trial. Those with more severe and frequent adverse events will have a higher toxicity index than those with less severe and infrequent adverse events. To construct the toxicity index for an individual patient, their observed adverse event grades over the study period are first ordered descending in severity, then Formula 1 is applied to these ordered data.

Formula 1.
Toxicity Index Statistic

$$\sum_{i \leq m} \frac{x_i}{\prod_{j<i}(1+x_j)}$$

Where m is the number of observed adverse events for a given patient, $x_{i=1}$ is the largest adverse event grade, $x_{i=2}$ is the second largest adverse event, and so on, up to the smallest adverse event grade, $x_{i=m}$. A detailed example of this calculation has been reported previously [8]. The resulting toxicity index statistic has two components: an integer and decimal portion. The integer portion is the patient's maximum adverse event grade. The decimal portion is considered the additional adverse event experience that this summary measure seeks to capture and allows for the patients to be ranked as such. For example, Patient A with observed PRO-CTCAE pain severity scores of 3, 3, 4, and 2 will have a toxicity index of 4.775, and Patient B with pain severity scores of 2, 3, and 4 will have a toxicity index of 4.700 (example calculation shown in Table 1). Thus, Patient A is ranked as having worse pain severity over the trial than Patient B. With possible PRO-CTCAE item scores of 0, 1, 2, 3, or 4, and composite grades of 0, 1, 2, or 3, the toxicity index statistic for PRO-CTCAE items and composite grades range from 0 to 4.999… and 0 to 3.999…, respectively. By design of the toxicity index, the accrual of additional adverse event experience will never result in the toxicity index increasing to the next whole unit score or grade above the maximum score or grade. This is convenient as interpreting a grade above the natural range of an adverse event measure (e.g., PRO-CTCAE, CTCAE, MedDRA) is not meaningful. As a result, PRO-CTCAE item and composite grade toxicity

index estimates reported here are not rounded to their respective upper bounds (e.g., a toxicity index estimate of 4.999 will be reported as 4.99 in lieu of rounding to 5.00). Those reporting the toxicity index may consider following this rounding exception to avoid interpretation-related confusion.

Adjusting for baseline symptoms

A variety of methods are available to account for pretreatment symptomatic burden measured at baseline using summary measures [14]. The typical approach is to compare the adverse event rates of only the most severe grade per patient after the baseline trial time point (post-baseline maximum). This approach incorporates all symptoms experienced during treatment, regardless of presumed pathology. Another method is to adjust this post-baseline maximum for pre-existing adverse events (baseline-adjusted maximum). Here emphasis is placed on adverse events worsening during treatment relative to the baseline trial time point and are deemed treatment emergent. The baseline-adjusted maximum is defined as the maximum grade post-baseline if there was at least a 1-grade increase in adverse event grade from baseline; otherwise the baseline-adjusted grade of 0 is given. With CTCAE grading, clinicians typically do not report an adverse event unless it is new or has worsened from baseline. Thus, this direct adjustment for patients' presenting symptomatic burden closely mimics how clinical adverse events are collected. This approach has previously been applied to these PRO-CTCAE data in the COMET-2 trial by Dueck and colleagues [15].

Along with the unadjusted toxicity index, novel post-baseline and baseline-adjusted versions of the toxicity index were evaluated. The post-baseline toxicity index only included observed PRO-CTCAE scores after the baseline trial time point. The baseline-adjusted toxicity index was defined as expressed in Formula 1 after including only those grades that were worse than the adverse event grade at baseline (similarly to the baseline-adjusted maximum). The baseline grade and subsequent grades that were not worse than the baseline value are excluded from the calculation of the baseline-adjusted toxicity index. An example of this calculation is shown in Table 2.

Statistical analysis

The proportions of patients with post-baseline and baseline-adjusted maximum PRO-CTCAE score 0 or higher, and 3 or higher, were compared between treatment arms using Fisher's exact tests. Nonparametric methods (Wilcoxon rank-sum tests) were utilized to compare toxicity indexes between arms, due to its multimodal distribution and inherent rank nature. For the same reason, the median was chosen to convey central tendency when summarizing at the arm level. The distributions of the decimal portion of the toxicity index were evaluated within each integer portion using Kolmogorov-Smirnov tests where appropriate. All presented p values are unadjusted for multiple testing and are provided for reference only. The intention is side-by-side presentation of the toxicity index summary measure applied in various fashions. Analyses were performed using the statistical software SAS version 9.4 (SAS Institute Inc., Cary, NC). PRO-CTCAE composite grades and longitudinal bar charts presented later and in Supplemental 1 were created using the statistical software R and the ProAE package [16].

Results

The COMET-2 trial enrolled a total of 119 male participants randomized to study treatment (cabozantinib n=61 or mitoxantrone-prednisone n=58). Of those enrolled, 107 completed a baseline PRO-CTCAE evaluation and at least one follow-up PRO-CTCAE evaluation (cabozantinib n=53 and mitoxantrone-prednisone n=54). Results here reflect these 107 participants. Among them, the number of PRO-CTCAE questionnaires completed ranged from 2 to 17 per participant, with 75% of participants completing 5 or more questionnaires. Figures showing all PRO-CTCAE individual item and composite grade distributions across trial time points can be found in Supplemental 1, as well as violin plots with overlaid density histograms displaying the toxicity index summary measure distributions. Demographic and disease-related characteristics for the COMET-2 trial are available elsewhere [12, 15].

PRO-CTCAE tolerability rates for present and severe adverse events (scores > 0 and scores ≥ 3, respectively) are reported by arm in Table 3. The mitoxantrone-prednisone arm showed generally favorable tolerability among individual PRO-CTCAE items compared to cabozantinib in both present and severe PRO adverse event rates. Tolerability rates among PRO-CTCAE composite grades were significantly higher (worse) in the cabozantinib arm for decreased appetite, diarrhea, nausea, and vomiting, for both post-baseline and baseline-adjusted rates.

