2019-1

Establishing the European Norm for the health-related quality of life domains of the computer-adaptive test EORTC CAT Core

G. Liegl
Humboldt University of Berlin

M. A. Petersen
Bispebjerg Hospital

M. Groenvold
University of Copenhagen

See next page for additional authors

Authors
G. Liegl, M. A. Petersen, M. Groenvold, N. K. Aaronson, A. Costantini, P. M. Fayers, B. Holzner, C. D. Johnson, G. Kemmler, K. A. Tomaszewski, A. Waldmann, T. E. Young, Matthias S. F. Rose, and S. Nolte

Keywords
Computer-adaptive test, EORTC CAT Core, General population, Item response theory, Norm data, Normative data, Patient-reported outcomes, Quality of life, Self-report, Survey

European Journal of Cancer 107 (2019) 133–141

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Establishing the European Norm for the health-related quality of life domains of the computer-adaptive test EORTC CAT Core

G. Liegl [a], M.A. Petersen [b], M. Groenvold [b,c], N.K. Aaronson [d], A. Costantini [e], P.M. Fayers [f], B. Holzner [g], C.D. Johnson [h], G. Kemmler [g], K.A. Tomaszewski [i], A. Waldmann [j,k], T.E. Young [l], M. Rose [a,m], S. Nolte [a,n,*] on behalf of the EORTC Quality of Life Group

[a] Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health, Berlin, Germany
[b] Department of Palliative Medicine, Bispebjerg Hospital, Copenhagen, Denmark
[c] Department of Public Health, University of Copenhagen, Copenhagen, Denmark
[d] Division of Psychosocial Research & Epidemiology, The Netherlands Cancer Institute, Amsterdam, the Netherlands
[e] Psycho-Oncology Unit, Sant'Andrea Hospital Sapienza, University of Rome, Rome, Italy
[f] Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom
[g] Department of Psychiatry, Psychotherapy and Psychosomatics, Innsbruck Medical University, Innsbruck, Austria
[h] University of Southampton, Southampton, United Kingdom
[i] Health Outcomes Research Unit, Department of Gerontology, Geriatrics, and Social Work, Faculty of Education, Ignatianum Academy, Krakow, Poland
[j] Institute of Social Medicine and Epidemiology, University of Luebeck, Luebeck, Germany
[k] Ministry for Health and Consumer Protection, Hamburg Cancer Registry, Hamburg, Germany
[l] East & North Hertfordshire NHS Trust, Mount Vernon Cancer Centre, Northwood, Middlesex, United Kingdom
[m] Quantitative Health Sciences, Outcomes Measurement Science, University of Massachusetts Medical School, Worcester, MA, USA
[n] Population Health Strategic Research Centre, School of Health and Social Development, Deakin University, Burwood, VIC, Australia

Received 7 November 2018; accepted 10 November 2018
Available online 18 December 2018

DOI of original article: https://doi.org/10.1016/j.ejca.2018.11.024.

* *Corresponding author:* Sandra Nolte, Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany.
E-mail address: sandra.nolte@charite.de (S. Nolte).

https://doi.org/10.1016/j.ejca.2018.11.023

KEYWORDS
Quality of life;
Computer-adaptive test;
Item response theory;
EORTC CAT Core;
Self-report;
Patient-reported outcomes;
General population;
Norm data;
Normative data;
Survey

Abstract ***Objective:*** The computer-adaptive test (CAT) of the European Organisation for Research and Treatment of Cancer (EORTC), the EORTC CAT Core, assesses the same 15 domains as the EORTC QLQ-C30 health-related quality of life questionnaire but with increased precision, efficiency, measurement range and flexibility. CAT parameters for estimating scores have been established based on clinical data from cancer patients. This study aimed at establishing the European Norm for each CAT domain based on general population data.

