ELSEVIER

ORIGINAL ARTICLE

Thresholds for clinical importance were established to improve interpretation of the EORTC QLQ-C30 in clinical practice and research

Johannes M. Giesinger[a,*], Fanny L.C. Loth[a], Neil K. Aaronson[b], Juan I. Arraras[c], Giovanni Caocci[d], Fabio Efficace[e], Mogens Groenvold[f], Marieke van Leeuwen[b], Morten Aa. Petersen[g], John Ramage[h], Krzysztof A. Tomaszewski[i], Teresa Young[j], Bernhard Holzner[a], on behalf of the EORTC Quality of Life Group

[a]*University Hospital of Psychiatry II, Medical University of Innsbruck, Innsbruck, Austria*
[b]*Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, the Netherlands*
[c]*Medical Oncology Department, Hospital of Navarre, Pamplona, Spain*
[d]*Department of Medical Sciences and Public Health, University of Cagliari, Cagliari, Italy*
[e]*Health Outcomes Research Unit, Italian Group for Adult Hematologic Diseases (GIMEMA) Data Center, Rome, Italy*
[f]*The Research Unit, Department of Palliative Medicine, Bispebjerg & Frederiksberg Hospital, University of Copenhagen, Copenhagen, Denmark*
[g]*Department of Public Health, University of Copenhagen, Copenhagen, Denmark*
[h]*Department of Gastroenterology and Hepatology, Hampshire Hospitals NHS Foundation Trust, Aldermaston Road, Basingstoke, RG24 9NA, UK*
[i]*Faculty of Medicine and Health Sciences, Andrzej Frycz Modrzewski Krakow University, Krakow, Poland*
[j]*Lynda Jackson Macmillan Centre, East & North Hertfordshire NHS Trust incorporating Mount Vernon Cancer Centre, Northwood, UK*

Abstract

Objective: The objective of this study was to establish thresholds for clinical importance (TCIs) for the five functioning and nine symptom scales of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30).

Study Design and Setting: In this diagnostic study, cancer patients with mixed diagnoses and treatments completed the EORTC QLQ-C30 and anchored the questions in each domain in terms of their clinical importance. The anchor questions, concerned limitations in daily life, need for help/care, and the worries of the patient and his/her partner/family. These questions allowed categorizing patients into whether they exhibited a clinically important symptom/functional impairment for each scale and performing a receiver operating characteristic curve analysis to establish TCIs.

Results: Data from 498 patients from six European countries (mean age 60.4 years, 55.2% women) were analyzed. For the TCIs generated using the patient questionnaire data, the EORTC QLQ-C30 scales showed sensitivity values between 0.71 and 0.97 and specificity values between 0.62 and 0.92 (area under the curve above 0.80 for all scales).

Conclusion: This EORTC Quality of Life Group study provides TCIs for the functioning and symptom scales of the EORTC QLQ-C30. These TCIs can increase the interpretability of the questionnaire results and foster its use in daily clinical practice and in clinical research. © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Quality of life; Clinical oncology; Patient-reported outcome measures; EORTC QLQ-C30; Clinical significance; Thresholds; Cut-offs

1. Background

The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) [1] is one of the most widely used patient-reported outcome (PRO) questionnaires for assessing quality of life (QOL), functional health, and symptom burden in patients with cancer worldwide [2]. Historically, it has served as an outcome measure in cancer clinical trials,

What is new?

Key findings
- In our study, we have developed thresholds for clinical importance for the EORTC QLQ-C30, one of the most widely used questionnaires in cancer research, that assesses functional health, symptoms and global quality of life. We have found that the questionnaire provides excellent diagnostic accuracy for identification of clinically important symptoms and functional health impairments.

What this adds to what was known?
- These thresholds are the first that rely on criteria developed systematically including the views of both patients with cancer and health professionals.

What is the implication and what should change now?
- Interpretation of scores from the EORTC QLQ-C30 which are given on a 0-100 metric can be challenging. The thresholds facilitate interpretation of such scores and can be used e.g., for symptom screening in daily practice, or for calculating symptom prevalence rates from the EORTC QLQ-C30.

but in recent years, it is increasingly being used for patient monitoring in daily clinical practice [3].