Differences in toxicity index between treatment arms shown in Table 4 were directionally consistent with tolerability rate comparisons, as expected. Among significantly different toxicity index distributions across PRO-CTCAE item groups, higher median toxicity indexes were observed in the cabozantinib arm for constipation, decreased apatite, diarrhea, numbness or tingling in hands or feet, and vomiting. Similar to rate comparisons, the cabozantinib arm had higher median toxicity index among composite grades for appetite, diarrhea, nausea, and vomiting, in both post-baseline and baseline-adjusted versions.

The median toxicity index was substantially reduced between post-baseline and baseline-adjustment methods within the constipation, fatigue, insomnia, and pain PRO-CTCAE item groups. For example, the post-baseline median pain severities were 3.50 and 3.67 for cabozantinib and mitoxantrone-prednisone, respectively. This indicates at least 50% of all participants reported multiple pain episodes with at least one being severe after the baseline visit. However, the baseline-adjusted medians for pain severity are each 0, indicating that 50% or more of participants did not experience treatment emergent pain. Unsurprisingly, this differing impact of baseline adjustment methods is equivalently observed in Table 3 using the dichotomous tolerability rates (i.e., scores > 0 and scores ≥ 3). The post-baseline tolerability rates for pain severity with maximum score 3 or higher were 60% and 67% for cabozantinib and mitoxantrone-prednisone, respectively. Again, this is consistent with the toxicity index result as roughly 50% or more reported at least one severe pain episode after baseline. Among the baseline-adjusted rates, the proportion of participants with maximum score greater than 0 were 19% and 30% for cabozantinib and mitoxantrone-prednisone, respectively (each arm below 50% incidence of treatment emergent pain). Reading the COMET-2 participants' symptomatic pain in this way shows that similar information in percentile description can be gathered from the toxicity index and PRO adverse event

rate and are equivalently impacted by the baseline adjustment methods. Figure 1 shows the longitudinal profiles of pain frequency, severity, and interference scores, as well as composite grade during the trial. Figure 2 shows the distributions of the toxicity index summary measure.

To further evaluate the characteristics of the toxicity index, the distribution of the decimal portion among unadjusted toxicity index estimates was assessed within each integer portion (Figure 3). The histograms in Figure 3 incorporate 3210 toxicity index estimates (one estimate for each of 30 PRO-CTCAE items and composite grades among the 107 participants). This graphically demonstrates how the toxicity index accumulates the additional toxicity (decimal portion) at differing rates within each maximum score (integer portion). Specifically, the set of possible ranks varies within each integer portion. Since interpreting arm medians with differing integer portions may be precarious as the decimal portions are scaled differently, comparisons of decimal portion distributions between treatment arms were carried out individually within integer groups. In the COMET-2 trial, the only statistically significant differences seen in decimal portion between treatment arms were that of decreased appetite within the maximum score groups of 3 (for severity, interference, and composite) and 4 (for interference).

Discussion

In this study, the toxicity index was applied to PRO-CTCAE data for the first time with adjustment for each patient's pre-existing symptoms and evaluated as a tolerability outcome in univariate analyses. Broad agreement was observed between the toxicity index and more typical summary measures like maximum score. However, the median and range of the toxicity index reported by arm were often challenging to interpret. We see this when comparing values between Table 3 to Table 4. Admittedly, care must be taken when reporting typical group estimates of the toxicity index directly (mean, median, etc.). Since the distribution of the decimal portion varies within each integer portion, a representation of the decimal portion such as the median toxicity index is not necessarily interpreted equivalently across integer portions. It remains unclear what group-level summary estimates of the toxicity index are most interpretable.

Capturing the longitudinal toxicity experience remains an emerging area of methodological research in treatment tolerability analysis. The toxicity index introduced by Rogatko and colleagues in 2004 is an innovative summary measure accounting for both the multiplicity and severity of adverse events. Rogatko demonstrated that it has useful potential in early-phase clinical trials by creating more sensitive dose limiting toxicity thresholds [8]. Some purposive methods have since been demonstrated to highlight amenable approaches accommodating the rank and multimodal nature of the toxicity index. For example, using CTCAE grades, Gresham et al showed that the toxicity index has increased power when using a probability index modeling approach [9] and Razaee et al present a novel framework showing increased power when using their derived method testing for a difference in mean Poisson-limit vector parameters between treatment arms [10]. Each of these methodologies are valued additions to the adverse event literature and may be considered when statistically discriminating treatment arms is of paramount concern. However, specifically for the

purposes of communicating comparative tolerability with patients and reporting to a wider scientific audience, we feel the trade-off between interpretability and statistical complexity for the sake of increased power is substantial. Simpler dichotomizations or categorizations of tolerability may have more practical communicative utility than the toxicity index.

It is evident that this ranking measure is statistically convenient where patients with similar adverse event profiles can be precisely ordered by rank. The toxicity index appears most useful for statistical comparisons between treatment arms or between subgroups where interpreting tolerability is of lesser importance relative to statistical power. Though, additional work is needed for more comprehensive applications of the toxicity index and to assess its ability to support clinical decision making. Direct interpretation and associated effect size recommendations need to be outlined. Considerations should also be defined for applying the toxicity index to serial versus episodic adverse event evaluations in the clinical trial setting. For example, PRO-CTCAE evaluations are more likely to record non-zero scores at scheduled visits, while CTCAE evaluations (when captured in a log-style format) typically record a single toxicity grade until the adverse event worsens or reoccurs after resolution. Approaches for handling missing data should also be investigated. A patient with any missing non-zero adverse event scores will have a lower toxicity index than if that data were observed. This implies that the existence of any missing symptomatic adverse event data will result in an underestimation of toxicity index. Simulations of the toxicity index's decimal portion accrual may illuminate these interests in repeated PRO-CTCAE evaluations, interpretable effect sizes, and missing data impacts.

