

Cultural adaptation of the Italian version of the Patient-Reported Outcomes Common Terminology Criteria for Adverse Event (PRO-CTCAE®)

Tumori Journal

1–11

© Fondazione IRCCS Istituto Nazionale dei Tumori 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03008916221099558

journals.sagepub.com/home/tmj



Caterina Caminiti¹ , Jane Bryce², Silvia Riva³, Diane Ng⁴,
Francesca Diodati¹, Elisa Iezzi¹, Lucia Sparavigna⁵, Silvia Novello⁶,
Camillo Porta^{7,8} , Lucia Del Mastro⁹, Giuseppe Procopio¹⁰ ,
Saverio Cinieri¹¹, Amalia Falzetta¹², Fabio Calabrò¹³, Vito Lorusso¹⁴,
Alessio Aligi Cogoni¹⁵, Giampaolo Tortora¹⁶, Marco Maruzzo¹⁷,
Rodolfo Passalacqua¹⁸, Francesco Cognetti¹⁹, Vincenzo Adamo²⁰,
Enrica Capelletto⁶, Alessandra Ferrari⁷, Michela Bagnalasta²¹,
Maurizio Bassi²¹, Annalisa Nicelli²¹, Davide De Persis²²,
Alessia D'Acunti²³, Elisabetta Iannelli (patient)²²,
Francesco Perrone^{±,5}  and Sandra A. Mitchell^{±24}

Abstract

Introduction: US National Cancer Institute's (NCI) Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE®) is a library of 78 symptom terms and 124 items enabling patient reporting of symptomatic adverse events in cancer trials. This multicenter study used mixed methods to develop an Italian language version of this widely accepted measure, and describe the content validity and reliability in a diverse sample of Italian-speaking patients.

¹Research and Innovation Unit, University Hospital of Parma, Parma, Italy

²Ascension St. John Clinical Research Institute, Tulsa, Oklahoma, USA

³Department of Psychology, St Mary's University, London, UK

⁴Westat Inc, Rockville, Maryland, USA

⁵Istituto Nazionale Tumori, IRCCS, Fondazione G. Pascale, Napoli, Italy

⁶Department of Oncology, Azienda Ospedaliero-Universitaria San Luigi Gonzaga, Orbassano, Italy

⁷Division of Medical Oncology, IRCCS San Matteo University Hospital Foundation, Pavia, Italy

⁸Department of Biomedical Sciences and Human Oncology, University of Bari A. Moro, Bari, Italy

⁹Department of Medical Oncology, IRCCS Azienda Ospedaliera Universitaria San Martino-IST, Genoa, Italy

¹⁰Fondazione IRCCS, Istituto Nazionale dei Tumori, Oncology Unit 1, Milan, Italy

¹¹Medical Oncology & Breast Unit, Antonio Perrino Hospital, Brindisi, Italy

¹²Ospedale della Val d'Elsa Senese, Poggibonsi, Italy

¹³Department of Oncology, San Camillo Forlanini Hospital, Rome, Italy

¹⁴Medical Oncology, Istituto di Ricovero e Cura a Carattere Scientifico Ospedale Oncologico, Bari, Italy

¹⁵Ospedale Civile SS. Annunziata, Sassari, Italy

¹⁶Department of Oncology, G.B. Rossi Hospital, University of Verona, Verona, Italy

¹⁷Oncology Unit 1, Istituto Oncologico Veneto IRCCS, Padova, Italy

¹⁸Division of Medical Oncology, Hospital of Cremona, Cremona, Italy

¹⁹Department of Clinical and Molecular Medicine, Università La Sapienza di Roma, Roma, Italy

²⁰Medical Oncology Unit, A.O. Papardo & Department of Human Pathology University of Messina, Messina, Italy

²¹Fondazione Smith Kline, Verona, Italy

²²Federazione Italiana delle Associazioni di Volontariato in Oncologia - F.A.V.O., Rome, Italy

²³Associazione Italiana Malati di Cancro, parenti ed amici - AIMaC, Rome Italy

²⁴Division of Cancer Control and Population Sciences, Outcomes Research Branch, National Cancer Institute, Rockville, Maryland, USA

± Both authors contributed equally

Corresponding author:

Francesco Perrone, Istituto Nazionale per lo Studio e la Cura dei Tumori, Fondazione G. Pascale IRCCS, Via Mariano Semmola, Napoli, 80131, Italy.

Email: f.perrone@istitutotumori.na.it

Methods: All PRO-CTCAE items were translated in accordance with international guidelines. Subsequently, the content validity of the PRO-CTCAE-Italian was explored and iteratively refined through cognitive debriefing interviews. Participants (n=96; 52% male; median age 64 years; 26% older adults; 18% lower educational attainment) completed a PRO-CTCAE survey and participated in a semi-structured interview to determine if the translation captured the concepts of the original English language PRO-CTCAE, and to evaluate comprehension, clarity and ease of judgement. Test-retest reliability of the finalized measure was explored in a second sample (n=135).

Results: Four rounds of cognitive debriefing interviews were conducted. The majority of PRO-CTCAE symptom terms, attributes and associated response choices were well-understood, and respondents found the items easy to judge. To improve comprehension and clarity, the symptom terms for nausea and pain were rephrased and retested in subsequent interview rounds. Test-retest reliability was excellent for 41/49 items (84%); the median intraclass correlation coefficient was 0.83 (range 0.64-0.94).

Discussion: Results support the semantic, conceptual and pragmatic equivalence of PRO-CTCAE-Italian to the original English version, and provide preliminary descriptive evidence of content validity and reliability.