However, interpreting the scores from this and other PRO questionnaires can be challenging for clinicians and researchers. A number of approaches have been used to facilitate interpretation of these scores, usually by making their meaning more tangible. Most efforts have focused on defining minimal important differences, which offer a reference for interpreting the clinical importance of differences in scores between groups or changes in scores over time [4−6]. Unfortunately, these approaches do not aid in the interpretation of absolute scores obtained from an individual patient at a single point in time. For such absolute scores, thresholds are necessary for converting metric PRO scores into symptom prevalence rates or to screen for clinically important symptoms and functional impairments in daily practice.

Previous studies have used different methodological approaches to develop thresholds that aid in the interpretation of absolute scores, including cutoff scores based on the wording of the response categories [7,8], identifying score distributions in specific patient populations [9,10], or referring to external criteria such as need for care [11−13]. However, the criteria for establishing thresholds in these studies were either created ad hoc or defined by experts only; in other words, they were not developed systematically including the views of both patient and health professionals.

To meet the need for well-defined, valid thresholds, the EORTC Quality of Life Group initiated a project to develop thresholds for clinical importance (TCIs) for the absolute scores on the QLQ-C30 based on external criteria reflecting the clinical importance of a health problem. Clinical importance is defined as any aspect of a health problem that makes it relevant for the clinical encounter. The specific external criteria (limitations in daily living, perceived need for treatment or care, and disease- or treatment-related worry) were established in the first part of the project, which included a Europe-wide mixed methods study of patients with cancer and health professionals [14].

Based on this groundwork, we sought to develop TCIs for the five functioning and nine symptom scales of the QLQ-C30. We expect that these TCIs will help improve interpretation of scores for the QLQ-C30 in both daily clinical practice and in research.

2. Methods

2.1. Sample

In this prospective study, we recruited a heterogeneous sample of patients with cancer from Austria, Italy, the Netherlands, Poland, Spain, and the United Kingdom. There was no restriction placed on cancer diagnosis, type of treatment or treatment status (on- or off-treatment), but patients were excluded if they had serious cognitive impairment that would prohibit them from completing questionnaires. For participation, patients had to be aged 18 years or older, speak the primary language of their country of residence, and provide written informed consent. Patients were assessed at a single time point with the QLQ-C30 and a questionnaire with anchor items on clinical importance.

For data collection, we used the Computer-based Health Evaluation System [15], which allows participants to complete case report forms and questionnaires electronically. Ethical approval was obtained from the local ethics committees according to national regulations.

2.2. The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30

The QLQ-C30 [1] comprises 30 items that can be summarized in 15 scales: Physical Functioning (PF), Role Functioning (RF), Social Functioning (SF), Emotional Functioning (EF), Cognitive Functioning (CF), Global QOL (QL), Fatigue (FA), Pain (PA), Nausea/Vomiting (NV), Appetite Loss (AP), Dyspnea (DY), Sleep Disturbances (SL), Diarrhea (DI), Constipation (CO), and Financial Impact of Disease (FI). For all scales, item scores are summed and linearly transformed into a scale ranging from 0 to 100. Higher scores on the functioning scales indicate higher levels of functioning, whereas higher scores on the symptom scales represent more symptom burden.

2.3. Assessment of clinical importance

In a previous mixed methods study [14], we identified various aspects that make a symptom or function clinically important—that is, deserving attention and discussion during the clinical encounter. The aspects most frequently mentioned by patients and health care providers were limitations in everyday life, the problem indicates/causes other problems, the emotional impact of a problem on the patient and/or family/partner, duration or frequency, change from normal, and the need for treatment/care.

Drawing on these aspects, we defined anchor items to be used as external criteria for developing TCIs for all QLQ-C30 scales (with the exception of the QL scale). The anchor items were defined via a consensus discussion within an EORTC Quality of Life Group meeting consisting of 23 members of different professional backgrounds (e.g., oncology, surgery, psychology, nursing, statistics). Before the meeting, a report was circulated among group members summarizing the findings of the aforementioned mixed methods study and detailing a number of considerations on the various aspects of clinical importance. These considerations helped to guide definition of the anchor items. Most importantly, we decided to use the same set of anchor items (with domain-specific wording) for all scales to ensure that the meaning of the TCIs was consistent across domains. For this reason, we did not consider associations between particular domains as possible anchors. We also excluded the "duration or frequency" and "change from normal" aspects because the QLQ-C30 measures symptom severity, which is conceptually different from these aspects.