This evaluation of the toxicity index has some limitations, several stemming from the characteristics which make PRO-CTCAE unique from CTCAE. For example, the maximum score summary measure is computed from a single observation, whereas a summary measure like the toxicity index is computed from a series of observations. This raises computational questions when applying the baseline adjustment approach and whether it should be applied at the summary measure level or applied to raw data prior to the calculation of the summary measure. Adding to this, patients on study for longer periods or having more frequent serial PRO-CTCAE evaluations will inherently have more opportunity to accrue toxicity index (e.g., weekly versus monthly evaluations per annum). This coupled with the self-reported nature of PRO data, trial participants with more frequent visits and better adherence to fully completing PRO-CTCAE questionnaires may be biased towards a higher toxicity index; specifically, within the decimal portion of the statistic where toxicity is accrued. The impact of missing data on the toxicity index was also not evaluated here. As referred to previously, PRO-CTCAE and CTCAE evaluations observe the absence of symptomatic adverse events differently. This inconsistency in how the respective tools collect data also extends to the means by which missing data are generated. We believe these potential impacts do not jeopardize this study's evaluation of treatment tolerability using toxicity index and are not yet addressed in the literature.

This study showed that it is possible to apply the toxicity index to PRO-CTCAE data, incorporate in tabular reporting of symptomatic adverse events in cancer clinical trials and be adjusted for patients' pre-existing symptoms. An appropriate interpretation of the toxicity index as well as foundational work to understand impacts of PRO assessment schedules

and missing data are needed for a broader implementation of the toxicity index with PRO adverse event data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Disclosures, Acknowledgments, and funding

This work was supported through the Cancer Moonshot initiative of the US National Cancer Institute (grant U01 CA233046). Data were provided by Exelixis Inc for this trial, but Exelixis Inc did not provide any funds and did not participate in the decision to submit this manuscript, analysis, or approval of the submission. The authors declare no conflicts of interest.

References

1. Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst. 2009 Dec 2;101(23):1624–32. doi: 10.1093/jnci/djp386. Epub 2009 Nov 17. PMID: 19920223; PMCID: PMC2786917. [PubMed: 19920223]
2. Justice AC, Rabeneck L, Hays RD, Wu AW, Bozzette SA. Sensitivity, specificity, reliability, and clinical validity of provider-reported symptoms: a comparison with self-reported symptoms. Outcomes Committee of the AIDS Clinical Trials Group. J Acquir Immune Defic Syndr. 1999 Jun 1;21(2):126–33. PMID: 10360804. [PubMed: 10360804]
3. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring (MD): Food and Drug Administration (US); 2016-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK326791/Co-published by National Institutes of Health (US), Bethesda (MD).
4. Friends of Cancer Research. Broadening the Definition of Tolerability in Cancer Clinical Trials to Better Measure the Patient Experience. Website: https://www.focr.org/publications/broadening-definition-tolerability-cancer-clinical-trials-better-measure-patient; 2018.
5. Thanarajasingam G, Minasian LM, Baron F, et al. Beyond maximum grade: modernising the assessment and reporting of adverse events in haematological malignancies. Lancet Haematol. 2018 Nov;5(11):e563–e598. doi: 10.1016/S2352-3026(18)30051-6. Epub 2018 Jun 18. Erratum in: Lancet Haematol. 2019 Mar;6(3):e121. PMID: 29907552; PMCID: PMC6261436. [PubMed: 29907552]
6. Langlais BT, Geyer H, Scherber R, Mesa RA, Dueck AC. Quality of life and symptom burden among myeloproliferative neoplasm patients: do symptoms impact quality of life? Leuk Lymphoma. 2019 Feb;60(2):402–408. doi: 10.1080/10428194.2018.1480768. Epub 2018 Jul 22. PMID: 30033837; PMCID: PMC6363896. [PubMed: 30033837]
7. Thanarajasingam G, Atherton PJ, Novotny PJ, et al. Longitudinal adverse event assessment in oncology clinical trials: the Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254. Lancet Oncol. 2016 May;17(5):663–70. doi: 10.1016/S1470-2045(16)00038-3. Epub 2016 Apr 12. PMID: 27083333; PMCID: PMC4910515. [PubMed: 27083333]
8. Rogatko A, Babb JS, Wang H, Slifker MJ, Hudes GR. Patient characteristics compete with dose as predictors of acute treatment toxicity in early phase clinical trials. Clin Cancer Res. 2004 Jul 15;10(14):4645–51. doi: 10.1158/1078-0432.CCR-03-0535. PMID: 15269136. [PubMed: 15269136]
9. Gresham G, Diniz MA, Razaee ZS, et al. Evaluating Treatment Tolerability in Cancer Clinical Trials Using the Toxicity Index. J Natl Cancer Inst. 2020 Dec 14;112(12):1266–1274. doi: 10.1093/jnci/djaa028. PMID: 32091598; PMCID: PMC7735773. [PubMed: 32091598]
10. Razaee ZS, Amini AA, Diniz MA, et al. On the properties of the toxicity index and its statistical efficiency. Stat Med. 2021 Mar 15;40(6):1535–1552. doi: 10.1002/sim.8858. Epub 2020 Dec 20. PMID: 33345351; PMCID: PMC7953898. [PubMed: 33345351]

11. Atkinson TM, Dueck AC, Satele DV, et al. Clinician vs Patient Reporting of Baseline and Postbaseline Symptoms for Adverse Event Assessment in Cancer Clinical Trials. JAMA Oncol. 2020 Mar 1;6(3):437–439. doi: 10.1001/jamaoncol.2019.5566. PMID: 31876902; PMCID: PMC6990818. [PubMed: 31876902]