Keywords

Linguistic validation, PRO-CTCAE, patient-reported outcomes, symptomatic adverse events, cognitive interviewing

Date received: 15 February 2022; revised: 4 April 2022; accepted: 21 April 2022

Background

The National Cancer Institutes's (NCI's) Common Terminology Criteria for Adverse Events (CTCAE) is an internationally accepted system for clinician grading and reporting of adverse events (AEs) in cancer clinical trials.¹ However, a clinician report may underestimate the prevalence and severity of symptomatic adverse events.²⁻⁴ Inclusion of direct patient self-reporting of symptomatic adverse events using patient-reported outcomes (PROs) in oncology has been widely endorsed as complementary to clinician-reporting,⁵⁻¹¹ thereby improving precision and reliability in the reporting of symptomatic AEs.¹²⁻¹⁴

To facilitate patient self-reporting of symptomatic adverse events in cancer clinical trials, NCI has developed the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE®).^{15,16} PRO-CTCAE is an item library comprised of 124 self-report items reflecting 78 symptomatic AEs drawn from the CTCAE. Each symptomatic AE is evaluated using a combination of one to three attributes (frequency, amount, severity, interference, presence/absence). A multi-site cognitive interviewing study found the English language PRO-CTCAE items to be well comprehended, acceptable, and meaningful to patients undergoing cancer treatment.^{17,18} PRO-CTCAE also demonstrates strong measurement properties, including construct validity and test-retest reliability.¹⁹ The validity of the 7-day recall period,²⁰ the discriminative validity of the PRO-CTCAE response choices,²¹ and the measurement equivalence of paper, web and interactive voice response modes of administration²² have also been confirmed. Translation and cross-cultural adaptation of PRO-CTCAE is being conducted in a number of different languages.²³

It is well recognized that if PRO measures are to be used in different countries and cultures they must demonstrate cultural acceptability, as well as linguistic and conceptual equivalence to the source measure.²⁴⁻²⁶ The aim of this study was to develop an Italian language version of PRO-CTCAE, and examine its content validity and test-retest reliability in a diverse sample of Italian-speaking patients undergoing cancer treatment.

Materials and methods

This multisite study recruited participants from 15 cancer treatment centers around Italy and used a mixed methods approach to develop and refine a culturally acceptable and well-comprehended Italian language version of PRO-CTCAE. The study was approved by the Ethics Committees of all participating centers; all participants provided written informed consent.

Translation

We followed internationally-established procedures for the translation and cross-cultural adaptation of PRO instruments.²⁴⁻²⁶ Two independent forward-translations of PRO-CTCAE were prepared by Italian-speaking bilingual translators. As recommended by Beaton et al.,²⁵ one translator was aware of the PRO-CTCAE item definitions, while the other was not provided any contextual details about the project, and had no medical background. The translators then produced a single reconciled Italian version. The reconciled PRO-CTCAE instrument then underwent dual independent back-translation into English by two bilingual translators who had no specific background in oncology, were not informed about the specific project aims, and had not seen the English PRO-CTCAE source

materials. In preparing the final reconciled version to be advanced for linguistic validation, each PRO-CTCAE-Italian item, including the attributes (e.g. severity), response choices, and phrasing for ‘severity at its worst’ were examined to ensure that the PRO-CTCAE-Italian was equivalent to the PRO-CTCAE-English with respect to semantic (similar meaning), conceptual (similar relevance of the constructs), and pragmatic (similar tone or communicative effect) equivalence.

Cognitive debriefing interviews

Interviews were conducted to explore the comprehensibility, clarity and ease of response of PRO-CTCAE-Italian. Eligible study participants were 18 years of age or older; were receiving (or had received within the past six months) anticancer therapy with curative or palliative intent; and were able to speak, read and comprehend Italian as their primary language. Patients judged by their clinicians as unable to recall symptoms during the past seven days due to cognitive impairment were excluded.

Given the intended context of use across a range of disease sites and treatment types, a diverse sample with respect to cancer site, treatment type, age, educational attainment, and geographic regions in Italy (North, Center and South/Island) was recruited from 15 cancer treatment centers around Italy. To avoid potential biases due to regional literacy patterns, we established an accrual goal that at least 25% of the participants from each geographic area should have lower educational attainment (defined as primary school level or lower, i.e., five or fewer years of schooling).

The interview methodology followed the approach used in the linguistic validation of other PRO-CTCAE language versions.^{17,27-31} In the first round of interviews (n=47), males were debriefed on all 74 gender-relevant symptom terms (120 items), and females were debriefed on 76 gender-relevant symptom terms (122 items). After a protocol amendment, in Rounds 2 through 4, we employed eight shorter interview schedules to reduce the duration of the interviews and address participant burden. Each of the eight shorter interview schedules (four per gender) contained 29-32 symptom terms for debriefing. Interview schedules were deployed in a manner that ensured that ultimately all PRO-CTCAE items received approximately equivalent attention during the debriefing process.

Participants were first asked to independently complete a series of PRO-CTCAE items. Subsequently, they were debriefed by an Italian-speaking, master’s prepared interviewer with experience interviewing research participants in medical settings. A semi-structured interview guide (see Supplementary Table 1) was used to examine the comprehension, clarity and acceptability of the PRO-CTCAE symptom terms, the attributes (e.g. severity, interference) and the associated response options, as well as the clarity

and ease of judgement for ‘severity at its worst’ and the 7-day recall period.¹⁷ For each PRO-CTCAE element, four categories related to cognitive processing were considered: comprehension, memory retrieval, ease of judgment, and response mapping.³¹ Cognitive interviews were audio-recorded, and extensive field notes were taken by interviewers. Items with > 10% respondent difficulties were considered for rephrasing and/or retesting in a subsequent interview round. Interviews continued until no additional comprehension difficulties were identified.

As recommended by Beaton et al.,²⁵ following the translation and adaptation process, it is important to explore quantitatively the measurement properties of the newly adapted instrument. Test-retest reliability is a fundamental property of a measure upon which construct validity, responsiveness, and sensitivity to change are based. Reliability may be demonstrated through either internal consistency reliability and/or test-retest reliability methods.³² Internal consistency reliability is only relevant when items are part of a unidimensional scale or factor. However, for a library of symptom items such as PRO-CTCAE, evidence of consistent scores over time in a population with stable symptoms supports a conclusion that the instrument is reliable. We therefore investigated the test-retest reliability of our cross-culturally adapted instrument in a second sample of adult participants (n=135) who were currently undergoing or had recently completed cancer treatment at one of six participating cancer centers, and were willing to complete a PRO-CTCAE survey on the day of study enrollment and again within 24-36 hours. To limit respondent burden, a subset of 26 symptomatic AEs (49 PRO-CTCAE items) was presented to each participant. To enhance comparability with previously published test-retest reliability coefficients, these were the same 26 symptomatic AEs that were previously evaluated by Dueck et al.¹⁹ in their study to evaluate the measurement properties of PRO-CTCAE in English. Intraclass correlation coefficients (ICC), and their 95% confidence intervals, were calculated based on a two-way mixed-effect analysis of variance model with interaction for the absolute agreement between single scores.³³ ICC values were interpreted as follows: Values less than 0.50, between 0.50 and 0.75, between 0.75 and 0.90, and greater than 0.90 were interpreted as indicating poor, moderate, good, and excellent reliability, respectively.³³⁻³⁵

Results

Translation

Comparison of the two forward-translations revealed semantic equivalence (‘same words’) and conceptual (‘same meaning’) equivalence for 29 of the 78 PRO-CTCAE symptom terms. Translation of the symptom term “fatigue” caused some difficulty, as there is no Italian

Table 1. Sample characteristics.