We created the following anchor items (with domain-specific wording) to assess clinical importance separately for each domain:

- Limitations: "Has your SYMPTOM/PROBLEM limited your daily life?"
- Need for help: "Have you needed any help or care because of your SYMPTOM/PROBLEM?"
- Worries: "Has your SYMPTOM/PROBLEM caused you or your family/partner to worry?"

Three domains (RF, SF, and PA) of the QLQ-C30 ask about both severity and interference in daily life; thus, we did not include the limitations anchor for these domains.

In line with a previous pilot study [16], the standard QLQ-C30 response format was used for the anchor items (4-point Likert scale with responses of not at all, a little, quite a bit, and very much). A patient was categorized as a positive case (i.e., as having a clinically important problem/symptom) if (s)he selected "quite a bit" or "very much" on at least one of the anchor items.

2.4. Statistical analysis

Descriptive statistics for the QLQ-C30 were calculated separately for positive and negative cases. We also calculated group differences in absolute scores and the corresponding effect sizes (Cohen's d statistic based on the pooled standard deviation from positive and negative cases). To evaluate the consistency of the anchor items, we calculated the ranges and medians of the Spearman rank correlation coefficients across the QLQ-C30 scales.

We defined the TCIs using receiver operating characteristic (ROC) analysis, setting the QLQ-C30 scales as predictors and using the binary criterion of clinical importance for definition of positive/negative cases (e.g., a patient was a "positive" for pain if (s)he answered quite a bit, or very much to at least one of the three criteria related to pain). For each scale, we selected the threshold with the highest Youden's J statistic [17] (the sum of sensitivity and specificity minus one), or if the Youden's J of two adjacent thresholds differed by less than 0.05, we selected the threshold with higher sensitivity. If these thresholds had a sensitivity below 0.70, we selected the closest threshold providing a sensitivity above this value.

The areas under the curve (AUC) were calculated to determine the diagnostic accuracy of each scale—that is, the ability of each scale to discriminate between negative and positive cases. As per Hosmer and Lemeshow [18], AUCs of 0.70-0.80 indicated acceptable discrimination and AUCs above 0.80 indicated excellent discrimination.

Sensitivity analyses were conducted to investigate the variability in TCIs and diagnostic accuracy across 14 patient groups, defined by age (below/above 60y), sex, treatment intention (curative/palliative), treatment status (on/off), comorbidity (no/yes), and European region (Western Europe [Austria and the Netherlands], Southern Europe [Italy and Spain], Eastern Europe [Poland], and the UK). To evaluate diagnostic accuracy, we calculated group-specific AUCs for each QLQ-C30 scale. Invariance in TCIs was investigated separately for each scale using multivariate binary logistic regression models with the above grouping variables and the QLQ-C30 scale as independent variables, and the binary criterion for clinical importance as the dependent variable. In such models, a statistically significant ($P < 0.01$) grouping variable indicates that TCIs are different between groups. For significant grouping variables, we determined the TCIs for each subgroup using the aforementioned decision rule based on Youden's J and compared these TCIs with those of the total sample.

An a priori power analysis indicated that 500 patients (assuming 33% positive cases) would provide a power of 0.80 to demonstrate (with a two-sided alpha of 0.05) that the AUC is above 0.80 if the observed AUC is 0.865. The power analysis was conducted with PASS 11.0 [19].

3. Results

3.1. Patient characteristics

Between November 2016 and November 2018, we recruited 502 patients at eight centers in the six countries.

