

## references

- Davies MA, Liu P, McIntyre S et al. Prognostic factors for survival in melanoma patients with brain metastases. *Cancer* 2011; 117: 1687–1696.
- Sperduto PW, Chao ST, Sneed PK et al. Diagnosis-specific prognostic factors, indexes, and treatment outcomes for patients with newly diagnosed brain metastases: a multi-institutional analysis of 4,259 patients. *Int J Radiat Oncol Biol Phys* 2010; 77: 655–661.
- Weber JS, D'Angelo SP, Minor D et al. Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol* 2015; 16: 375–384.
- Long GV, Stroyakovskiy D, Gogas H et al. Dabrafenib and trametinib versus dabrafenib and placebo for Val600 BRAF-mutant melanoma: a multicentre, double-blind, phase 3 randomised controlled trial. *Lancet* 2015; 386: 444–451.
- Long GV, Trefzer U, Davies MA et al. Dabrafenib in patients with Val600Glu or Val600Lys BRAF-mutant melanoma metastatic to the brain (BREAK-MB): a multicentre, open-label, phase 2 trial. *Lancet Oncol* 2012; 13: 1087–1095.
- Margolin K, Ernstoff MS, Hamid O et al. Ipilimumab in patients with melanoma and brain metastases: an open-label, phase 2 trial. *Lancet Oncol* 2012; 13: 459–465.
- Kefford GMR, Arance A, Nathan P et al. Vemurafenib in metastatic melanoma patients with brain metastases: an open-label, single-arm, phase 2, multicenter study. *Pigment Cell Melanoma Res* 2013; 26: 965.
- Kiess AP, Wolchok JD, Barker CA et al. Stereotactic radiosurgery for melanoma brain metastases in patients receiving ipilimumab: safety profile and efficacy of combined treatment. *Int J Radiat Oncol Biol Phys* 2015; 92: 368–375.
- Mathew M, Tam M, Ott PA et al. Ipilimumab in melanoma with limited brain metastases treated with stereotactic radiosurgery. *Melanoma Res* 2013; 23: 191–195.
- Ahmed KA, Freilich JM, Sloot S et al. LINAC-based stereotactic radiosurgery to the brain with concurrent vemurafenib for melanoma metastases. *J Neurooncol* 2015; 122: 121–126.
- Ahmed KA, Stallworth DG, Kim Y et al. Clinical outcomes of melanoma brain metastases treated with stereotactic radiation and anti-PD-1 therapy. *Ann Oncol* 2016; 27: 434–441.
- Eisenhauer EA, Therasse P, Bogaerts J et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45: 228–247.
- Patel BG, Ahmed KA, Johnstone PA et al. Initial experience with combined BRAF and MEK inhibition with stereotactic radiosurgery for BRAF mutant melanoma brain metastases. *Melanoma Res* 2016; 26: 382–386.
- Lin NU, Lee EQ, Aoyama H et al. Response assessment criteria for brain metastases: proposal from the RANO group. *Lancet Oncol* 2015; 16: e270–e278.
- Chang J, Atkinson H, A'Hern R et al. A phase II study of the sequential administration of dacarbazine and fotemustine in the treatment of cerebral metastases from malignant melanoma. *Eur J Cancer* 1994; 30A: 2093–2095.
- Agarwala SS, Kirkwood JM, Gore M et al. Temozolomide for the treatment of brain metastases associated with metastatic melanoma: a phase II study. *J Clin Oncol* 2004; 22: 2101–2107.
- Topalian SL, Sznol M, McDermott DF et al. Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab. *J Clin Oncol* 2014; 32: 1020–1030.
- Robert C, Karaszewska B, Schachter J et al. Improved overall survival in melanoma with combined dabrafenib and trametinib. *N Engl J Med* 2015; 372: 30–39.
- Larkin J, Chiarion-Sileni V, Gonzalez R et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 2015; 373: 23–34.
- Berghoff AS, Ricken G, Widhalm G et al. Tumour-infiltrating lymphocytes and expression of programmed death ligand 1 (PD-L1) in melanoma brain metastases. *Histopathology* 2015; 66: 289–299.
- Gaudy-Marqueste C, Carron R, Delsanti C et al. On demand Gamma-Knife strategy can be safely combined with BRAF inhibitors for the treatment of melanoma brain metastases. *Ann Oncol* 2014; 25: 2086–2091.
- Narayana A, Mathew M, Tam M et al. Vemurafenib and radiation therapy in melanoma brain metastases. *J Neurooncol* 2013; 113: 411–416.
- Fontanella AN, Boss MK, Hadsell M et al. Effects of high-dose microbeam irradiation on tumor microvascular function and angiogenesis. *Radiat Res* 2015; 183: 147–158.
- Yu C, Chen JC, Apuzzo ML et al. Metastatic melanoma to the brain: prognostic factors after gamma knife radiosurgery. *Int J Radiat Oncol Biol Phys* 2002; 52: 1277–1287.