Characteristic	Cognitive debriefing (N=96)		Test-retest (N=135)	
	N	(%)	N	(%)
Geographic Region of Residence				
Northern Italy	32	33.3	112	83.0
Central Italy	31	32.3	17	12.6
Southern Italy/Islands	33	34.4	6	4.4
Age (in years), median [IQR]	64 [46-75]		63 [54-71]	
Age (in years)				
18 – 44	23	24.0	12	8.9
45 – 59	23	24.0	42	31.1
60 – 74	25	26.0	65	48.1
75 +	25	26.0	16	11.9
Gender, N. (%)				
Female	46	48	55	40.7
Male	50	52	80	59.3
Education, N (%)				
Primary school or less	17	17.7	12	8.9
Middle school	35	36.4	67	49.6
High school / University Degree	44	45.9	56	41.5
Cancer type, N (%)*				
Lung, head, or neck	-	-	63	46.7
Breast	-	-	14	10.4
Genitourinary or gynecologic	-	-	41	30.4
Gastrointestinal	-	-	5	3.7
Hematologic	-	-	1	0.7
Other or unknown	-	-	11	8.1

Note: *Data about disease site was not captured in cognitive interviewing participants.

word that encompassed all of the meanings of the English term, and an explanatory elaboration of the Italian term chosen for fatigue (*senso di fatica, stanchezza o mancanza di energia*) was included to improve conceptual and pragmatic equivalence. In addition, the phrase “at its worst” proved problematic to translate because of syntactic differences between English and Italian, (translated as *nel momento peggiore*). In the two independent back-translations, 29 of the 78 PRO-CTCAE symptom terms exhibited both semantic and conceptual equivalence to the English source; 46 of the 78 PRO-CTCAE symptom terms were conceptually equivalent. There were three symptom terms (pain in the abdomen [belly]; breast area enlargement or tenderness; and feeling that nothing could cheer you up) that required careful deliberation among the translators and an expert committee of clinicians and methodologists in order to settle on final Italian language phrasings that were conceptually equivalent to the English source material.

Cognitive debriefing interviews

Participants (n=96) were recruited at 15 cancer centers throughout Italy to participate in one of four consecutive

rounds of cognitive interviewing. The sample was diverse with respect to age, educational attainment, and geographic region (Table 1). 54% of the sample had completed eight or fewer years of education. Participants enrolled in Round 1 (n=47) completed a survey comprised of all the PRO-CTCAE items (122 for women and 120 for men), followed by cognitive debriefing on all of the relevant symptom terms. The average time required to complete each PRO-CTCAE item was 14 seconds (SD ±6 seconds). Cognitive debriefing interviews in Round 1 lasted an average of 86 minutes (SD ± 27 minutes). Six of 47 participants experienced the length of the survey as taxing, with two respondents unable to complete the survey, one unable to complete the interview, and three requiring a rest break during the study-related procedures.

In Rounds 2 through 4, participants completed a shorter survey comprised of a subset of 29-34 PRO-CTCAE symptom terms, followed by cognitive debriefing only on the items included in that survey. The average time to complete each PRO-CTCAE item in Rounds 2 through 4 was 22 seconds (SD ± 8), 14 seconds (SD ± 5), and 15 seconds (SD ± 6), respectively. The average time for cognitive debriefing interviews in Rounds 2 through 4 was 55 minutes (SD ±18 minutes), 47 minutes (SD ±15 minutes),

Table 2. Comprehension difficulties across four rounds of cognitive interviews.

Symptom Term in English	Symptom Term in Italian	Round 1 (N=47)	Round 2 (N=15)	Decision after Round 2	Round 3 and 4 (N=34)
		Participants with difficulty	Participants with difficulty		Participants with difficulty
		N (%)	N (%)		
Achieve and maintain erection	Difficoltà a raggiungere o mantenere l'erezione	2/12 (16.6)	-	No suitable alternative; retested with adjusted probes to allow for greater participant comfort when paraphrasing	2/4 (50.0)
Bed/pressure sores	Piaghe da decubito	6/44 (13.6)	3/5 (60.0)	No suitable alternative	0/14 (0)
Ejaculation	Problemi di eiaculazione	2/17 (11.8)	0/2 (0)	No suitable alternative; retested	1/4 (25.0)
General pain	Dolore	10/46 (21.7)	1/15 (6.7)	Rephrased to: dolore fisico	0/34 (0)
Flashing lights	Lampi davanti agli occhi	7/45 (15.6)	0/4 (0)	No suitable alternative	0/17 (0)
Missed expected menstrual period	Regolarità delle mestruazioni	2/28 (7.1)	0/1 (0)	No suitable alternative; retested	1/3 (33.3)
Nausea	Nausea	6/47 (12.8)	2/15 (13.3)	Rephrased to add parenthetical: nausea (senso di malessere allo stomaco)	0/28 (0)
Blurred vision	Appannamento della vista	3/47 (6.4)	0/4 (0)	No suitable alternative; retested	1/7 (14.3)
Itching	Prurito	4/46 (8.7)	0/5 (0)	Phrasing retained and retested	0/13 (0)
Vomiting	Vomito	3/46 (6.5)	2/15 (13.3)	No suitable alternative; retested	0/33 (0)
Joint pain	Dolore alle articolazioni	1/46 (2.2)	0/5 (0)	Phrasing retained and retested	0/13 (0)

and 47 minutes (SD \pm 28 minutes), respectively. In Rounds 2 through 4, all participants were able to complete both the survey and the debriefing interview without becoming unduly fatigued.