Table 1. Descriptive statistics for sociodemographic and clinical variables (*n* = 498)

Variable		Mean (SD)	Range
Age		60.4 (12.7)	19-87
		N	%
Sex	Women	272	55.2
	Men	221	44.8
	Missing data	5	
Diagnosis	Breast cancer	117	23.6
	Hematological malignancy	66	13.3
	Lung cancer	49	9.9
	Prostate cancer	48	9.7
	Colorectal cancer	42	8.5
	Head and neck cancer	39	7.9
	Lymphoma	37	7.5
	Gynecologic cancer	29	5.9
	Stomach cancer	12	2.4
	Brain cancer	10	2.0
	Other	46	9.3
	Missing data	3	
UICC stage[a]	I	61	16.1
	II	100	26.4
	III	79	20.8
	IV	139	36.7
	Missing data	16	
Comorbidity	No	272	59.0
	Yes	189	41.0
	Missing data	37	
Treatment intention	Curative	282	60.6
	Palliative	183	39.4
	Missing data	33	
Current treatment	No current treatment	115	23.3
	Current treatment	379	76.7
	Surgery[b]	135	35.6
	Chemotherapy[b]	232	61.2
	Radiotherapy[b]	127	33.5
	Other[b]	81	21.4
	Missing data	4	

Abbreviation: UICC, Union for International Cancer Control.
[a] Only reported for patients with solid tumors.
[b] More than one treatment is possible, so the percentages for the treatment modalities do not sum up to 100%.

Four patients with only partially completed questionnaires were excluded from further analysis. The remaining 498 patients had a mean age of 60.4 years (SD = 12.7) and 55.2% were women. The most frequent diagnoses were breast cancer (23.6%), hematological malignancies (13.3%), and lung cancer (9.9%). Most (76.7%) of patients were receiving treatment during the assessment, and 60.6%

Table 2. Number of patients with clinically important problems/symptoms (positive cases) and those without (negative cases) for each EORTC QLQ-C30 scale

EORTC QLQ-C30 scale	Cases	*N*	%
Physical Functioning (PF)	Positive	270	54.5
	Negative	225	45.5
Role Functioning (RF)[a]	Positive	122	24.6
	Negative	373	75.4
Social Functioning (SF)[a]	Positive	92	18.5
	Negative	404	81.5
Emotional Functioning (EF)	Positive	139	28.0
	Negative	357	72.0
Cognitive Functioning (CF)	Positive	57	11.5
	Negative	439	88.5
Fatigue (FA)	Positive	176	35.5
	Negative	320	64.5
Pain (PA)[a]	Positive	89	18.0
	Negative	405	82.0
Nausea/Vomiting (NV)	Positive	51	10.3
	Negative	443	89.7
Sleep Disturbances (SL)	Positive	95	19.2
	Negative	399	80.8
Dyspnea (DY)	Positive	90	18.2
	Negative	404	81.8
Appetite Loss (AP)	Positive	54	10.9
	Negative	440	89.1
Constipation (CO)	Positive	49	9.9
	Negative	444	90.1
Diarrhea (DI)	Positive	41	8.3
	Negative	453	91.7
Financial Impact (FI)	Positive	70	14.2
	Negative	423	85.8

Cases were defined using all external criteria (worries of patient/partner/family, limitations in daily life, and need for help/care).
[a] Please note that the anchor item for limitations in daily life was not used for Role Functioning, Social Functioning, and Pain because these scales already cover such limitations.

were receiving treatment with curative intention. Comorbid conditions were present in 41.0% of the patients. Further details are reported in Table 1.

3.2. Thresholds for clinical importance

Looking at the internal correlation between anchors, the correlations between the limitations and the need for help anchors ranged from 0.53 (SL) to 0.75 (PF) (median = 0.58), correlations between the worry and limitations anchors ranged from 0.47 (CO) to 0.64 (EF) (median = 0.56), and those between worry and need for help ranged from 0.46 (PF) to 0.70 (SF) (median = 0.65).