Methods: We collected representative general population data across 11 European Union (EU) countries, Russia, Turkey, Canada and the United States ($n \geq 1000$/country; stratified by sex and age). We selected item subsets from each CAT domain for data collection (totalling 86 items). Differential item functioning (DIF) analyses were conducted to investigate cross-cultural measurement invariance. For each domain, means and standard deviations from the EU countries (weighted by country population, sex and age) were used to establish a T-metric with a European general population mean = 50 (standard deviation = 10).

Results: A total of 15,386 respondents completed the online survey (n = 11,343 from EU countries). EORTC CAT Core norm scores for all 15 countries were calculated. DIF had negligible impact on scoring. Domain-specific T-scores differed significantly across countries with small to medium effect sizes.

Conclusion: This study establishes the official European Norm for the EORTC CAT Core. The European CAT Norm can be used globally and allows for meaningful interpretation of scores. Furthermore, CAT scores can be compared with sex- and age-adjusted norm scores at a national level within each of the 15 countries.

1. Introduction

The assessment of health-related quality of life (HRQoL) using patient-reported outcome (PRO) measures has become increasingly important to evaluate, monitor and improve the quality of cancer care [1–3]. One of the most frequently used PRO measures for cancer patients is the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life core questionnaire QLQ-C30 [4,5]. It consists of 30 items covering five multi-item function scales (physical, role, social, cognitive and emotional function), three multi-item symptom scales (fatigue, nausea/vomiting and pain), six single-item scales assessing further aspects of HRQoL (dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties) and a scale on overall health/HRQoL.

The QLQ-C30 has been evaluated in many different cancer populations, and adequate psychometric properties were largely supported [6]. However, some studies found floor and ceiling effects [7,8]. Also, it is a well-known limitation of traditional HRQoL questionnaires, such as the QLQ-C30, that some questions might be irrelevant to the patient, which potentially increases respondent burden and results in less precise score estimates [9]. A promising solution to overcome such limitations is the use of computer-adaptive tests (CATs). CAT is methodologically based on item response theory (IRT). Using IRT, domain-specific item banks (i.e. lists of items measuring the same domain) can be calibrated on a common scale [9,10]. Fitting an IRT model provides item parameters that reflect the statistical relationship between an individual's response to a given item and his/her position on a domain scale. A major advantage of using IRT for scoring is that any item subset of an item bank can be used to estimate a person's domain score on the same continuous metric [11]. This enables CATs, in which HRQoL assessment is tailored to the individual, which increases measurement precision and range while reducing respondent burden, sample size requirements and study costs [10,12,13].

In 2006, the EORTC Quality of Life Group started developing the EORTC CAT Core, which is based on domain-specific item banks measuring the same dimensions as the QLQ-C30 [14,15]. In the item bank development process, different sources of information were collated, including literature reviews, qualitative input from various stakeholders and psychometric analyses of large international samples of cancer patients [14]. Item bank development for all domains was completed in 2016 [12,16–22]. To calibrate items of each bank, IRT models were estimated using data obtained from these clinical samples [22]. After item calibration, parameters are based on a z-score metric with a study population mean of 0 and standard deviation (SD) of 1. Such scores are 'arbitrary', hampering interpretation. Thus, the next step of item calibration is to link the CAT algorithm to general population norm data to simplify score interpretation. In this final step, it has become common practice to transform scores to a T-score metric with a general population mean of 50 (SD = 10).

The EORTC CAT Core development did not include this final step of transforming scores to T-scores. Therefore, for a more meaningful and sensible score interpretation, this study aimed at collecting representative data of the European general population to transform the current scoring to a T-score metric. Any score obtained via the EORTC CAT Core can then be interpreted in relation to this European mean. In addition, CAT norm scores are established for sex-, age- and country-specific subpopulations.