As shown in Figure 1, approximately half of all PRO-CTCAE symptom terms posed comprehension difficulties for at least one respondent. Rounds 1 and 2 of cognitive interviews (see Table 2) revealed four PRO-CTCAE-Italian items presenting difficulties in more than 10% of participants: bedsores (9/49, 18.4%), general pain (11/62, 17.7%), flashing lights (7/49, 14.3%), and nausea (8/62, 12.9%). Respondents were uncertain if “Pain”, originally translated as *Dolore*, should be interpreted as referring to physical pain or emotional pain. The item was revised to *Dolore Fisico* (physical pain) and was subsequently well-comprehended by all 23 participants in Round 3. Difficulty with comprehension of the symptom term for “Nausea” appeared to be caused by respondents conflating the concept of nausea with vomiting; comprehension difficulties with “Vomiting” were also noted in 5/61 patients (8.2%). An elaboration in parentheses that clarified that nausea referred to “feeling sick to your stomach”, or *senso di malessere allo stomaco*, was added. The revised phrasing for nausea incorporating this parenthetical was well comprehended by all participants (n=28) in Round 3. The phrasing for “Vomiting” was not modified as no suitable

alternative was identified, and there were no further difficulties with comprehension of “Vomiting” in Round 3. The Italian phrasing for “Bedsores” and “Flashing Lights” posed comprehension difficulties for 18% and 16% of respondents in the first two interview rounds. However suitable alternative phrasing could not be identified by the study team, and both of these symptom terms appear to be well understood by patients who had experienced these symptoms. Two additional terms (“Joint Pain” and “Itchy Skin”) also received additional debriefing in Rounds 3 and 4. These terms were well understood by respondents and as a result no phrasing changes were made. Across four interview rounds, no respondents demonstrated difficulties understanding the PRO-CTCAE attributes (e.g. frequency, severity, interference) or the associated response options. All respondents well-comprehended the intended meaning of “at its worst”, and there were no expressed difficulties in making judgements of “at its worst” or using the recall period of the past seven days.

Test-retest reliability

The test-retest reliability substudy enrolled 137 unique participants at 11 cancer centers in Italy; 135 participants had complete data. Participants completed 46 PRO-CTCAE items on two occasions (median interval between

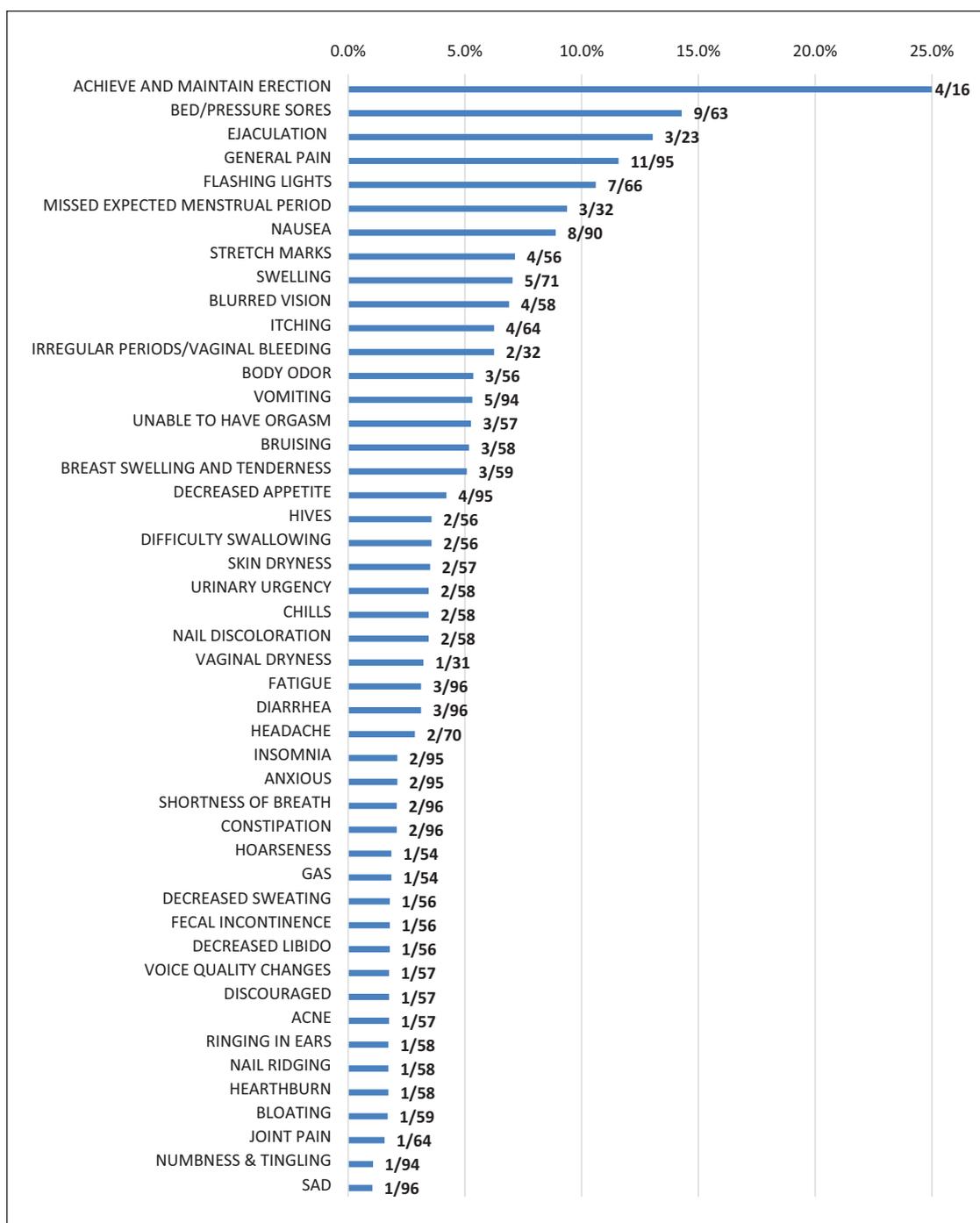


Figure 1. Waterfall plot of rates of comprehension difficulties with PRO-CTCAE-Italian symptom terms across four rounds of interviews (n=96) and the total number of participants debriefed on each item.

two surveys was 1 day; range 1-3 days). Due to rapid accrual, the sample size exceeded the minimum required based on the a priori power analysis. Characteristics of the participants are summarized in Table 1. Median age was 63 years; 42% were female; 58.4% had less than a high school education. A variety of tumor sites were reflected in the sample, including lung (47.4%), genitourinary (29.9%),

breast (10.2%), gastrointestinal (9.5%) and other (2.9%). In the 135 respondents who completed PRO-CTCAE on two consecutive business days, the test-retest reliability for the 49 prespecified items ranged from ICC 0.64 to 0.94 (median ICC 0.83) (Supplementary Table 2); 84% of the items examined demonstrated an ICC greater than or equal to 0.75 (Figure 2).