12. Basch EM, Scholz M, de Bono JS, et al. Cabozantinib Versus Mitoxantrone-prednisone in Symptomatic Metastatic Castration-resistant Prostate Cancer: A Randomized Phase 3 Trial with a Primary Pain Endpoint. Eur Urol. 2019 Jun;75(6):929–937. doi: 10.1016/j.eururo.2018.11.033. Epub 2018 Dec 4. PMID: 30528222; PMCID: PMC6876845. [PubMed: 30528222]

13. Basch E, Becker C, Rogak LJ, et al. Composite grading algorithm for the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Clin Trials. 2021 Feb;18(1):104–114. doi: 10.1177/1740774520975120. Epub 2020 Dec 1. PMID: 33258687; PMCID: PMC7878323. [PubMed: 33258687]

14. Basch E, Rogak LJ, Dueck AC. Methods for Implementing and Reporting Patient-reported Outcome (PRO) Measures of Symptomatic Adverse Events in Cancer Clinical Trials. Clin Ther. 2016 Apr;38(4):821–30. doi: 10.1016/j.clinthera.2016.03.011. Epub 2016 Apr 2. PMID: 27045992; PMCID: PMC4851916. [PubMed: 27045992]

15. Dueck AC, Scher HI, Bennett AV, et al. Assessment of Adverse Events From the Patient Perspective in a Phase 3 Metastatic Castration-Resistant Prostate Cancer Clinical Trial. JAMA Oncol. 2020 Feb 1;6(2):e193332. doi: 10.1001/jamaoncol.2019.3332. Epub 2020 Feb 13. PMID: 31556911; PMCID: PMC6764147. [PubMed: 31556911]

16. Langlais B, Dueck A. ProAE: Tools for PRO-CTCAE data management, analysis, and graphical display. R package version 0.1.1. https://CRAN.R-project.org/package=ProAE. 2021.

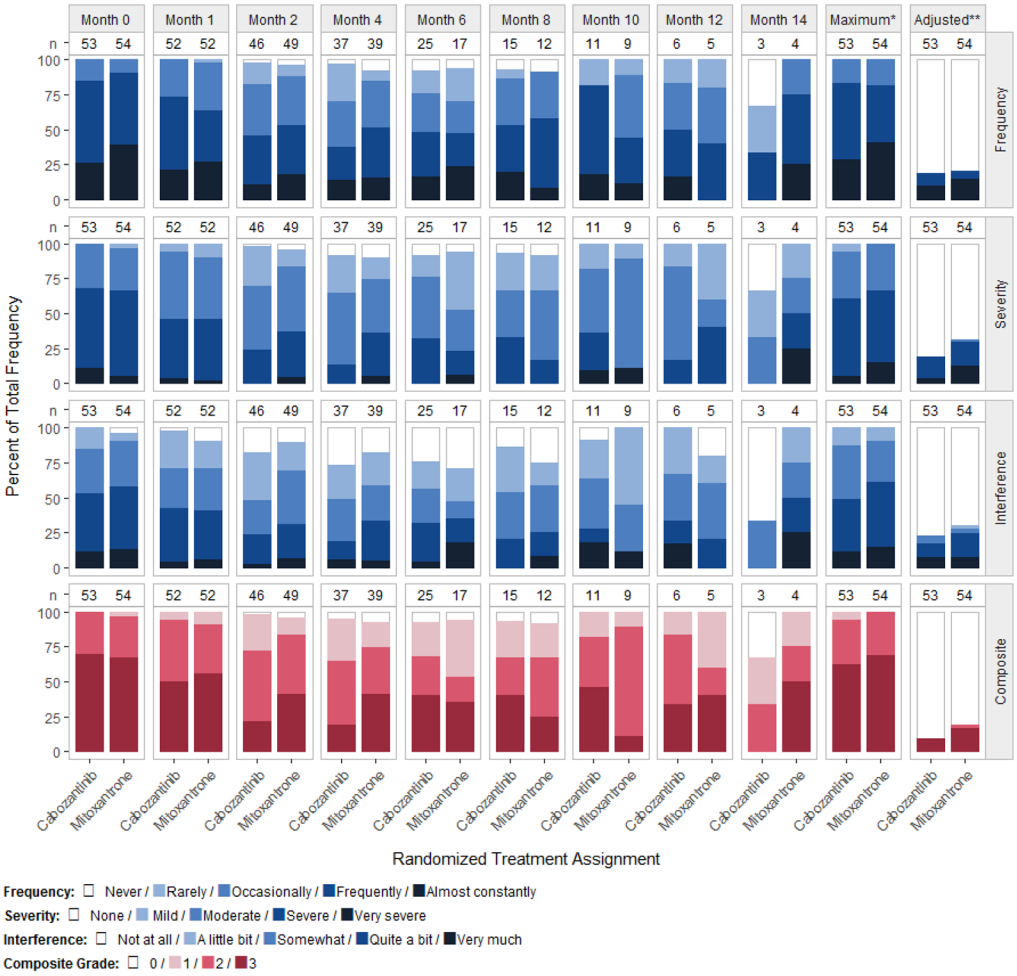

Figure 1.

Distribution of the PRO-CTCAE Pain item group at successive time points during the COMET-2 trial and maximum score post-baseline without and with baseline adjustment

Column labels (n) show the number of subjects with an observed symptom score or grade.

*Maximum score or grade reported post-baseline per patient.

**Maximum score or grade reported post-baseline per patient when including only scores that were worse than the patient's baseline score.