*Annals of Oncology* 27: 2294–2299, 2016  
doi:10.1093/annonc/mdw422  
Published online 28 September 2016

## Validation of the German patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE™)

V. Hagelestein<sup>1</sup>, I. Ortland<sup>1</sup>, A. Wilmer<sup>1</sup>, S. A. Mitchell<sup>2</sup> & U. Jaehde<sup>1\*</sup>

<sup>1</sup>Institute of Pharmacy, Clinical Pharmacy, University of Bonn, Bonn, Germany; <sup>2</sup>Division of Cancer Control and Population Sciences, Outcomes Research Branch, National Cancer Institute, Rockville, USA

Received 10 April 2016; revised 30 July 2016; accepted 29 August 2016

**Background:** Integrating the patient's perspective has become an increasingly important component of adverse event reporting. The National Cancer Institute has developed a Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE™). This instrument has been translated into German and linguistically validated; however, its quantitative measurement properties have not been evaluated.

**Patients and methods:** A German language survey that included 31 PRO-CTCAE items, as well as the EORTC QLQ-C30 and the Oral Mucositis Daily Questionnaire (OMDQ), was distributed at 10 cancer treatment settings in Germany and

\*Correspondence to: Prof. Ulrich Jaehde, Institute of Pharmacy, Clinical Pharmacy, University of Bonn, An der Immenburg 4, D-53121 Bonn, Germany. Tel: +49-228-73-5252; Fax: +49-228-73-9757; E-mail: u.jaehde@uni-bonn.de

Austria. Item quality was assessed by analysis of acceptability and comprehensibility. Reliability was evaluated by using Cronbach's  $\alpha$  and validity by principal components analysis (PCA), multitrait-multimethod matrix (MTMM) and known groups validity techniques.

**Results:** Of 660 surveys distributed to the study centres, 271 were returned (return rate 41%), and data from 262 were available for analysis. Participants' median age was 59.7 years, and 69.5% of the patients were female. Analysis of item quality supported the comprehensibility of the 31 PRO-CTCAE items. Reliability was very good; Cronbach's  $\alpha$  correlation coefficients were  $>0.9$  for almost all item clusters. Construct validity of the PRO-CTCAE core item set was shown by identifying 10 conceptually meaningful item clusters via PCA. Moreover, construct validity was confirmed by the MTMM: monotrait-heteromethod comparison showed 100% high correlation, whereas heterotrait-monomethod comparison indicated 0% high correlation. Known groups validity was supported; PRO-CTCAE scores were significantly lower for those with impaired versus preserved health-related quality of life.

**Conclusion:** A set of 31 items drawn from the German PRO-CTCAE item library demonstrated favourable measurement properties. These findings add to the body of evidence that PRO-CTCAE provides a rigorous method to capture patient self-reports of symptomatic toxicity for use in cancer clinical trials.