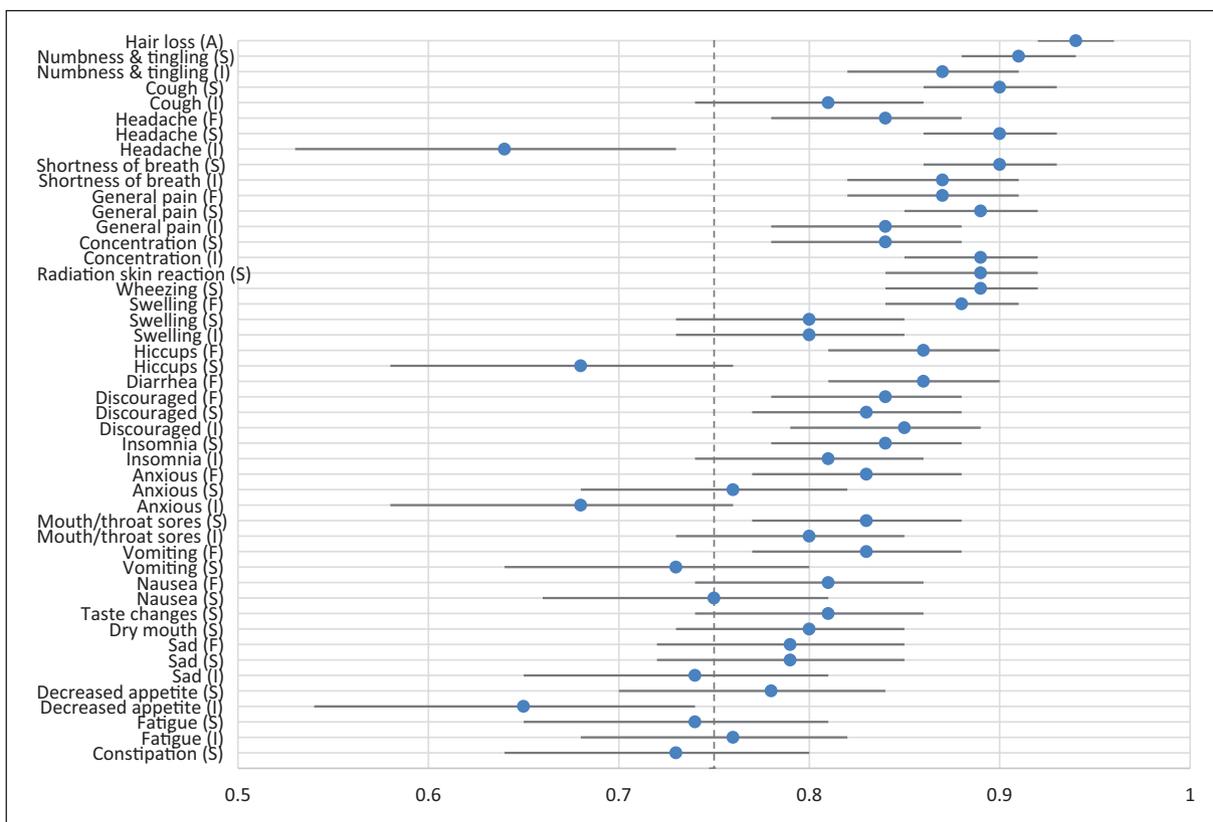


Figure 2. Test-retest reliability of 49 PRO-CTCAE-Italian Items: Intraclass Correlation Coefficient and 95% CI (n=135).

A, Amount; F, Frequency; I, Inference with usual or daily activities; PRO-CTCAE, Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events; S, Severity.

Note: Rash [Presence] (n=135): Proportion of agreement =84%, Range: 65%-100%; Skin darkening [Presence] (n=134): Proportion of agreement = 86%, Range: 74%-97%.

Participants in the test-retest sample were highly symptomatic, endorsing the presence of a median of 20 symptoms (IQR 15-28). 97% of the sample (131/135) reported the presence of five or more concurrent symptoms at the first visit. Of the 49 PRO-CTCAE items evaluated in this study, 48 (98%) were reported as present by at least 10% of respondents at both visits, with 37 of 49 (76%) items having at least 25% prevalence (Supplementary Table 3). The distribution of item scores for the set of 20 commonly occurring symptomatic AEs, as identified by Dueck et al.,¹⁹ is depicted in Figure 3.

Discussion

We report the cross-cultural adaptation, linguistic validity and test-retest reliability of the PRO-CTCAE-Italian. The translation process emphasized the use of plain language phrasing to promote comprehension by diverse respondents, including those with lower educational attainment. Review of the back-translations demonstrated that PRO-CTCAE-Italian was semantically and conceptually consistent with the English source. Four rounds of cognitive interviews were conducted, resulting in adjusted phrasing

in two of the PRO-CTCAE symptom terms to achieve satisfactory comprehension. Results of the cognitive testing confirm the comprehensibility, clarity and ease of judgment of the Italian PRO-CTCAE, and support its conceptual equivalence to the English source. PRO-CTCAE-Italian also exhibited excellent test-retest reliability as indicated by a median ICC of 0.83.¹⁹ The absence of prominent floor and ceiling effects is supported by the fact that more than 95% of respondents reported the presence of 5 or more symptoms, and exhibited high scores (>3) for at least one PRO-CTCAE symptom (91/135, 67.4%) at first visit, and 84/135 (62.2%) at second visit. These findings also support a conclusion that reliability coefficients were not erroneously amplified due to low variability in a symptom-free sample. Our test-retest reliability parameters are comparable to those reported by Dueck et al.¹⁹ for the English language version of PRO-CTCAE, and provide encouraging preliminary evidence that PRO-CTCAE-Italian yields reproducible scores.