2. Methods

2.1. Sampling

To collect general population data to establish the 'European CAT Norm', we subcontracted GfK SE (http://www.gfk.com/). Data were collected via online surveys in March/April 2017 in 11 European Union (EU) countries (Austria, Denmark, France, Germany, Hungary, Italy, the Netherlands, Poland, Spain, Sweden and the United Kingdom), Russia, Turkey, Canada and the United States for comparative purposes. We stratified data collection by sex and age groups (18–39, 40–49, 50–59, 60–69, 70+ years), with a target sample size of each sex × age × country subgroup of n = 100, leading to an anticipated sample size of n = 1000/country. Assuming a T-scale with a mean = 50 (SD = 10), this sample size allows for estimating the population T-score mean of each country (n = 1000) within ±0.6 T-scores (95% confidence interval), which was considered to be sufficiently precise. Moreover, this sampling design is sufficient to investigate differential item functioning (DIF) using logistic regression analysis [23].

2.2. Selection of items

The full item pool of the EORTC CAT Core consists of 14 item banks for the function and symptom-related HRQoL domains in the QLQ-C30 plus two global items forming the 15th scale for *overall HRQoL*. The number of items per bank ranges between 7 (*appetite loss*) and 34 items (*fatigue, cognitive function*). In total, 260 items are available for CAT assessment. For establishing CAT norm values, 86 items were selected, consisting of all QLQ-C30 items plus four additional items per domain (56 items). The selection of the 56 items was based on high measurement precision and adequate range of measurement as indicated by the items' psychometric properties and content validity considerations, that is, all aspects of a given domain had to be covered.

2.3. Differential item functioning

DIF analyses are used to evaluate whether items measure the same underlying construct in different subpopulations [24], a central requirement for establishing a common norm across subpopulations. We investigated DIF regarding country, sex and age groups using ordinal logistic regression [23,25]. A regression was modelled for each item, with the item response as the dependent variable and the IRT-based domain score as the independent variable. If adding the grouping variable of interest (country/sex/age) to this model as an independent variable leads to a change in the Nagelkerke R^2 coefficient $\geq$.035, this indicates potentially relevant DIF [20,26]. If DIF was identified, we evaluated its practical importance by calculating the standardised mean difference (SMD) between scores derived from all available items of a domain versus scores derived from a reduced item set, that is, excluding items showing DIF. If SMD was $\leq$.2 (small effect size [ES] [27]), the practical impact of DIF was considered to be negligible and affected items were kept.

2.4. Establishing the European CAT norm

For establishing the 'European CAT Norm', we used general population data from the 11 EU countries. In a first step, we scored the data based on the previously established 'arbitrary' IRT-based z-score metric and calculated means/SDs for each CAT domain. To correct for over- or under-representation of subgroups, we weighted scores by country population size, sex and age distribution, with the youngest age group further divided into 18–29 and 30–39 years. Individual weighting factors were calculated for each country × sex × age group based on general population distribution statistics for 2015 [28] using the formula:

Weighting factor = percentage of subgroup in population/ percentage of subgroup in sample.

After estimating weighted means and SDs, scores were transformed to a T-score metric using linear transformation to establish the 'European CAT Norm' with mean = 50 and SD = 10 using the formula:

$$\text{T-score} = 10 * (\text{z-score} - \text{z-score mean})/\text{z-score SD} + \mathbf{50}.$$

Using these formulas, we calculated European norm scores (means, SDs) for each CAT domain overall and by sex and age. Furthermore, to establish norm scores for each of the 15 countries, national T-score means/SDs were calculated using country-specific sex- and age-weighting factors.

2.5. Determining the extent of subgroup differences

For each CAT domain, we investigated T-score differences between countries, sex, age groups and dichotomised educational levels (less than postcompulsory education and at least some postcompulsory education). We conducted covariance analyses for each independent variable, entering the remaining three variables as covariates. When used as a covariate, *age* was entered as continuous variable. Statistical significance was implied by P value <.01. We interpreted partial eta^2 values of .01 ($\approx R^2 = 1\%$), .06 ($\approx R^2 = 6\%$) and .14 ($\approx R^2 = 14\%$) as small, medium and large ES, respectively [27].