Applying our binary criterion for clinical importance, the proportion of positive cases ranged from 8.3% (DI) to

Table 3. Comparison of EORTC QLQ-C30 scores in patients with clinically important problems/symptoms (positive cases) and those without (negative cases)

EORTC QLQ-C30 scale	Negative cases		Positive cases		Difference	Effect size[a]
	Mean	SD	Mean	SD		
Functioning scales						
Physical Functioning (PF)	88.0	12.3	61.8	24.8	26.2	1.30
Role Functioning (RF)	75.6	24.8	32.5	25.3	43.1	1.73
Social Functioning (SF)	75.5	24.0	37.5	30.5	38.0	1.50
Emotional Functioning (EF)	83.3	15.0	52.4	22.5	30.9	1.77
Cognitive Functioning (CF)	84.4	19.0	54.8	27.1	29.6	1.48
Symptom scales						
Fatigue (FA)	25.8	18.2	66.7	22.2	−40.9	−2.08
Pain (PA)	15.0	20.2	62.5	27.0	−47.6	−2.20
Nausea/Vomiting (NV)	5.8	12.1	46.0	33.4	−40.2	−2.58
Sleep Disturbances (SL)	24.0	25.5	69.5	27.4	−45.5	−1.76
Dyspnea (DY)	14.8	21.0	56.7	25.2	−41.9	−1.92
Appetite Loss (AP)	16.5	23.3	64.2	25.8	−47.7	−2.02
Constipation (CO)	14.2	23.2	64.6	34.3	−50.4	−2.06
Diarrhea (DI)	10.0	20.5	50.4	30.8	−40.4	−1.88
Financial Impact (FI)	8.0	17.7	61.9	32.2	−53.9	−2.65

[a] Cohen's d based on pooled standard deviation.

54.5% (PF). The differences in scores on QLQ-C30 scales between negative and positive cases (in terms of effect size) were largest for FI (Cohen's d = 2.65) and NV (d = 2.58) and smallest for PF (d = 1.30) and CF (d = 1.48). The details are reported in Tables 2 and 3.

All scales had AUCs above 0.80, with the lower limits of their 95% confidence intervals being above 0.75. The highest AUCs were observed for FA (0.92) and FI (0.91) and the lowest for CF (0.81) and PF (0.82). The TCIs given in

Table 4 and Figure 1 had sensitivities ranging from 0.71 (SF) to 0.97 (DY), and specificities ranging from 0.62 (DY) to 0.92 (CO).

3.3. Sensitivity analysis

Estimating the diagnostic accuracy of each of the 14 QLQ-C30 scales within the 14 patient groups all 196 AUCs were ≥0.70 (acceptable discrimination), and 178 were at

Table 4. Results of the receiver operating characteristic (ROC) analysis and thresholds for clinical importance

QLQ-C30 scale	TCI	Sensitivity	95% CI	Specificity	95% CI	AUC	95% CI
Functioning scales							
Physical Functioning	83	0.79	0.74-0.84	0.70	0.64-0.76	0.82	0.78-0.85
Role Functioning	58	0.82	0.75-0.89	0.77	0.73-0.82	0.87	0.84-0.91
Social Functioning	58	0.71	0.62-0.81	0.77	0.73-0.81	0.82	0.77-0.87
Emotional Functioning	71	0.80	0.73-0.86	0.76	0.72-0.81	0.87	0.84-0.91
Cognitive Functioning	75	0.77	0.66-0.88	0.73	0.68-0.77	0.81	0.75-0.88
Symptom scales							
Fatigue	39	0.89	0.85-0.94	0.82	0.77-0.86	0.92	0.89-0.94
Pain	25	0.92	0.87-0.98	0.71	0.66-0.75	0.90	0.87-0.94
Nausea/Vomiting	8	0.86	0.76-0.96	0.75	0.71-0.79	0.87	0.81-0.93
Sleep Disturbances	50	0.79	0.71-0.87	0.84	0.81-0.88	0.87	0.82-0.91
Dyspnea	17	0.97	0.93-1.00	0.62	0.58-0.67	0.88	0.84-0.92
Appetite Loss	50	0.78	0.67-0.89	0.90	0.87-0.93	0.89	0.84-0.94
Constipation	50	0.73	0.61-0.86	0.92	0.89-0.94	0.86	0.79-0.92
Diarrhea	17	0.88	0.78-0.98	0.77	0.74-0.81	0.85	0.79-0.92
Financial Impact	17	0.91	0.85-0.98	0.81	0.77-0.84	0.91	0.86-0.95

Abbreviations: TCI, threshold for clinical importance; AUC, area under curve; CI, confidence interval.