Both the strengths of this study as well as some limitations should be considered in interpreting our findings. First, because our study was conducted in patients currently receiving cancer treatment, and because some

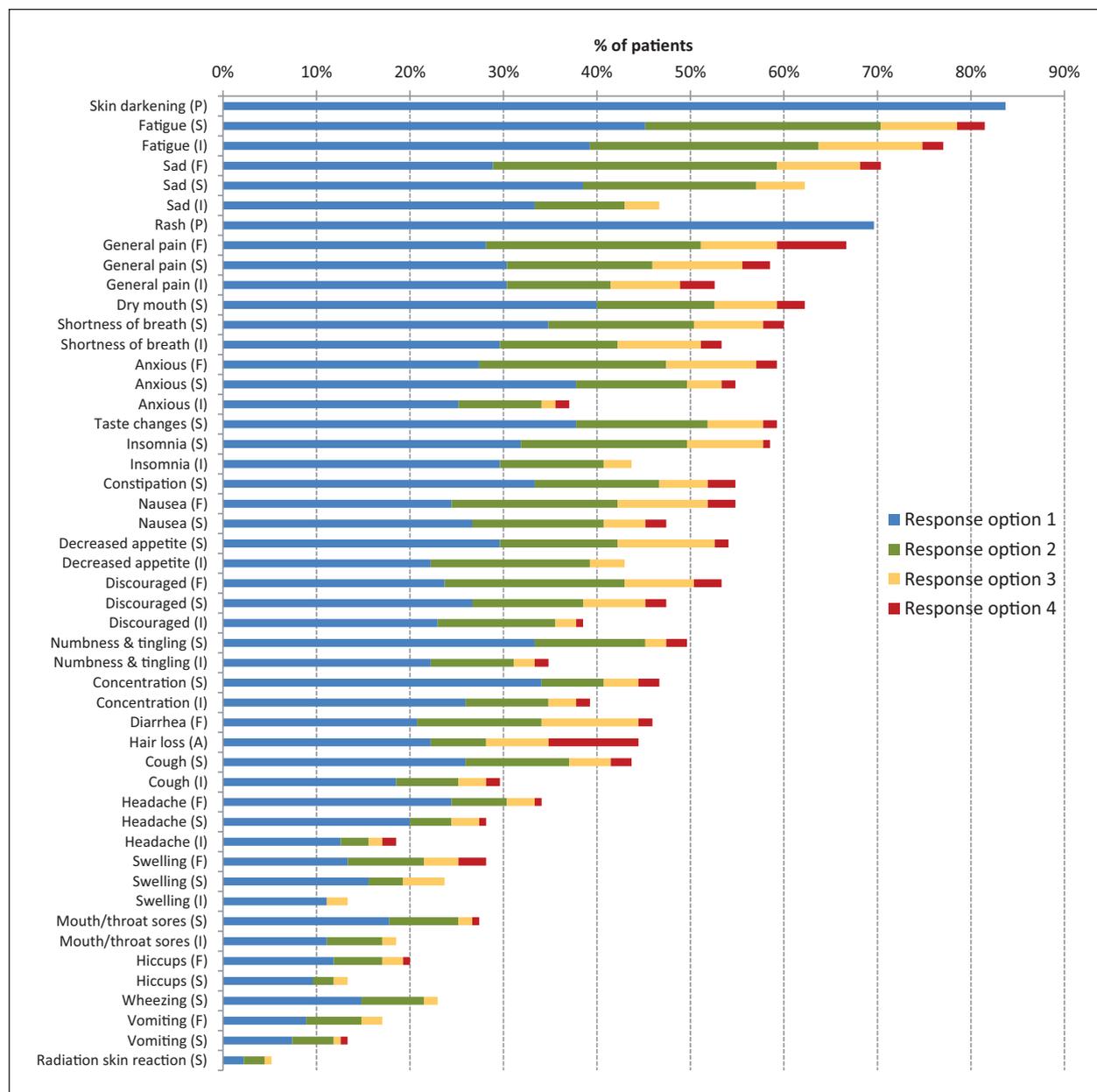


Figure 3. Distribution of PRO-CTCAE Item Scores for Symptomatic Toxicities Prevalent in $\geq 10\%$ of Respondents at Visit 1 (n=135).

A, Amount; F, Frequency; I, Inference with usual or daily activities; P, Presence; PRO-CTCAE, Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events; S, Severity.

Response options:

Frequency: 0= "Never", 1= "Rarely", 2= "Occasionally", 3= "Frequently", 4= "Almost constantly".

Severity: 0= "None", 1= "Mild", 2= "Moderate", 3= "Severe", 4= "Very severe".

Inference: 0= "Not at all", 1= "A little bit", 2= "Somewhat", 3= "Quite a bit", 4= "Very much".

Amount: 0= "Not at all", 1= "A little bit", 2= "Somewhat", 3= "Quite a bit", 4= "Very much".

Presence/Absence: 0= "No", 1= "Yes".

interviews lasted more than an hour, participants had to have sufficient endurance to complete study-related procedures. Thus, our sample may have been biased toward inclusion of participants with better performance status. Although we did not formally collect data to evaluate this possibility, we do note that the majority of our respondents

were experiencing disease- or treatment-related symptoms. Second, we did not capture cancer site in the participants in the cognitive debriefing component of the study. While the sample in the test-retest substudy was somewhat diverse with respect to disease site, we acknowledge that almost 50% of the test-retest sample were being treated for