**Key words:** cancer, German, patient-reported outcomes, PRO-CTCAE, questionnaire, validation

## introduction

Adverse event (AE) monitoring is an essential aspect of any cancer clinical trial, ensuring that study participants are not harmed by treatment, and allowing for conclusions to be made about a treatment regimen's safety and tolerability profile. For systematically grading and reporting treatment-related toxicity in cancer clinical trials, AEs are assessed with the Common Terminology Criteria for Adverse Events (CTCAE). The latest version 4.03 of the CTCAE released by the National Cancer Institute (NCI) contains 790 items including laboratory tests, clinical events, and symptom evaluation [1]. Typically, the grading of symptoms in clinical trials is carried out by research staff. As an underestimation of AE severity by healthcare professionals has been repeatedly demonstrated [2], there has been expanding interest to incorporate the patient's perspective via patient-reported outcomes (PRO) [3]. Covering different aspects of AE reports, both clinician-reported and patient-reported approaches are complementary [1, 2].

In 2010, the NCI initiated development of the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE<sup>TM</sup>) [4] and has evaluated the PRO-CTCAE item library using both qualitative and quantitative techniques [5]. The PRO-CTCAE item library contains 124 items evaluating 78 symptomatic toxicities, and the English language version has demonstrated generally favourable measurement properties [5].

The PRO-CTCAE item library has been translated into German and linguistically validated in a sample of patients undergoing haematopoietic stem cell transplantation [6]; however, its quantitative measurement properties (item quality, reliability, and validity) have not been evaluated.

The aim of this study was to assess item quality, reliability, and validity of a subset of items from the PRO-CTCAE German item library, thereby supporting the continued implementation of PRO-CTCAE in cancer clinical trials in German-speaking countries.

## patients and methods

### setting and sample

This study was conducted at 10 cancer treatment centres in Germany and Austria between October 2012 and April 2013. Participants had to be able to

read, write, and comprehend German. There were no eligibility limitations with respect to the type of treatment currently being received, disease site, or inpatient/outpatient setting. Participants who, in the opinion of the treating clinician, had clinically significant cognitive impairment were excluded from the study. The study received approval of the Ethics Committee of the University of Bonn.

A total of 660 questionnaires were distributed to the 10 participating centres. Totally, 271 participants were enrolled and provided their informed consent. They were provided with the study instruments and could complete study procedures independently by paper and pencil during their visit at the centre, or at home and return by surface mail.

### study measures

Of the 124 items contained in the PRO-CTCAE German language item library, 31 were selected for validation and defined as a 'core item set'. The selected items reflect 14 symptomatic toxicities and were selected a priori on the basis of their prevalence across cancer treatment types and disease sites [5, 7] and based on expert consultation. PRO-CTCAE attributes include frequency (F) (e.g. how often did you have nausea), severity (S) (e.g. what was the severity of your pain), and interference with daily activities (I) (e.g. how much did fatigue interfere with your usual or daily activities). Responses are provided on a five-point Likert scale. The recall period for PRO-CTCAE is the past 7 days. PRO-CTCAE symptomatic toxicities included in this study were those crossing disease sites and treatment modalities (shown in Table 1). Examples of the PRO-CTCAE items are shown in supplementary Table S1, available at *Annals of Oncology* online. For more information about PRO-CTCAE and for permission to use, visit <http://healthcaredelivery.cancer.gov/pro-ctcae/>

Participants also completed the Quality of Life Questionnaire of the European Organisation for Research and Treatment of Cancer (EORTC QLQ-C30) and the Oral Mucositis Daily Questionnaire (OMDQ). The EORTC QLQ-C30 [8] includes 30 items summarised to a global health status subscale, functional subscales, and symptom subscales. The OMDQ [9] contains eight items covering the subscales general well-being, mouth and throat pain, and diarrhoea. In the present study, the recall period of the OMDQ was extended from 24 h to the past 7 days, allowing comparability across PRO-CTCAE, EORTC QLQ-C30, and the OMDQ.

For comparability and only for the purpose of the MTMM analysis, PRO-CTCAE, EORTC QLQ-C30, and OMDQ scores were calculated according to the following equations [10].