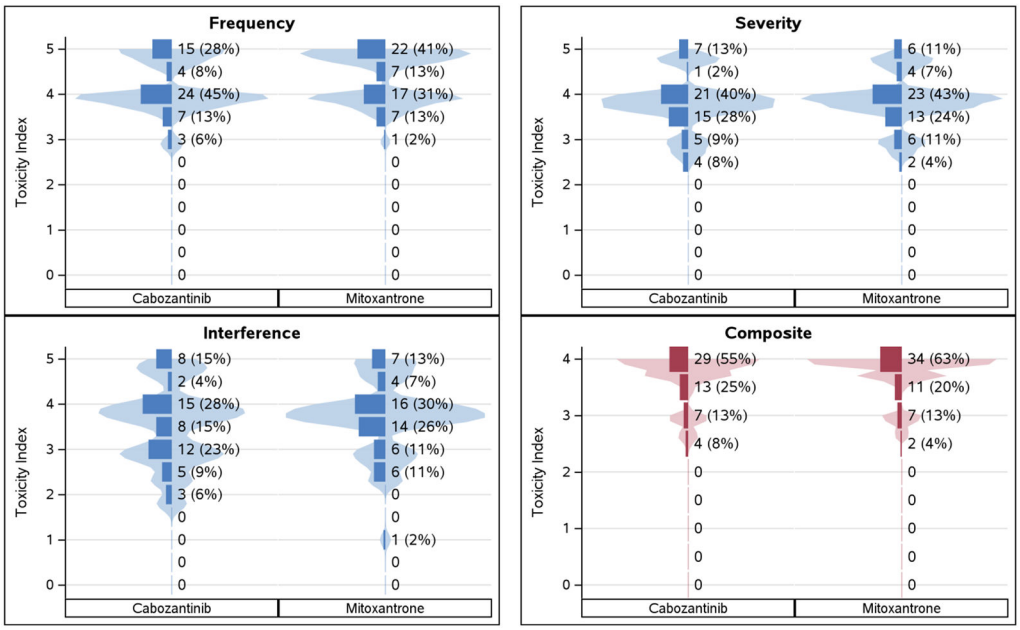

Figure 2.

Violin plots with overlaid density histograms of the unadjusted toxicity index summary measure distribution for the PRO-CTCAE Pain item group

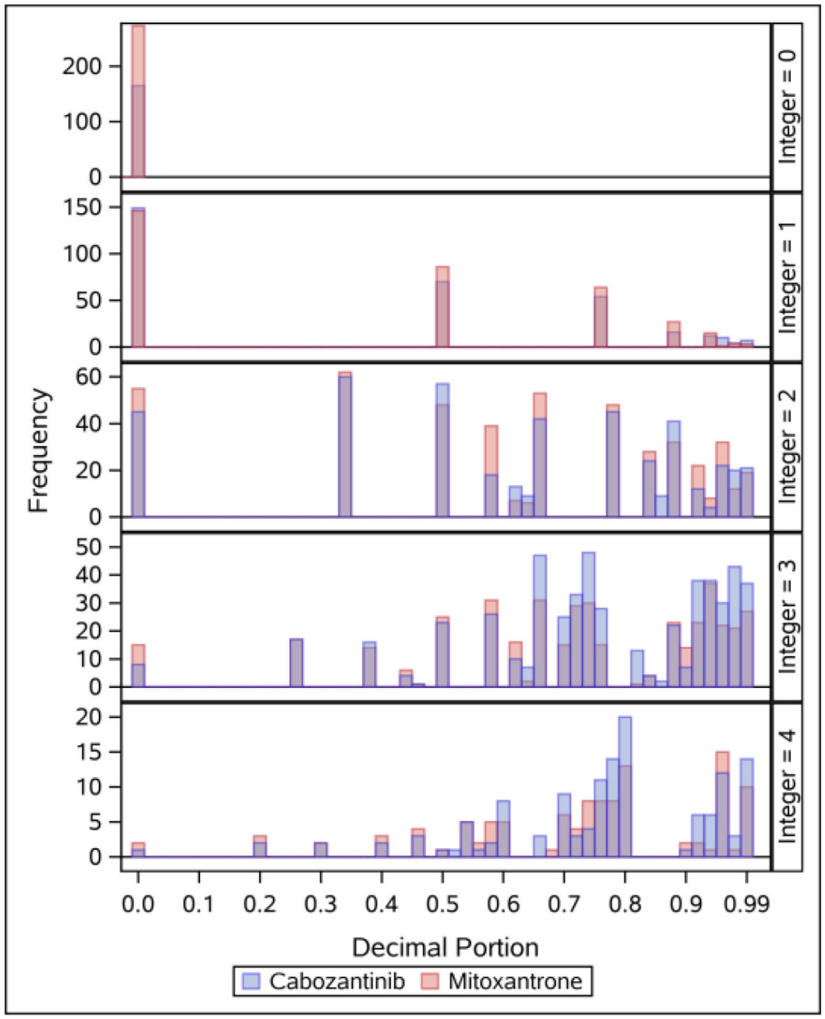

Figure 3.

Histograms of the toxicity index decimal portion by integer portion, across all patients and PRO-CTCAE items and composites.

Histograms show all 3210 unadjusted toxicity estimates – one for each of the 30 PRO-CTCAE items and composites across the 107 patients. The y-axis shows the frequency for each toxicity integer group (maximum score; 0, 1, 2, 3, or 4). The x-axis shows the decimal portion of the toxicity index (accumulated toxicity in addition to the maximum score)

Table 1.

Example calculation of the toxicity index

	Patient A		Patient B	
	Score	**Accrued Toxicity**	**Score**	**Accrued Toxicity**
	4 (cycle 2)	4.000	4 (cycle 2)	4.000
	3 (cycle 1)	+ 0.600	3 (cycle 1)	+ 0.600
	3 (baseline)	+ 0.150	2 (baseline)	+ 0.100
	2 (cycle 3)	+ 0.025	-	-
Toxicity Index:		**4.775**		**4.700**

Adverse event scores are show here ordered descending in severity to further illustrate the calculation, with each associated time point in parentheses; i.e., score (time point).

Table 2.