Please note that for the functioning scales, scoring below the TCI indicates a clinically important problem, whereas for the symptom scales, scores above the TCI indicate such a problem. TCIs are between values that individual patients can obtain on the QLQ-C30 scales [20].
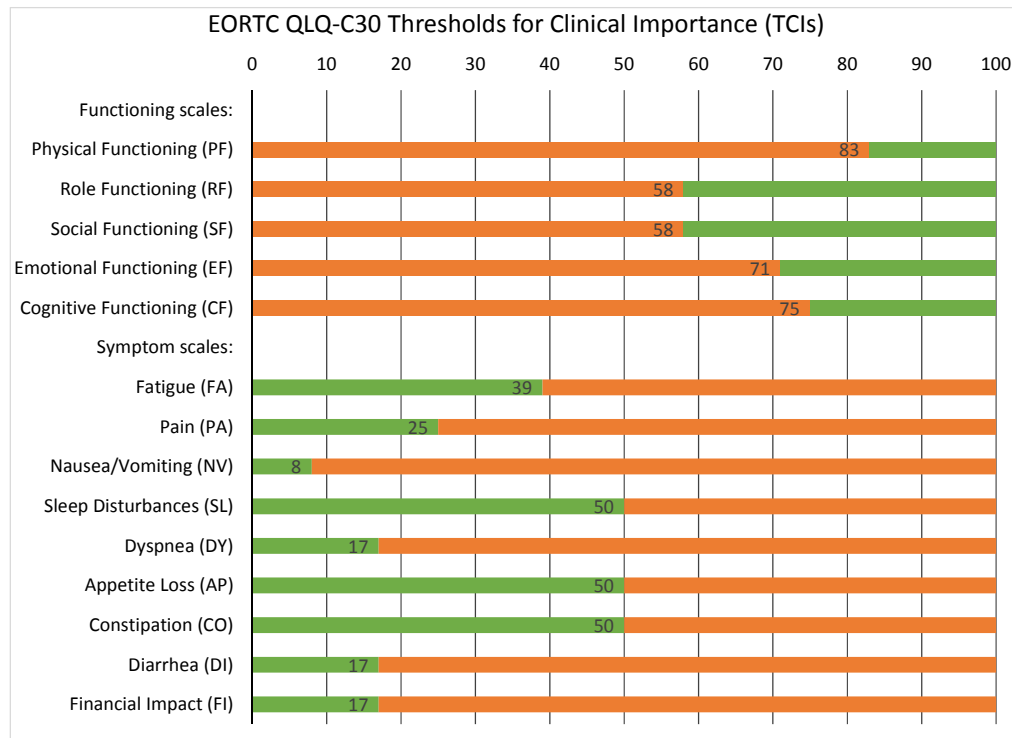
Fig. 1. Thresholds for clinical importance (TCIs) for the functioning and symptom scales of the EORTC QLQ-C30. TCIs are shown inside the bars. Patient scores in the orange range of the bar (i.e., below the TCI for functioning scales, or above the TCI for symptom scales) indicate clinically important problems or symptoms. EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

least 0.80 (excellent discrimination). The five lowest AUCs were for CF (0.70) and CO (0.73) in Southern Europe, PF in Western (0.74) and Eastern (0.76) Europe, and CF (0.76) among patients on-treatment.

The sensitivity analyses of the robustness of TCIs across patient groups indicated that only 5 of the 196 comparisons resulted in group-specific TCIs differing from the total sample TCIs. Specifically, for SF, a TCI of 42 provided higher sensitivity and specificity among patients below 60 years of age (sensitivity = 0.74, specificity = 0.87) and among patients from the UK (sensitivity = 0.82, specificity = 0.82) and Eastern Europe (sensitivity = 0.68, specificity = 0.83). For PA, a TCI of 42 was more discriminative among patients with comorbidities (sensitivity = 0.78, specificity = 0.90) and patients from Eastern Europe (sensitivity = 0.83, specificity = 0.97). Finally, for CO, the optimal TCI in Southern Europe was 17 (sensitivity = 0.64, specificity = 0.75).