**Table 1.** PRO-CTCAE item clusters

Item cluster	Number of items	Item dimensions
Anxiety and sadness	6	Frequency, severity, interference
Nausea and vomiting	4	Frequency, severity
Appetite loss	2	Severity, interference
Fatigue	2	Severity, interference
Pain	3	Frequency, severity, interference
Mucositis and xerostomia	4	Severity, interference
Dyspnoea	2	Severity, interference
Mental concentration	2	Severity, interference
Numbness and tingling	2	Severity, interference
Insomnia	2	Severity, interference
Constipation	1	Severity
Diarrhoea	1	Frequency

PRO-CTCAE, Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events.

PRO-CTCAE item cluster scores (clusters indicated in Table 1) and EORTC QLQ-C30 subscale scores were calculated by averaging the component items of each of the clusters or subscales according to

$$RS = \frac{I_1 + I_2 + \dots + I_n}{n} \quad (1)$$

where  $I_{1,2,\dots,n}$  are the item values and  $n$  is the number of items.

To ease the interpretation of MTMM models, two PRO-CTCAE item clusters (problems with mental concentration and anxiety/sadness) and the analogous EORTC QLQ-C30 functional constructs (cognitive function and emotional function) were linearly transformed to a 0–100 scale using the following formula:

$$\text{Score} = \left\{ 1 - \frac{RS - 1}{\text{Range}} \right\} \times 100 \quad (2)$$

All of the other PRO-CTCAE item cluster scores and the EORTC QLQ-C30 symptom subscale scores and the global health status score were linearly transformed to a 0–100 scale so that higher scores indicate worse symptoms and better quality of life, using the following formula:

$$\text{Score} = \left\{ \frac{RS - 1}{\text{Range}} \right\} \times 100 \quad (3)$$

In the above equations, range refers to the theoretical range of the response options. In the case of missing data, scores were calculated if a minimum of 50% of the values were present [10].

### item quality

A raw data analysis was carried out. Acceptability and comprehensibility were assessed by inspecting for illogical endorsement patterns and missing values. To evaluate item redundancy, weighted kappa with a weighting scheme according to Agresti [11] and Bowker's test ( $P < 0.05$ ) were used [12].

### reliability

Reliability reflects the consistency and stability of test scores and was determined by calculating Cronbach's alpha for each PRO-CTCAE item cluster indicated in Table 1.

### validity

Validity refers to the degree to which an instrument measures what it is supposed to measure. Construct validity was evaluated using principal component analysis (PCA) with varimax rotation, multitrait-multimethod matrix (MTMM) methods, and known groups validity techniques to assess whether PRO-CTCAE items measure the same constructs as corresponding items from the EORTC QLQ-C30 and OMDQ. The MTMM correlation coefficients were calculated as rank correlation coefficients according to Spearman's hypothesising that different instruments measuring similar constructs would show high correlations ( $r > 0.6$ ), whereas different instruments measuring different traits would not correlate strongly [13]. Known groups validity evaluated the criteria age ( $\leq 60$  and  $> 61$  years), diagnosis (breast cancer and all other types of cancer), and quality of life (EORTC QLQ-C30 item 30 score  $\leq 4$  and  $> 4$  [8]), using the Mann-Whitney  $U$  test ( $P < 0.05$ ).

Statistical analyses were carried out using SPSS® and Microsoft Excel®. Data were treated as interval-scaled data.  $P$  values were adjusted for multiplicity by Benjamini-Hochberg's step-up procedure (false discovery rate of 0.05) [14].

## results

### patients

The return rate was ~41%. Totally, 271 patient questionnaires were returned, and 262 were available for psychometric analyses.

The median age of the sample was 60.0 years (mean age 59.7 years, range 24–91); a majority of the sample was female (69.5%). The native language of almost all respondents was German. In the sample, 56.7% had a school leaving certificate as the highest educational degree. A majority of the sample had breast cancer (42.1%) due to a high proportion of participating breast cancer units, and ~80% had been treated on an outpatient basis. The EORTC QLQ-C30 item 30 score suggested that the sample had a generally moderate level of health-related quality of life (HRQL) of the population (mean 55.9, standard deviation 23.4).