For analysing DIF, R 3.1.2 was applied using the package *lordif* version 0.3-3 [25,29]. For all other statistical analyses, we used IBM SPSS Statistics®, version 22 [30].

3. Results

3.1. Sample

The total sample size was N = 15,386 for the full sample and n = 11,343 for the EU sample (Table 1). Further details on sampling and sociodemographic data are provided elsewhere [35].

Table 1
EORTC CAT Core general population norm data sample characteristics.

Sociodemographic variable	Full sample (15 countries; N = 15,386) n (%)	Norm sample (11 EU countries; n = 11,343) n (%)
Age, years		
18–29	1177 (7.6)	883 (7.8)
30–39	1902 (12.4)	1370 (12.1)
40–49	3049 (19.8)	2248 (19.8)
50–59	3059 (19.9)	2253 (19.9)
60–69	3138 (20.4)	2337 (20.6)
70+	3061 (19.9)	2252 (19.9)
Sex		
Female	7650 (49.7)	5623 (49.6)
Male	7736 (50.3)	5720 (50.4)
Education		
Less than compulsory education	183 (1.2)	95 (.8)
Compulsory (left school at the minimum school leaving age)	1509 (9.8)	897 (7.9)
Some postcompulsory (some school after reaching school leaving age without reaching university entrance qualifications [e.g. A levels])	2050 (13.3)	1954 (17.2)
Postcompulsory below university (e.g. reaching A levels)	4405 (28.6)	3408 (30.0)
University degree (bachelor's or equivalent level)	3716 (24.2)	2689 (23.7)
Postgraduate degree (master's, doctorate or equivalent level)	3337 (21.7)	2131 (18.8)
Prefer not to answer	186 (1.2)	169 (1.5)
Country		
Austria	1002 (6.5)	1002 (8.8)
Denmark	1003 (6.5)	1003 (8.8)
France	1001 (6.5)	1001 (8.8)
Germany	1006 (6.5)	1006 (8.9)
Hungary	1053 (6.8)	1053 (9.3)
Italy	1036 (6.7)	1036 (9.1)
The Netherlands	1000 (6.5)	1000 (8.8)
Poland	1024 (6.7)	1024 (9.0)
Spain	1165 (7.6)	1165 (10.3)
Sweden	1027(6.7)	1027 (9.1)
The United Kingdom	1026 (6.7)	1026 (9.0)
Russia	1007 (6.5)	–
Turkey	1023 (6.6)	–
Canada	1004 (6.5)	–
The United States	1009 (6.6)	–

3.2. Psychometric properties

DIF by country and age was detected in two items of the *physical function* (PF) scale. As the ES of excluding versus including the DIF-items when estimating PF scores in the full sample were very small (SMD = .01), all items of the PF scale were retained for further analyses. For the Hungarian data, one item of the *sleep problems* scale (SL4) had to be excluded from further analyses due to a translation error.

3.3. European CAT norm scores

Table 2 presents domain-specific *z-score* distributions in the EU countries, weighted by country population size, sex and age. In all CAT domains, except for *dyspnoea*, all mean scores indicate better HRQoL in the general population compared with scores in the original cancer populations (which by way of model estimation has mean = 0).

In Table 3, the final European CAT Norm *T-scores* (means and SDs) for each domain are reported for the EU sample overall (by definition with mean = 50) and by sex and age. Covariance analyses indicated higher HRQoL in men than in women ($P < .001$) in all domains but ES were small, ranging from eta^2 = .001 for *diarrhoea* and *nausea/vomiting* to eta^2 = .020 and .022 for *emotional function* and *PF*, respectively. Age also had a statistically significant effect on HRQoL ($P < .001$ for each domain); however, the relational patterns were inconsistent across CAT scales, and ES were small for most domains, ranging from eta^2 = .002 for *pain* to eta^2 = .020 for *fatigue*. In three domains, a larger and relatively linear age effect was found: *emotional function* (eta^2 = .047) and *nausea/vomiting* (eta^2 = .034) improved while *PF* scores (eta^2 = .066) worsened with age.