4. Discussion

In this study, we established TCIs for all QLQ-C30 functioning and symptom scales and demonstrated their high diagnostic accuracy in identifying functional impairments and symptoms that limit a patient's daily life, cause worry to the patient and/or his/her partner or family, or require help or care. These thresholds will facilitate the use of the QLQ-C30 in daily clinical practice and in research, as

they allow for the conversion of absolute scores to prevalence rates, thereby making QLQ-C30 data more accessible to clinicians and researchers who find it challenging to interpret metric data on a 0-100 scale. The sensitivity analysis investigating the robustness of TCIs and diagnostic accuracy indicated that invariance can be assumed for the vast majority of patient groups, with only the language/country being a possible minor source of variability.

The robustness of the TCIs developed in this study is further supported by our previous study [16], wherein we developed thresholds for four QLQ-C30 scales (PF, EF, FA, PA) using three anchor items based on expert opinions (burden, limitations in daily activities, and need for help). Despite rather different anchor items, the established TCIs were identical for all four domains in that independent sample. As such, minor variations in external criteria might not have a substantial impact on TCIs.

Other studies have attempted to define thresholds for the QLQ-C30 using ad hoc criteria. Hickmann et al. [11] developed thresholds for six functioning and symptom domains using distress and supportive care needs separately as criteria. For the latter criterion, Hickmann et al. reported the same thresholds for RF, CF, and SF as in our study, whereas thresholds for PF and FA were higher. As for supportive care needs, the same thresholds were found for RF, EF, and SF, whereas the PF and FA thresholds indicated higher severity than in our study. Snyder et al. [12] reported thresholds for PF, RF, EF, QL, PA, and FA related to supportive care needs in two

separate samples that reflected similar or lower severity levels as compared with our results.

The QLQ-C30's coverage of important domains of functional health and key cancer symptoms [21], coupled with these TCIs, makes it especially suitable for symptom screening in daily clinical practice. The TCIs can also be integrated into electronic PRO data collection software to enable color-coding of graphical PRO result presentation [3], which is frequently used in such software to mark scores that require clinicians' attention.

In a research context, TCIs may aid in the interpretation of group-level data and allow for the calculation of prevalence rates. However, we should note that statistical analysis should not rely on these dichotomized scores because of their more limited statistical power [22] and ability to detect change over time.

A limitation of our study is that the various aspects of clinical importance were premised on the fact that clinical importance is related to the relevance of a symptom or problem to patient−clinician communication in clinical encounters. Although we believe this assumption to be valid, there might be other definitions of clinical importance linked to prognostic endpoints. Nevertheless, to the best of our knowledge, this is the first study defining clinical importance including systematically the perspectives of patients and health professionals.

In conclusion, this EORTC Quality of Life Group study has successfully defined TCIs for all functioning and symptoms scales of the QLQ-C30, thereby facilitating interpretation of the PRO scores derived from this measure. The thresholds might help make the QLQ-C30 scores more accessible and further increase the questionnaire's application in daily clinical practice and research.

CRediT authorship contribution statement

Johannes M. Giesinger: Conceptualization, Methodology, Formal analysis, Writing - original draft, Data curation, Writing - review & editing. **Fanny L.C. Loth:** Formal analysis, Writing - original draft, Data curation, Writing - review & editing. **Neil K. Aaronson:** Conceptualization, Methodology, Data curation, Writing - review & editing. **Juan I. Arraras:** Data curation, Writing - review & editing. **Giovanni Caocci:** Data curation, Writing - review & editing. **Fabio Efficace:** Data curation, Writing - review & editing. **Mogens Groenvold:** Writing - review & editing. **Marieke van Leeuwen:** Data curation, Writing - review & editing. **Morten Aa. Petersen:** Writing - review & editing. **John Ramage:** Data curation, Writing - review & editing. **Krzysztof A. Tomaszewski:** Data curation, Writing - review & editing. **Teresa Young:** Data curation, Writing - review & editing. **Bernhard Holzner:** Conceptualization, Methodology, Data curation, Writing - review & editing.