### item quality

Inspection of the item-level endorsement patterns indicated that PRO-CTCAE, EORTC QLQ-C30, and OMDQ responses were not distributed normally. The frequency distributions of responses for all items are shown in supplementary table S2, available at *Annals of Oncology* online. Illogical endorsement patterns of 0.71% and missing values of 0.78% suggest that PRO-CTCAE is acceptable and comprehensible to German speakers. Redundancy assessment according to weighted kappa [15] revealed that mucositis, dyspnoea, and mental concentration (severity versus interference), as well as vomiting and pain (frequency versus severity), exhibited almost perfect agreement. Only one item showed only moderate agreement: sadness (interference versus frequency). However, Bowker's test identified significant differences between the attributes for all items except nausea, vomiting, and pain (frequency and severity). Detailed

results are shown in supplementary table S3, available at *Annals of Oncology* online.

### reliability

In general, Cronbach's alpha should be  $>0.8$ , but values  $>0.7$  are still acceptable [16]. All Cronbach's alpha values for the PRO-CTCAE item clusters were  $>0.9$  indicating very good reliability, except for the item clusters nausea and vomiting as well as appetite loss (good reliability with values  $>0.8$ ), and mucositis and xerostomia (values  $>0.7$ ); see supplementary table S4, available at *Annals of Oncology* online.

### validity

PCA with varimax rotation identified 10 components explaining 81.5% of the total variance (based on 231 questionnaires). Most of the items contributed strongly to only one of the extracted factors. Four PRO-CTCAE items, specifically fatigue (severity and interference), difficulty swallowing (severity) and dry mouth (severity), exhibited cross-loadings with other PRO-CTCAE items. Fatigue demonstrated significant cross-loadings with dyspnoea and difficulties with mental concentration, and difficulty swallowing and dry mouth cross-loaded with PRO-CTCAE symptom terms reflecting gastrointestinal symptoms (nausea and anorexia). Detailed PCA results are shown in supplementary figure S1, available at *Annals of Oncology* online.

The MTMM according to Campbell and Fiske (based on 210 questionnaires) confirmed construct validity [13]. To facilitate interpretation of the MTMM results, Table 2 presents a summary of the main results, showing the proportion of the correlation coefficients that were moderate (0.4–0.6) or high ( $>0.6$ ), as well as the proportion of non-significant correlation coefficients. For full results, see supplementary figure S2, available at *Annals of Oncology* online. All monotrait-heteromethod (MTHM) correlation coefficients were  $\geq 0.7$ . The mean of all MTHM correlation coefficients quantifies convergent validity. With a mean of 0.80 (median 0.83, range 0.70–0.86), convergent validity was high. Heterotrait-monomethod (HTMM) correlation coefficients, representing discriminant validity, were smaller compared with the MTHM. Of the PRO-CTCAE HTMM correlations and of the EORTC QLQ-C30 HTMM correlations 21% and 27%, respectively, were moderate, and none were high. The heterotrait-heteromethod (HTHM) correlation coefficients were the smallest. The proportion of non-significant coefficients was higher than in the HTMM.

The assessment of known groups validity indicated that PRO-CTCAE scores were statistically significantly different in the two groups that also differed with respect to HRQL. Those respondents with better HRQL (EORTC QLQ-C30 item 30 score  $>4$ ) showed statistically significantly lower PRO-CTCAE scores compared with those patients with lower HRQL (EORTC QLQ-C30 item 30 score  $\leq 4$ ). These differences in PRO-CTCAE scores were largest for fatigue (mean 1.41 versus 2.54) and least pronounced for diarrhoea (mean 0.51 versus 0.83) and mucositis (mean 0.45 versus 0.82), see Figure 1.