Example of baseline-adjusted toxicity index procedure (1) and subsequent calculation (2)

	(1) PRO adverse event scores overtime for Patient A and B			
	Patient A		Patient B	
Time point	Score	Worse than baseline?	Score	Worse than baseline?
Baseline	**3**	-	**2**	-
Cycle 1	3	No	3	Yes
Cycle 2	4	Yes	4	Yes
Cycle 3	2	No	-	-
	(2) Baseline-adjusted toxicity index (only include scores worse than baseline)			
	Patient A		Patient B	
	Score	Accrued Toxicity	Score	Accrued Toxicity
	4 (cycle 2)	4.000	4 (cycle 2)	4.000
	3 (cycle 1)	-	3 (cycle 1)	+ 0.600
	3 (baseline)	-	2 (baseline)	-
	2 (cycle 3)	-	-	-
Toxicity Index:		**4.000**		**4.600**

PRO: Patient reported outcome. Within sub-table (2), adverse event scores are show here ordered descending in severity to further illustrate the calculation, with each associated time point in parentheses; i.e., score (time point).

Table 3.

Rates of PRO-CTCAE item scores and composite grades greater than 0 and 3 or higher, by treatment arm

PRO-CTCAE Item Group	Cabozantinib (Cabo) N	Mitoxantrone-prednisone (Mito) N	Post-Baseline Maximum Score						Baseline-Adjusted Maximum Score					
			Score > 0			Score ≥ 3			Score > 0			Score ≥ 3		
			Cabo n (%)	Miton (%)	p	Cabo n (%)	Miton (%)	p	Cabo n (%)	Miton (%)	p	Cabo n (%)	Miton (%)	p
Constipation														
Severity	53	54	50 (94)	47 (87)	0.32	19 (36)	11 (20)	0.09	25 (47)	16 (30)	0.08	14 (26)	7 (13)	0.09
Composite	53	54	50 (94)	47 (87)	0.32	19 (36)	11 (20)	0.09	24 (45)	16 (30)	0.11	13 (25)	7 (13)	0.14
Decreased Appetite														
Severity	52	54	50 (96)	48 (89)	0.27	27 (52)	10 (19)	**0.0005**	34 (65)	18 (33)	**0.0017**	20 (38)	8 (15)	**0.0079**
Interference	52	54	48 (92)	39 (72)	**0.0103**	23 (44)	11 (20)	**0.0122**	34 (65)	19 (35)	**0.0034**	18 (35)	9 (17)	**0.0449**
Composite	52	54	50 (96)	48 (89)	0.27	18 (35)	7 (13)	**0.0116**	32 (62)	17 (31)	**0.0033**	14 (27)	5 (9)	**0.0228**
Diarrhea														
Frequency	52	54	48 (92)	34 (63)	**0.0004**	23 (44)	6 (11)	**0.0002**	42 (81)	26 (48)	**0.0006**	23 (44)	6 (11)	**0.0002**
Composite	52	54	48 (92)	34 (63)	**0.0004**	9 (17)	1 (2)	**0.0075**	38 (73)	21 (39)	**0.0005**	9 (17)	1 (2)	**0.0075**
Fatigue														
Severity	53	54	53 (100)	54 (100)	>.99	39 (74)	32 (59)	0.15	24 (45)	17 (31)	0.17	19 (36)	14 (26)	0.30
Interference	53	54	53 (100)	54 (100)	>.99	40 (75)	35 (65)	0.29	27 (51)	19 (35)	0.12	23 (43)	17 (31)	0.23
Composite	53	54	53 (100)	54 (100)	>.99	34 (64)	28 (52)	0.24	18 (34)	10 (19)	0.08	12 (23)	8 (15)	0.33
Insomnia														
Severity	53	54	44 (83)	47 (87)	0.60	10 (19)	8 (15)	0.61	20 (38)	22 (41)	0.84	7 (13)	7 (13)	>.99
Interference	53	54	36 (68)	41 (76)	0.40	10 (19)	10 (19)	>.99	15 (28)	22 (41)	0.22	5 (9)	7 (13)	0.76
Composite	53	54	44 (83)	47 (87)	0.60	6 (11)	5 (9)	0.76	14 (26)	23 (43)	0.10	3 (6)	4 (7)	>.99
Nausea														
Frequency	52	54	49 (94)	38 (70)	**0.0018**	25 (48)	11 (20)	**0.0039**	35 (67)	20 (37)	**0.0021**	23 (44)	10 (19)	**0.0061**
Severity	52	54	49 (94)	36 (67)	**0.0004**	21 (40)	9 (17)	**0.0094**	39 (75)	20 (37)	**0.0001**	20 (38)	8 (15)	**0.0079**
Composite	52	54	49 (94)	38 (70)	**0.0018**	18 (35)	5 (9)	**0.0020**	35 (67)	18 (33)	**0.0009**	17 (33)	5 (9)	**0.0038**
Numbness/ Tingling in Hands/Feet														
Severity	52	54	44 (85)	40 (74)	0.23	16 (31)	7 (13)	**0.0341**	28 (54)	18 (33)	**0.0495**	12 (23)	4 (7)	**0.0307**