Except for *diarrhoea*, scores were also significantly associated with educational level ($P < .01$) with higher educated individuals reporting better HRQoL scores (data not shown). However, these ES were very small (all eta$^2 \leq .015$).

Table 4 presents CAT T-scores for all 15 countries. Within EU countries, domain scores in Poland indicated relatively low HRQoL, while scores were comparatively high in Austria and the Netherlands. In the United States and Canada, score distributions were relatively

Table 2
z-score distribution in the EU countries (n = 11,343) of the EORTC CAT Core domain scales.

EORTC CAT Core domain	Number of items	Mean	SD	Median	Minimum	Maximum	Skewness
Physical function	9	**.2327**	**.81453**	.1960	−2.69	1.51	−.312
Role function	6	**.4491**	**.82772**	.6990	−2.19	1.18	−.912
Emotional function	8	**.0813**	**.98832**	.0982	−2.76	1.38	−.394
Cognitive function	6	**.2466**	**.97733**	.3490	−2.98	1.24	−.777
Social function	6	**.3203**	**.93293**	1.0150	−2.62	1.02	−1.097
Fatigue	7	**−.1346**	**.92938**	−.1080	−1.45	2.42	.375
Nausea/vomiting	6	**−1.8961**	**.78304**	−2.2130	−2.21	2.14	2.748
Pain	6	**−.8193**	**1.20929**	−.9480	−2.03	2.43	.612
Dyspnoea	5	**.1787**	**.94265**	.2050	−.68	2.91	.674
Sleep problems	5	**−.3120**	**1.08615**	−.2320	−1.68	2.36	.461
Appetite loss	5	**−.2141**	**.77220**	−.6760	−.68	2.61	1.453
Constipation	5	**−.4222**	**.74434**	−.6020	−1.07	2.21	.928
Diarrhoea	5	**−.4564**	**.67354**	−.8460	−.85	2.20	1.535
Financial difficulties	5	**−.3617**	**.78364**	−.8310	−.83	2.38	1.497
Overall HRQoL	2	**.1163**	**.93597**	.0390	−2.58	1.91	−.089

EORTC CAT = The computer-adaptive test of the European Organisation for Research and Treatment of Cancer; EU = European Union; HRQoL = health-related quality of life. Bold values = The z-score mean and SD values were used in the formula described in section 2.4 for transforming the z-scores to T-scores.

similar to the EU countries. In contrast, mean scores of the Russian and Turkish general populations indicated worse HRQoL in most CAT domains compared to the EU average. In the covariance analyses, T-scores differed significantly across countries in all CAT domains ($P < .001$). However, ES were small for each domain ($eta^2 < .06$).

4. Discussion

The EORTC CAT Core is the first disease-specific computer-adaptive PRO assessment system developed across different countries for measuring a wide range of HRQoL aspects relevant to cancer patients. In an extensive development and psychometric evaluation process, the EORTC CAT has been proven to be a more precise, efficient and flexible measurement instrument compared to the traditional QLQ-C30 static questionnaire [22].

This study established the official 'European CAT Norm' based on general population data from 11 EU countries for a more meaningful and sensible interpretation of EORTC CAT scores. The domain-specific means and SDs presented herein are now implemented in the EORTC CAT scoring algorithm using a standardised scale centred to the European general population with a mean of 50 (SD = 10). Additionally, we present norm scores per country and for sex- and age-specific subgroups. This allows for a meaningful and detailed interpretation of cancer patients' scores.

Similar to our findings presented in the EORTC QLQ-C30 norm data paper [35], some group differences were observed. For example, men tended to score somewhat better than women, which is consistent with other QLQ-C30 norm data studies [31]. Furthermore, some observed age differences were counterintuitive, with the youngest participants showing lowest/worst scores in some function scales, which has also been observed by others in the application of item banks [32]. Due to these group differences, we recommend the use of sex- and age-matched norm data for the most sensible and meaningful score interpretation of data from cancer patients obtained via the EORTC CAT Core.