For the known groups analyses based on age ( $>60$  versus  $\leq 60$  years of age) or disease site (breast cancer versus all other tumour sites), there were no statistically significant differences in PRO-

**Table 2.** Proportion of correlations in the multitrait-multimethod matrix

Correlation type	High correlation (%)	Moderate correlation (%)	Non-significant correlation (%)
Monotrait-heteromethod	100	0	0
Heterotrait-monomethod PRO-CTCAE	0	21	15
Heterotrait-monomethod EORTC QLQ-C30	0	27	20
Heterotrait-heteromethod	0	15	26

CTCAE mean scores, see supplementary figure S3 and figure S4, available at *Annals of Oncology* online.

## discussion

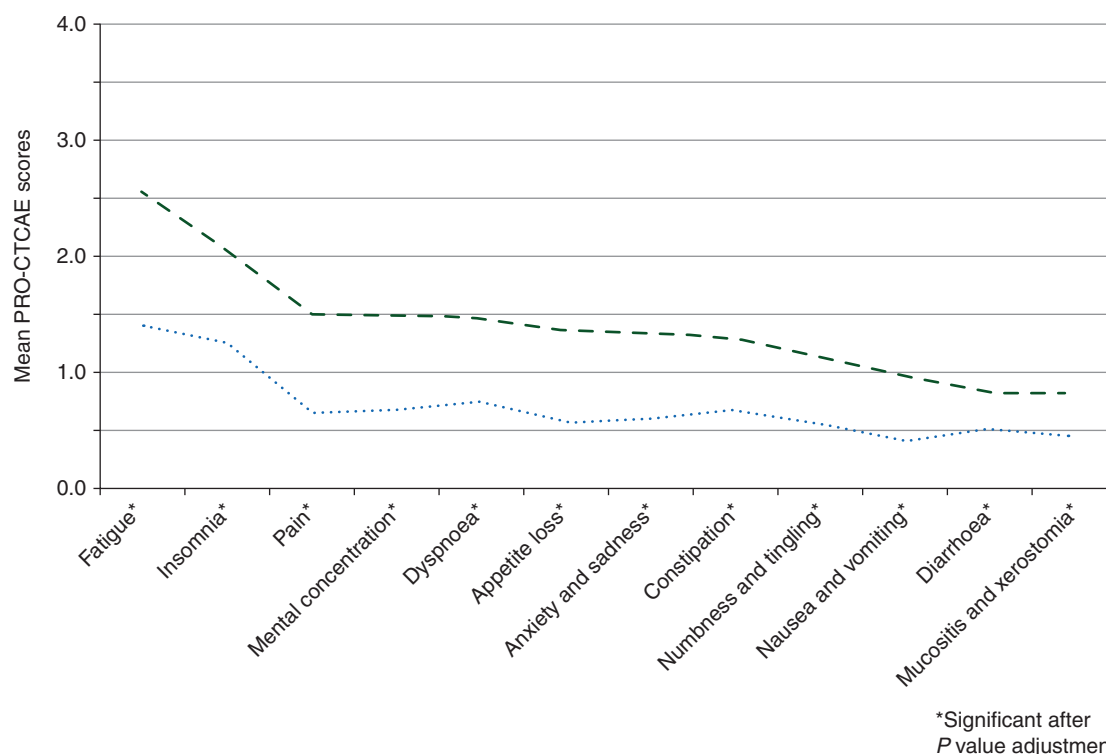
A German language core item set of the PRO-CTCAE was psychometrically tested using a variety of approaches, and the examined items demonstrated generally favourable measurement properties. These results are the first quantitative evidence that a subset of the PRO-CTCAE item library in the German language meets accepted criteria with respect to item quality, reliability, and validity for use as a patient-reported measure of symptomatic toxicity in cancer clinical trials.

The return rate of 41% was just acceptable since rates of 50%–70% are generally recommended. However, the sample size of 262 participants was sufficiently large to adequately power these analyses.