lung or head and neck cancers. At the same time, patients with lung and head/neck cancers often have prominent symptom burden,^{36,37} and the high level of symptom burden observed in our sample helps to mitigate the possibility that our reliability estimates were inflated in the absence of symptoms.³⁸ We also acknowledge that the small sample used to examine test-retest reliability was drawn from a single center, and as such future study to determine that our reliability estimates are replicable is warranted. Third, comprehension problems for two items, bedsores and flashing lights, persisted in a small proportion of participants; these could not be resolved because no suitable phrasing alternatives could be identified for these symptom terms. Encouragingly, we noted during cognitive interviews that both symptom terms were generally well understood by respondents who had direct experience with these symptoms. Lastly, all of our interviewers were female and there was some hesitancy and embarrassment observed on the part of male respondents during the part of the interview when they were asked to rephrase in their own words symptoms terms reflecting urinary or sexual dysfunction. While this improved in Rounds 3 and 4 with adjustments to the interview probes as recommended by Arthur et al.,³⁹ we cannot exclude the possibility that gender mismatch between interviewers and study participants may have influenced the accuracy and quality of our data. At the same time, strengths of the study include compliance with rigorous procedures for translation and cross-cultural adaptation of PRO instruments. An additional strength was the large and diverse sample, which included approximately 20% of respondents with an extremely low level of educational attainment, a proportion that reflects the Italian general population where approximately 18% have completed only primary school education or less.⁴⁰ Lastly, inclusion of the evaluation of test-retest reliability provides supportive evidence that the comprehensibility of PRO-CTCAE-Italian is reproducible. Additional research examining the psychometric properties and responsiveness of PRO-CTCAE-Italian in a large sample of Italian speakers undergoing cancer treatment is planned.

Conclusions

We have developed and content validated an Italian language version of PRO-CTCAE for use in cancer clinical trials. These study results advance the ongoing efforts to make the PRO-CTCAE measurement system available to patients of different linguistic and cultural backgrounds.²³ Currently, PRO-CTCAE is available in more than 35 languages, with additional languages in development and testing. This availability of the PRO-CTCAE measurement system in a wide variety of languages and developed using rigorous procedures for translation and cultural adaptation will allow an expanded number of patients to

provide their perspective on the tolerability of cancer treatment, and supports the multinational conduct of trials of new therapies for cancer.

Acknowledgements

We gratefully acknowledge the four Italian translators, Enrica Bondavalli, Anne Prudence Collins, Claudia Di Loreto, and Stephanie Johnson. We also thank Luca Gattinoni, MD, Experimental Transplantation and Immunology Branch, National Cancer Institute, for his expert review of the PRO-CTCAE-Italian instrument prior to the cognitive interviews.

Authors' contributions

The authors confirm contributions to the paper as follows:

Study conception and design: CC, JB, CP, LDM, RP, MBag, MBas, AN, DDP, Elan FP, SAM.

Participant recruitment and data collection: JB, LS, SN, CP, LDM, GP, SC, AFal, FCal, VL, AAC, GT, MM, RP, FCog, VA, EC, AFer, ADA

Analysis and Interpretation of data: CC, JB, SR, FD, Elez, CP, LDM, RP, MBag, MBas, AN, DDP, Elan, FP, SAM.

Drafting of the manuscript: CC, JB, DN, FD, FP, SAM.

All authors reviewed and approved the final version of the manuscript.

Availability of data and material

The datasets used and analyzed for the current study are available at https://zenodo.org/record/6033723#.YgTW_ZbSKUk.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article:

SN received payments or honoraria from Eli Lilly, Sanofi, Amgen, Beigene, Roche, Takeda, Pfizer, AstraZeneca, Boehringer Ingelheim, MSD, Novartis.

CP received payments or honoraria from Eisai, MSD, BMS, Astra Zeneca, Pfizer, Novartis, Merck, Janssen, Ipsen, EUSA Pharma, Angelini, General Electric.

LDM received payments or honoraria from Roche, Novartis, Eli Lilly, MSD, Pfizer, Ipsen, Celgene, Genomic Helath, Pierre Fabre, Daiichi Sankyo, Seagen, AstraZeneca, Eisai.

GP received payments or honoraria BMS, MSD, Bayer, Astra Zeneca, Merck, Ipsen, Janssen, Novartis, Pfizer, Sanofi.

SC received payments or honoraria from Eli Lilly.

FCal received payments or honoraria from MSD, BMS, Pfizer, AstraZeneca, Ipsen.

GT received payments or honoraria from BMS, MSD, Servier, Astra Zeneca, Merck.

RP received payments or honoraria from Regeneron, Ipsen, Sanofi-Aventis, Amgen, MSD.

FCog received payments or honoraria from Genomic Helath, Biolitec Pharma, Pfizer, Abbott, Celgene, Glaxo, Roche, Bayer, Novartis, Amgen, Pfizer, AstraZeneca, Eisai, Merck-Seron, Boehringer Ingelheim, MSD, BMS, Takeda, Astella, Eli Lilly, Seagen.

VA received payments or honoraria from Amgen, Astra Zeneca, BMS, Eli Lilly, MSD, Novartis, Pfizer, Roche, Sandoz, Sanofi, Servier, Takeda.

DDP received grants to his institution from Fondazione Smith Kline.

Elan received grants to her institution from Fondazione Smith Kline.

FP received payments or honoraria from Incyte, GSK, Eli Lilly, Ipsen, Astellas, Astra Zeneca, Roche, BMS, Bayer, Clovis, Pierre Fabre; FP also received grants to his institution from Roche, Astra Zeneca, Pfizer, MSD, Bayer, Incyte Taiho, Janssen, Exelixis, Aileron, Daiichi Sankyo.

CC, JB, SR, DN, FD, Elez, LS, AFal, VL, AAC, MM, EC, AFer, MBag, MBas, AN, ADA, SAM did not declare competing interests.

Ethics approval and consent to participate

Study oversight and ethics approval was provided by the Independent Ethics Committee of the National Cancer Institute of Naples and by Ethics Committee at each Italian participating centre. All study participants provided written consent to participate in this study.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this research was supported in part by an unrestricted grant of the Fondazione Smith-Kline.

ORCID iDs

Caterina Caminiti  <https://orcid.org/0000-0002-3196-2435>

Camillo Porta  <https://orcid.org/0000-0003-2412-1563>

Giuseppe Procopio  <https://orcid.org/0000-0002-2498-402X>

Francesco Perrone  <https://orcid.org/0000-0002-9738-0526>

Supplemental material

Supplemental material for this article is available online.