The observed differences between countries need to be taken at face value. It is conceivable that these differences reflect 'true' differences in HRQoL; however, it is also possible that some of these differences either reflect differences due to slightly different meanings between language versions or they reflect cultural differences, for example, in terms of culture-related health perceptions, expectations or response styles. Given the vast experience with questionnaire translation and cultural adaptation of items at the EORTC headquarters and findings in the literature showing language-related DIF to be negligible [33], we assume the observed differences to be 'true' country differences in HRQoL until further evidence is found that supports or refutes our hypothesis. Furthermore, our tests of country-DIF as presented herein show minimal impact of DIF, providing sufficient support for our assumption of true intercountry differences.

Our study has some limitations. First, it is not clear whether online panels are truly representative of the general population despite panel research companies claiming they are. As an increasingly large proportion of people have access to the Internet, the potential problem of representativeness is getting smaller but still remains, especially in countries such as Turkey where GfK had to carry out telephone interviews to achieve sampling quotas (for details see [35]). Our data suggest representativeness regarding most sample characteristics except for educational status; however, when testing for the influence of educational level, we found that the practical consequences were negligible. These findings

Table 3
European norm T-scores (based on 11 EU countries) for each EORTC CAT Core domain: mean scores (*M*) and standard deviations (SDs) by sex and age groups[a].