Inspection of the endorsement patterns did not reveal logically untenable response patterns (e.g. endorsing none for severity but a little for interference), and there were only few missing values for PRO-CTCAE. These observations suggest that PRO-CTCAE was well comprehended by German speakers.

While Bowker's test results suggested some redundancy between the frequency and severity attributes for three symptoms (nausea, vomiting, and pain), these observations should not be over-interpreted. Very few respondents were experiencing vomiting, and the low variability in this symptom may have produced an inflated level of concordance. The inclusion of multiple attributes is intended to improve the precision of PRO-CTCAE in capturing the latent construct (e.g. pain that is severe but infrequent). Replication of these results across multiple samples is needed in order to draw conclusions about whether these PRO-CTCAE items are understood by German speakers as reflecting distinct attributes of their symptom experience. Cronbach's alpha values revealed an excellent reliability of the PRO-CTCAE core item set.

Findings from the PCA, MTMM, and known groups analyses support the good validity of the German language PRO-CTCAE items and extend the linguistic validation of the PRO-CTCAE



**Figure 1.** Known groups validity for the criterion health-related quality of life (— EORTC QLQ-C30 item 30 score  $\leq 4$ ;  $n = 131$ ; ..... EORTC QLQ-C30 item 30 score  $> 4$ ;  $n = 114$ ),  $P$  value adjustment for multiplicity.

German as carried out by Kirsch et al. [6]. Validity is the most fundamental consideration in evaluating a new PRO measure [17]. The PCA yielded a set of 10 factors that were conceptually consistent with the intended dimensionality of the PRO-CTCAE item library [18]. For each PRO-CTCAE symptom term, the items reflecting attributes of frequency, severity, and interference generally loaded most strongly on a single factor. Interpretation of the very limited number of cross-loadings observed was conceptually consistent with the available knowledge of symptom clusters [19] in that fatigue severity and interference cross-loaded with problems with mental concentration and with dyspnoea, whereas the severity of swallowing problems and xerostomia each cross-loaded with other gastrointestinal symptoms.

Construct validity was also supported by the results of the MTMM analysis. Convergent validity as *first criterion* was fulfilled, with a mean MTHM correlation coefficient of 0.8 (median 0.83, range 0.70–0.86). Correlation coefficients approaching 1.0 indicate that a new measure may be too similar to the comparison measure and therefore redundant [20]. For the analysis of PRO-CTCAE, the mean correlation coefficient, however, clearly deviated from a perfect correlation so that redundancy is not given. Discriminant validity as *second criterion* was met since HTMM correlation coefficients were significantly smaller compared with the MTHM. The *third criterion*, HTHM correlation coefficients being the smallest, was also met. However, the *fourth criterion*, requiring similarity of the HTMM and HTHM pattern, could not be evaluated due to the complexity of the correlation matrix. In summary, the MTMM of PRO-CTCAE met three of the four criteria of Campbell and Fiske [13], supporting the construct validity of a PRO-CTCAE core item set in the German language. Additional evidence of the construct validity of PRO-

CTCAE German was demonstrated via known groups validity technique, which revealed that there were statistically significant differences in PRO-CTCAE scores between those respondents with impaired and preserved HRQL.

Several caveats should be considered in interpreting these findings. First, as only a subset of the PRO-CTCAE German language item library was evaluated, results may not generalise to the entire German language PRO-CTCAE item library. Second, due to the selection of the study centres, there was over-representation of women with breast cancer in the study sample, and the sample mean age was somewhat younger than the average age (69 years) of cancer onset in Germany [21]. Lastly, clinical data about current receipt of treatment were not collected. Thus, the degree to which the sample is sufficiently representative of a diverse population of cancer patients and survivors is not known. Therefore, additional studies are needed to extend the findings of this study.

## conclusion

In conclusion, a subset of items from the German language version of PRO-CTCAE was found to be reliable and valid suggesting that PRO-CTCAE German can be used for patient self-reporting of symptomatic AEs. Data derived from this instrument can supplement CTCAE toxicity evaluations in future clinical cancer trials in German-speaking countries.