PRO-CTCAE Item Group	Cabozantinib (Cabo) N	Mitoxantrone-prednisone (Mito) N	Post-Baseline Maximum Score						Baseline-Adjusted Maximum Score					
			Score > 0			Score ≥ 3			Score > 0			Score ≥ 3		
			Cabon (%)	Miton (%)	p	Cabon (%)	Miton (%)	p	Cabon (%)	Miton (%)	p	Cabon (%)	Miton (%)	p
Interference	52	54	34 (65)	26 (48)	0.08	11 (21)	5 (9)	0.11	24 (46)	16 (30)	0.11	7 (13)	4 (7)	0.35
Composite	52	54	44 (85)	40 (74)	0.23	9 (17)	4 (7)	0.1463	22 (42)	19 (35)	0.55	6 (12)	3 (6)	0.32
Pain														
Frequency	53	54	53 (100)	54 (100)	>.99	44 (83)	44 (81)	>.99	10 (19)	11 (20)	>.99	10 (19)	11 (20)	>.99
Severity	53	54	53 (100)	54 (100)	>.99	32 (60)	36 (67)	0.55	10 (19)	17 (31)	0.18	10 (19)	16 (30)	0.26
Interference	53	54	53 (100)	54 (100)	>.99	26 (49)	33 (61)	0.25	12 (23)	16 (30)	0.51	9 (17)	13 (24)	0.47
Composite	53	54	53 (100)	54 (100)	>.99	33 (62)	37 (69)	0.55	5 (9)	10 (19)	0.27	5 (9)	9 (17)	0.39
Shortness of Breath														
Severity	50	54	40 (80)	43 (80)	>.99	8 (16)	9 (17)	>.99	29 (58)	21 (39)	0.08	7 (14)	7 (13)	>.99
Interference	50	54	36 (72)	37 (69)	0.83	12 (24)	12 (22)	>.99	29 (58)	20 (37)	**0.0488**	10 (20)	9 (17)	0.80
Composite	50	54	40 (80)	43 (80)	>.99	7 (14)	7 (13)	>.99	28 (56)	20 (37)	0.08	7 (14)	6 (11)	0.77
Vomiting														
Frequency	52	54	40 (77)	26 (48)	**0.0027**	6 (12)	4 (7)	0.52	33 (63)	18 (33)	**0.0033**	6 (12)	4 (7)	0.52
Severity	52	54	37 (71)	20 (37)	**0.0005**	11 (21)	4 (7)	0.05	33 (63)	12 (22)	**<.0001**	11 (21)	4 (7)	0.05
Composite	52	54	40 (77)	26 (48)	**0.0027**	4 (8)	2 (4)	0.43	33 (63)	18 (33)	**0.0033**	4 (8)	2 (4)	0.43

PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events. P values reflect Fisher's exact tests comparing frequencies between treatment arms. P values less than 0.05 are bolded.

Table 4.

Toxicity Index for PRO-CTCAE item scores and composite grades, by treatment arms

	Toxicity Index					Toxicity Index - Post Baseline					Toxicity Index - Baseline Adjusted				
PRO-CTCAE Item Group	**n**	**Cabozantinib median (range)**	**n**	**Mitoxantrone median (range)**	**P Value**	**n**	**Cabozantinib median (range)**	**n**	**Mitoxantrone median (range)**	**P Value**	**n**	**Cabozantinib median (range)**	**n**	**Mitoxantrone median (range)**	
Constipation															
Severity	53	3.25 (0.00-4.95)	54	2.67 (0.00-4.96)	**0.0421**	53	2.58 (0.00-4.95)	54	2.00 (0.00-4.79)	0.08	53	0.00 (0.00-4.92)	54	0.00 (0.00-4.00)	
Composite	53	3.25 (0.00-3.98)	54	2.67 (0.00-3.99)	**0.0371**	53	2.58 (0.00-3.98)	54	2.00 (0.00-3.98)	0.08	53	0.00 (0.00-3.94)	54	0.00 (0.00-3.94)	
Decreased Appetite															
Severity	53	3.67 (0.00-4.99)	54	2.78 (0.00-4.73)	**0.0006**	52	3.13 (0.00-4.98)	54	2.33 (0.00-4.70)	**<.0001**	52	2.67 (0.00-4.95)	54	0.00 (0.00-4.60)	
Interference	53	2.98 (0.00-4.80)	54	2.17 (0.00-4.58)	**0.0045**	52	2.91 (0.00-4.79)	54	1.50 (0.00-4.53)	**<.0001**	52	2.00 (0.00-4.79)	54	0.00 (0.00-4.00)	
Composite	53	2.83 (0.00-3.99)	54	2.33 (0.00-3.92)	**0.0043**	52	2.72 (0.00-3.99)	54	1.75 (0.00-3.67)	**<.0001**	52	2.00 (0.00-3.99)	54	0.00 (0.00-3.50)	
Diarrhea															
Frequency	53	2.93 (0.00-4.96)	54	1.50 (0.00-4.56)	**<.0001**	52	2.63 (0.00-4.96)	54	1.00 (0.00-4.47)	**<.0001**	52	2.54 (0.00-4.95)	54	0.00 (0.00-4.00)	
Composite	53	1.88 (0.00-3.94)	54	1.50 (0.00-3.44)	**<.0001**	52	1.81 (0.00-3.93)	54	1.00 (0.00-3.38)	**<.0001**	52	1.63 (0.00-3.91)	54	0.00 (0.00-3.00)	
Fatigue															
Severity	53	3.93 (2.65-4.99)	54	3.81 (2.33-4.96)	0.13	53	3.88 (2.00-4.99)	54	3.50 (1.00-4.95)	0.05	53	0.00 (0.00-4.96)	54	0.00 (0.00-4.80)	
Interference	53	3.94 (1.75-4.99)	54	3.88 (1.00-4.99)	0.68	53	3.88 (1.50-4.99)	54	3.58 (1.00-4.99)	0.16	53	1.94 (0.00-5.00)	54	0.00 (0.00-4.96)	
Composite	53	3.75 (1.97-3.99)	54	3.68 (1.50-3.99)	0.30	53	3.63 (1.75-3.99)	54	3.00 (1.00-3.99)	0.15	53	0.00 (0.00-3.94)	54	0.00 (0.00-3.75)	
Insomnia															
Severity	53	2.66 (0.00-4.99)	54	2.67 (0.00-3.98)	0.91	53	2.00 (0.00-4.96)	54	2.33 (0.00-3.94)	0.47	53	0.00 (0.00-3.67)	54	0.00 (0.00-3.75)	
Interference	53	2.33 (0.00-4.80)	54	2.42 (0.00-4.73)	0.74	53	1.50 (0.00-4.80)	54	2.00 (0.00-4.53)	0.40	53	0.00 (0.00-4.75)	54	0.00 (0.00-4.00)	
Composite	53	2.50 (0.00-3.99)	54	2.33 (0.00-3.94)	0.82	53	1.75 (0.00-3.99)	54	2.00 (0.00-3.92)	0.54	53	0.00 (0.00-3.67)	54	0.00 (0.00-3.75)	
Nausea															
Frequency	53	3.50 (0.00-4.95)	54	2.63 (0.00-4.96)	**0.0264**	52	2.96 (0.00-4.95)	54	2.00 (0.00-4.96)	**0.0008**	52	2.72 (0.00-4.95)	54	0.00 (0.00-4.80)	
Severity	53	2.88 (0.00-4.70)	54	2.54 (0.00-4.47)	**0.0121**	52	2.81 (0.00-4.70)	54	1.88 (0.00-4.40)	**0.0002**	52	2.67 (0.00-4.70)	54	0.00 (0.00-4.40)	
Composite	53	2.78 (0.00-3.93)	54	2.33 (0.00-3.98)	**0.0108**	52	2.67 (0.00-3.92)	54	1.63 (0.00-3.98)	**0.0004**	52	2.17 (0.00-3.88)	54	0.00 (0.00-3.94)	
Numbness/Tingling in Hands/Feet															
Severity	53	2.50 (0.00-4.95)	54	2.33 (0.00-4.96)	0.06	52	2.33 (0.00-4.92)	54	1.75 (0.00-4.79)	**0.0367**	52	1.00 (0.00-4.80)	54	0.00 (0.00-3.92)	
Interference	53	1.75 (0.00-4.99)	54	1.00 (0.00-4.78)	**0.0443**	52	1.50 (0.00-4.99)	54	0.00 (0.00-4.78)	**0.0280**	52	0.00 (0.00-3.81)	54	0.00 (0.00-4.75)	