Domain	Full sample, *M* (SD)			18–29 years, *M* (SD)			30–39 years, *M* (SD)			40–49 years, *M* (SD)			50–59 years, *M* (SD)			60–69 years, *M* (SD)			70+ years, *M* (SD)		
	All	Male	Female	All	Male	Female	All	Male	Female	All	Male	Female	All	Male	Female	All	Male	Female	All	Male	Female
Physical function	50.00	51.52	48.56	52.83	53.75	51.86	52.54	53.82	51.23	51.47	52.95	49.99	49.34	50.89	47.83	47.40	48.60	46.30	45.36	47.24	44.02
	(10.00)	(10.23)	(9.56)	(9.53)	(10.38)	(8.44)	(9.83)	(9.93)	(9.55)	(10.00)	(9.93)	(9.87)	(9.74)	(9.87)	(9.38)	(9.26)	(9.60)	(8.79)	(9.29)	(9.62)	(8.81)
Role function	50.00	50.44	49.59	51.20	50.12	52.32	50.91	51.16	50.66	50.78	51.26	50.29	49.78	50.59	48.99	49.48	50.13	48.89	47.47	49.16	46.27
	(10.00)	(9.83)	(10.14)	(9.74)	(10.56)	(8.66)	(9.73)	(9.54)	(9.91)	(10.09)	(9.84)	(10.32)	(10.17)	(9.81)	(10.45)	(9.96)	(9.50)	(10.34)	(9.86)	(9.18)	(10.14)
Emotional function	50.00	51.19	48.88	48.50	49.86	47.09	48.21	48.98	47.42	49.19	50.40	47.96	49.71	51.05	48.40	52.16	53.40	51.02	52.92	54.81	51.57
	(10.00)	(9.84)	(10.02)	(10.84)	(10.93)	(10.55)	(10.36)	(10.36)	(10.32)	(10.15)	(9.71)	(10.43)	(9.60)	(9.54)	(9.49)	(9.08)	(8.65)	(9.33)	(8.48)	(7.46)	(8.90)
Cognitive function	50.00	50.65	49.39	48.33	48.73	47.91	49.57	50.24	48.89	49.82	50.78	48.85	50.26	51.40	49.14	51.56	51.84	51.31	51.11	51.80	50.61
	(10.00)	(10.04)	(9.93)	(11.11)	(11.57)	(10.61)	(10.93)	(10.83)	(10.99)	(10.35)	(10.08)	(10.53)	(9.78)	(9.66)	(9.78)	(8.45)	(8.41)	(8.49)	(8.15)	(7.74)	(8.40)
Social function	50.00	50.38	49.64	49.70	49.74	49.66	49.10	49.41	48.77	48.98	49.46	48.51	49.75	50.62	48.91	51.19	51.68	50.73	51.56	52.08	51.19
	(10.00)	(9.84)	(10.14)	(10.44)	(10.58)	(10.30)	(10.60)	(10.53)	(10.67)	(10.55)	(10.20)	(10.87)	(10.06)	(9.67)	(10.36)	(9.21)	(8.84)	(9.53)	(8.44)	(7.95)	(8.75)
Fatigue	50.00	48.84	51.10	50.89	49.96	51.87	51.18	50.06	52.33	50.67	49.21	52.15	50.09	48.88	51.28	48.31	47.42	49.13	48.38	46.54	49.69
	(10.00)	(9.81)	(10.06)	(9.61)	(9.71)	(9.41)	(10.15)	(9.67)	(10.50)	(10.35)	(9.89)	(10.60)	(10.06)	(9.98)	(10.01)	(9.84)	(9.63)	(9.97)	(9.64)	(9.46)	(9.56)
Nausea/vomiting	50.00	49.96	50.04	52.11	53.27	50.90	52.11	52.28	51.94	50.06	49.42	50.71	49.07	48.48	49.65	48.16	47.57	48.70	47.79	47.00	48.36
	(10.00)	(10.27)	(9.74)	(12.70)	(14.14)	(10.86)	(11.88)	(12.40)	(11.33)	(9.97)	(9.16)	(10.69)	(8.31)	(7.59)	(8.93)	(7.09)	(6.00)	(7.93)	(6.36)	(4.33)	(7.43)
Pain	50.00	49.35	50.61	49.17	49.25	49.09	49.87	49.79	49.95	50.06	49.41	50.71	50.80	49.79	51.80	49.87	49.30	50.39	50.39	48.44	51.77
	(10.00)	(9.74)	(10.20)	(9.83)	(10.01)	(9.63)	(10.15)	(9.96)	(10.35)	(10.21)	(9.84)	(10.52)	(10.15)	(9.85)	(10.35)	(9.92)	(9.42)	(10.33)	(9.69)	(9.06)	(9.88)
Dyspnoea	50.00	49.54	50.44	49.12	49.52	48.70	49.50	49.00	50.01	49.18	48.50	49.88	50.01	49.40	50.62	50.53	50.24	50.80	51.98	51.00	52.68
	(10.00)	(10.01)	(9.97)	(9.67)	(10.39)	(8.84)	(9.98)	(10.09)	(9.85)	(9.87)	(9.55)	(10.14)	(9.90)	(9.69)	(10.08)	(9.98)	(9.90)	(10.05)	(10.38)	(10.18)	(10.47)
Sleep problems	50.00	48.97	50.97	48.52	47.40	49.68	50.31	49.77	50.86	50.56	49.76	51.36	51.27	49.70	52.81	49.89	48.88	50.83	49.75	48.64	50.54
	(10.00)	(9.82)	(10.08)	(9.92)	(9.86)	(9.85)	(10.25)	(10.20)	(10.28)	(10.23)	(9.95)	(10.45)	(10.45)	(10.14)	(10.52)	(9.64)	(9.31)	(9.85)	(9.17)	(8.86)	(9.31)
Appetite loss	50.00	49.69	50.30	51.95	52.52	51.34	51.71	51.63	51.78	49.84	48.96	50.73	49.19	48.52	49.85	48.41	47.83	48.95	48.28	47.24	49.02
	(10.00)	(9.84)	(10.14)	(11.24)	(11.78)	(10.61)	(11.05)	(10.91)	(11.20)	(9.85)	(8.92)	(10.64)	(9.23)	(8.85)	(9.55)	(8.72)	(7.98)	(9.33)	(8.47)	(7.53)	(9.01)
Constipation	50.00	49.62	50.36	51.40	51.34	51.46	51.44	50.69	52.22	49.72	49.03	50.42	49.41	48.56	50.24	48.59	48.38	48.79	48.97	48.98	48.95
	(10.00)	(9.70)	(10.26)	(10.69)	(10.95)	(10.41)	(10.78)	(10.51)	(11.01)	(10.01)	(9.27)	(10.66)	(9.50)	(8.93)	(9.97)	(9.01)	(8.62)	(9.35)	(9.23)	(8.64)	(9.62)
Diarrhoea	50.00	50.53	49.50	51.21	52.64	49.71	52.01	52.70	51.31	50.15	50.72	49.56	49.58	49.60	49.55	48.22	48.41	48.05	48.36	47.74	48.80
	(10.00)	(10.34)	(9.64)	(11.12)	(12.00)	(9.90)	(11.11)	(11.39)	(10.78)	(10.10)	(10.40)	(9.76)	(9.59)	(9.73)	(9.46)	(8.21)	(8.14)	(8.27)	(8.41)	(7.22)	(9.14)
Financial difficulties	50.00	49.90	50.10	50.25	51.32	49.13	51.07	50.96	51.20	50.83	50.22	51.45	50.24	49.45	51.01	48.92	48.56	49.24	48.45	47.92	48.82
	(10.00)	(9.84)	(10.15)	(10.52)	(10.96)	(9.92)	(10.86)	(10.78)	(10.94)	(10.66)	(10.14)	(11.13)	(9.86)	(9.19)	(10.42)	(9.07)	(8.71)	(9.38)	(8.25)	(7.51)	(8.72)
Overall HRQoL	50.00	50.83	49.21	51.11	52.28	49.89	49.66	50.46	48.85	49.26	49.95	48.55	49.13	49.84	48.44	50.17	50.44	49.92	50.46	51.77	49.53
	(10.00)	(10.01)	(9.93)	(10.05)	(10.58)	(9.32)	(9.87)	(9.82)	(9.86)	(9.96)	(9.57)	(10.30)	(10.33)	(10.35)	(10.26)	(9.96)	(9.59)	(10.29)	(9.64)	(9.55)	(9.60)