## acknowledgements

We would like to express our gratitude to Klaus Meier and the German Society of Oncology Pharmacy (DGOP) for their great

cooperation and support for this study. Special thanks go to all of the patients who participated and to all of the staff in the participating centres for the distribution of the questionnaires: Evangelisches Waldkrankenhaus, Berlin; Dr med. C. Kurbacher, Bonn; Dr med. P. Schwindt, Bonn; Universitätsklinikum Freiburg; ABC Apotheke, Gelsenkirchen; Universitätsklinikum Heidelberg; St Elisabeth Krankenhaus, Köln; Heidekreisklinikum, Soltau; St Petrus Krankenhaus, Wuppertal and Kaiser Franz Josef Spital, Wien, Austria.

## funding

None declared.

## disclosure

The authors have declared no conflicts of interest.

## references

- Basch E, Bennett A, Pietanza MC. Use of patient-reported outcomes to improve the predictive accuracy of clinician-reported adverse events. *J Natl Cancer Inst* 2011; 103: 1808–1810.
- Basch E. Patient-reported outcomes in drug safety evaluation. *Ann Oncol* 2009; 20: 1905–1906.
- Trotti A, Colevas AD, Setser A, Basch E. Patient-reported outcomes and the evolution of adverse event reporting in oncology. *J Clin Oncol* 2007; 25: 5121–5127.
- Division of Cancer Control and Population Sciences, National Cancer Institute. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE™). <http://healthcaresdelivery.cancer.gov/pro-ctcae/> (10 April 2016, date last accessed).
- Dueck AC, Mendoza TR, Mitchell S 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.
- Kirsch M, Mitchell S, 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.
- Reeve BB, Mitchell S, Dueck AC et al. Recommended Patient-Reported Core Set of Symptoms to Measure in Adult Cancer Treatment Trials. *J Natl Cancer Inst* 2014; 106.
- Aaronson NK, Ahmedzai S, Bergman B et al. The European Organisation for Research and Treatment of Cancer QLQ-C30. A quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993; 85: 365–376.
- Stiff PJ, Erder H, Bensinger WI et al. Reliability and validity of a patient self-administered daily questionnaire to assess impact of oral mucositis (OM) on pain and daily functioning in patients undergoing autologous hematopoietic stem cell transplantation (HSCT) Bone Marrow Transplant 2006; 37: 393–401.
- Fayers PM, Aaronson NK, Bjordal K et al. on behalf of the European Organisation for Research and Treatment of Cancer Quality of Life Group. The EORTC QLQ-C30 Scoring Manual. 3rd Edition. Brussels, 2001. <http://www.eortc.be/qol> (10 April 2016, date last accessed).
- Agresti A. *Categorical Data Analysis*. 2nd Edition. New York, Wiley-Interscience, 2002.
- Bowker A. A test for symmetry in contingency tables. *J Am Stat Assoc* 1948; 43: 572–574.
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959; 56: 81–105.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995; 57: 289–300.
- Landis J, Richard KGG. The measurement of observer agreement for categorical data. *Biometrika* 1977; 33: 159–174.
- DeVellis RF. *Scale Development. Theory and Applications*. 3rd Edition. Thousand Oaks, Sage Publications, 2012.
- American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME). *Standards for Educational and Psychological Testing*. 2014 edition. Washington, American Educational Research Association 2014.
- 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.
- Dong ST, Butow PN, Costa DS, Lovell MR, Agar M. Symptom clusters in patients with advanced cancer: a systematic review of observational studies. *J Pain Symptom Manage* 2014; 48: 411–450.
- Coyne KS, Tubaro A, Brubaker L, Bavendam T. Development and validation of patient-reported outcomes measures for overactive bladder. A review of concepts. *Urology* 2006; 68(2 Suppl): 9–16.
- Robert Koch-Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. *Krebs in Deutschland 2009/2010*, 9th Edition. Berlin, 2013.