References

[1] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization 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: 365−76.

[2] Fayers P, Bottomley A, EORTC Quality of Life Group and Quality of Life Unit. Quality of life research within the EORTC-the EORTC QLQ-C30. European Organisation for Research and Treatment of Cancer. Eur J Cancer 2002;38(Suppl 4):S125−33.

[3] Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of patient-reported outcomes in routine medical care. Am Soc Clin Oncol Educ Book 2018;38:122−34.

[4] Musoro ZJ, Hamel JF, Ediebah DE, Cocks K, King MT, Groenvold M, et al. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. BMJ Open 2018;8(1):e019117.

[5] Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102−9.

[6] Cocks K, King MT, Velikova G, de Castro G Jr, Martyn St-James M, Fayers PM, et al. Evidence-based guidelines for interpreting change scores for the European organisation for the research and treatment of cancer quality of life questionnaire core 30. Eur J Cancer 2012; 48(11):1713−21.

[7] Johnsen AT, Petersen MA, Pedersen L, Groenvold M. Symptoms and problems in a nationally representative sample of advanced cancer patients. Palliat Med 2009;23(6):491−501.

[8] Oberguggenberger A, Hubalek M, Sztankay M, Meraner V, Beer B, Oberacher H, et al. Is the toxicity of adjuvant aromatase inhibitor therapy underestimated? Complementary information from patient-reported outcomes (PROs). Breast Cancer Res Treat 2011;128(2): 553−6.

[9] Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004;22:714−24.

[10] Scott N, Fayers P, Aaronson N, Bottomley A, De Graeff A, Groenvold M, et al. EORTC QLQ-C30 Reference Values. Brussels: EORTC Quality of Life Group; 2008.

[11] Hickmann AK, Hechtner M, Nadji-Ohl M, Janko M, Reuter AK, Kohlmann K, et al. Evaluating patients for psychosocial distress and supportive care needs based on health-related quality of life in primary brain tumors: a prospective multicenter analysis of patients with gliomas in an outpatient setting. J Neurooncol 2017;131(1): 135−51.

[12] Snyder CF, Blackford AL, Okuyama T, Akechi T, Yamashita H, Toyama T, et al. Using the EORTC-QLQ-C30 in clinical practice for patient management: identifying scores requiring a clinician's attention. Qual Life Res 2013;22:2685−91.

[13] Madsen UR, Groenvold M, Petersen MA, Johnsen AT. Comparing three different approaches to the measurement of needs concerning fatigue in patients with advanced cancer. Qual Life Res 2015;24: 2231−8.

[14] Giesinger JM, Aaronson NK, Arraras JI, Efficace F, Groenvold M, Kieffer JM, et al. A cross-cultural convergent parallel mixed methods study of what makes a cancer-related symptom or functional health problem clinically important. Psychooncology 2018; 27(2):548−55.

[15] Holzner B, Giesinger JM, Pinggera J, Zugal S, Schopf F, Oberguggenberger AS, et al. The computer-based health evaluation system (CHES): a software for electronic patient-reported outcome monitoring. BMC Med Inform Decis Mak 2012;12(1):126.

[16] Giesinger JM, Kuijpers W, Young T, Tomaszewski KA, Friend E, Zabernigg A, et al. Thresholds for clinical importance for four key

domains of the EORTC QLQ-C30: physical functioning, emotional functioning, fatigue and pain. Health Qual Life Outcomes 2016;14:87.

[17] Youden WJ. Index for rating diagnostic tests. Cancer 1950;3(1): 32–5.

[18] Hosmer D, Lemeshow S. Applied logistic regression. New York: John Wiley & Sons; 1989.

[19] NCSS-LCC. PASS - Power Analysis and Sample Size 2011: Kaysville (US). Available at https://www.ncss.com. Accessed November 7, 2019.

[20] Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, EORTC quality of life group. The EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: European Organisation for Research and Treatment of Cancer; 2001.

[21] Reeve BB, Mitchell SA, Dueck AC, Basch E, Cella D, Reilly CM, et al. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. J Natl Cancer Inst 2014;106:1–8.

[22] Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25:127–41.