References

- National Cancer Institute, National Institutes of Health. US Department of Health and Human Services: Common Terminology Criteria for Adverse Events (CTCAE), <https://evs.nci.nih.gov/ftp1/CTCAE/About.html> (accessed 25 January 2022).
- Fromme EK, Eilers KM, Mori M, et al. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J Clin Oncol* 2004; 22: 3485-3490.
- Basch E. The missing voice of patients in drug-safety reporting. *N Engl J Med* 2010; 362: 865-869.
- Basch E, Artz D, Dulko D, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. *J Clin Oncol* 2005; 23: 3552-3561.
- Kirkova J, Davis MP, Walsh D, et al. Cancer symptom assessment instruments: a systematic review. *J Clin Oncol* 2006; 24: 1459-1473.
- Trotti A, Colevas AD, Setser A, et al. Patient-reported outcomes and the evolution of adverse event reporting in oncology. *J Clin Oncol* 2007; 25: 5121-5127.
- Sloan JA, Berk L, Roscoe J, et al. Integrating patient-reported outcomes into cancer symptom management clinical trials supported by the National Cancer Institute-sponsored clinical trials networks. *J Clin Oncol* 2007; 25: 5070-5077.
- Bruner DW. Should patient-reported outcomes be mandatory for toxicity reporting in cancer clinical trials? *J Clin Oncol* 2007; 25: 5345-5347.
- Basch E, Abernethy AP, Mullins CD, et al. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. *J Clin Oncol* 2012; 30: 4249-4255.
- Johnson JR, Williams G and Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol* 2003; 21: 1404-1411.
- European Medicines Agency Committee for Medicinal Products for Human Use. Preauthorisation evaluation of medicines for human use: Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products, https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-regulatory-guidance-use-healthrelated-quality-life-hrql-measures-evaluation_en.pdf (accessed 25 January 2022).
- King A, Daniels J, Lim J, et al. Time to listen: a review of methods to solicit patient reports of adverse events. *Qual Saf Health Care* 2010; 19: 148-157.
- Stacey D, Bakker D, Ballantyne B, et al. Managing symptoms during cancer treatments: evaluating the implementation of evidence-informed remote support protocols. *Implement Sci* 2012; 7: 110.
- Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *Jama* 2012; 307: 1636-1640.
- Basch E, Reeve BB, Mitchell SA, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst* 2014; 106(9): dju244.
- National Institutes of Health, National Cancer Institute, Division of Cancer Control & Population Sciences. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE™), <http://healthcaredelivery.cancer.gov/pro-ctcae> (accessed 25 January 2022).
- Hay JL, Atkinson TM, Reeve BB, et al. Cognitive interviewing of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Qual Life Res* 2014; 23: 257-269.
- Reeve BB, McFatrigh M, Pinheiro LC, et al. Cognitive interview-based validation of the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events in adolescents with cancer. *J Pain Symptom Manage* 2017; 53: 759-766.
- Dueck AC, Mendoza TR, Mitchell SA, et al. Validity and reliability of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *JAMA Oncol* 2015; 1: 1051-1059.
- Mendoza TR, Dueck AC, Bennett AV, et al. Evaluation of different recall periods for the US National Cancer Institute's PRO-CTCAE. *Clin Trials* 2017; 14: 255-263.

21. Atkinson TM, Hay JL, Dueck AC, et al. What do "None," "Mild," "Moderate," "Severe," and "Very Severe" mean to patients with cancer? Content validity of PRO-CTCAE™ response scales. *J Pain Symptom Manage* 2018; 55: e3-e6.
22. Bennett AV, Dueck AC, Mitchell SA, et al. Mode equivalence and acceptability of tablet computer-, interactive voice response system-, and paper-based administration of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Health Qual Life Outcomes* 2016; 14: 24.
23. Klutz PG, Chingos DT, Basch EM, et al. Patient-reported outcomes in cancer clinical trials: Measuring symptomatic adverse events with the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Am Soc Clin Oncol Educ Book* 2016; 35: 67-73.
24. Guillemin F. Cross-cultural adaptation and validation of health status measures. *Scand J Rheumatol* 1995; 24: 61-63.
25. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000; 25: 3186-3191.
26. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005; 8: 94-104.
27. Arnold B, Mitchell SA, Lent L, et al. Linguistic validation of the Spanish version of the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Support Care Cancer* 2016; 24: 2843-2851.
28. Kirsch M, Mitchell SA, Dobbels F, et al. Linguistic and content validation of a German-language PRO-CTCAE-based patient-reported outcomes instrument to evaluate the late effect symptom experience after allogeneic hematopoietic stem cell transplantation. *Eur J Oncol Nurs* 2015; 19: 66-74.
29. Bæksted C, Nissen A, Pappot H, et al. Danish translation and linguistic validation of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *J Pain Symptom Manage* 2016; 52: 292-297.
30. Cho J, Yoon J, Kim Y, et al. Linguistic validation of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events in Korean. *J Glob Oncol* 2019; 5: 1-10.
31. Willis GB. *Analysis of the Cognitive Interview in Questionnaire Design*. New York: Oxford University Press; 2015.
32. Revicki DA, Gnanasakthy A and Weinfurt K. Documenting the rationale and psychometric characteristics of patient reported outcomes for labeling and promotional claims: the PRO Evidence Dossier. *Qual Life Res* 2007; 16: 717-723.
33. Koo TK and Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016; 15: 155-163.
34. Altman DG. *Practical Statistics for Medical Research*. London: CRC press; 1990.
35. Qin S, Nelson L, McLeod L, et al. Assessing test-retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. *Qual Life Res* 2019; 28: 1029-1033.
36. Mendoza TR, Kehl KL, Bamidele O, et al. Assessment of baseline symptom burden in treatment-naïve patients with lung cancer: an observational study. *Support Care Cancer* 2019; 27: 3439-3447.
37. Allen-Ayodabo CO, Eskander A, Davis LE, et al. Symptom burden among head and neck cancer patients in the first year after diagnosis: Association with primary treatment modality. *Oral Oncol* 2019; 99: 104434.
38. Paiva CE, Barroso EM, Carnesecca EC, et al. A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: a systematic review. *BMC Med Res Methodol* 2014; 14: 8.
39. Arthur EK, Menon U, Browning K, et al. Challenges of Cognitive Interviewing in Sensitive Health Topic Research. *Nurs Res* 2021; 70: 376-382.
40. I.Stat Statistics. Italian National Institute of Statistics, <http://dati.istat.it> (accessed 25 January 2022).