	Toxicity Index					Toxicity Index - Post Baseline					Toxicity Index - Baseline Adjusted				
PRO-CTCAE Item Group	**n**	**Cabozantinib median (range)**	**n**	**Mitoxantrone median (range)**	**P Value**	**n**	**Cabozantinib median (range)**	**n**	**Mitoxantrone median (range)**	**P Value**	**n**	**Cabozantinib median (range)**	**n**	**Mitoxantrone median (range)**	
Composite	53	1.88 (0.00-3.99)	54	1.75 (0.00-3.98)	0.14	52	1.75 (0.00-3.99)	54	1.50 (0.00-3.97)	0.06	52	0.00 (0.00-3.98)	54	0.00 (0.00-3.94)	
Pain															
Frequency	53	3.97 (2.78-4.99)	54	4.59 (2.96-4.99)	0.18	53	3.88 (2.00-4.99)	54	3.88 (2.00-4.99)	0.47	53	0.00 (0.00-4.00)	54	0.00 (0.00-4.96)	
Severity	53	3.75 (2.50-4.91)	54	3.90 (2.67-4.95)	0.56	53	3.50 (1.50-4.72)	54	3.67 (2.00-4.80)	0.34	53	0.00 (0.00-4.00)	54	0.00 (0.00-4.00)	
Interference	53	3.71 (1.75-4.99)	54	3.75 (1.00-4.99)	0.38	53	2.99 (1.00-4.99)	54	3.50 (1.00-4.99)	0.40	53	0.00 (0.00-4.99)	54	0.00 (0.00-4.96)	
Composite	53	3.75 (2.50-3.99)	54	3.89 (2.67-3.99)	0.58	53	3.50 (1.50-3.99)	54	3.69 (2.00-3.99)	0.46	53	0.00 (0.00-3.75)	54	0.00 (0.00-3.94)	
Shortness of Breath															
Severity	53	2.00 (0.00-4.95)	54	2.54 (0.00-4.75)	0.46	50	2.00 (0.00-4.93)	54	2.00 (0.00-4.00)	0.95	50	1.00 (0.00-4.92)	54	0.00 (0.00-4.00)	
Interference	53	1.75 (0.00-4.95)	54	2.17 (0.00-4.80)	0.57	50	1.91 (0.00-4.95)	54	1.97 (0.00-4.69)	0.93	50	1.00 (0.00-4.95)	54	0.00 (0.00-4.60)	
Composite	53	1.75 (0.00-3.93)	54	2.16 (0.00-3.91)	0.33	50	1.81 (0.00-3.93)	54	1.92 (0.00-3.87)	0.80	50	1.00 (0.00-3.92)	54	0.00 (0.00-3.75)	
Vomiting															
Frequency	53	1.50 (0.00-4.20)	54	1.00 (0.00-3.90)	**0.0443**	52	1.00 (0.00-4.00)	54	0.00 (0.00-3.81)	**0.0039**	52	1.00 (0.00-4.00)	54	0.00 (0.00-3.75)	
Severity	53	1.00 (0.00-4.00)	54	0.50 (0.00-4.72)	**0.0372**	52	1.00 (0.00-4.00)	54	0.00 (0.00-4.65)	**0.0009**	52	1.00 (0.00-4.00)	54	0.00 (0.00-4.60)	
Composite	53	1.50 (0.00-3.91)	54	1.00 (0.00-3.90)	**0.0278**	52	1.00 (0.00-3.91)	54	0.00 (0.00-3.81)	**0.0018**	52	1.00 (0.00-3.91)	54	0.00 (0.00-3.75)	

PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events. P values reflect Wilcoxon rank-sum tests comparing ranks between treatment arms. P values less than 0.05 are bolded.