EORTC CAT = The computer-adaptive test of the European Organisation for Research and Treatment of Cancer; EU = European Union; HRQoL = health-related quality of life.
[a] The European general population has a mean T-score of 50 (SD = 10).

Table 4
Country-specific EORTC CAT Core T-score[a] distributions.

Domain	AUT, M (SD)	CAN, M (SD)	DNK, M (SD)	FRA, M (SD)	DEU, M (SD)	HUN, M (SD)	ITA, M (SD)	NLD, M (SD)	POL, M (SD)	RUS, M (SD)	ESP, M (SD)	SWE, M (SD)	TUR, M (SD)	GBR, M (SD)	USA, M (SD)	Partial η^2
Physical function	52.31 (9.20)	50.19 (10.02)	50.46 (10.01)	51.24 (9.71)	48.33 (10.64)	51.30 (8.84)	50.37 (8.61)	52.74 (9.54)	48.34 (8.35)	44.69 (6.85)	51.35 (9.42)	52.33 (9.16)	46.76 (6.77)	49.02 (11.71)	49.18 (12.66)	.054
Role function	52.91 (8.39)	50.04 (9.91)	49.55 (9.99)	51.42 (9.35)	49.13 (10.76)	51.66 (8.25)	50.59 (9.55)	52.52 (9.30)	48.33 (9.12)	47.35 (8.51)	50.02 (9.14)	52.31 (8.45)	47.30 (8.41)	48.57 (11.33)	49.14 (11.49)	.029
Emotional function	52.35 (8.97)	50.76 (9.53)	51.95 (10.62)	51.62 (9.89)	50.63 (9.92)	49.67 (8.89)	48.49 (9.22)	53.47 (9.31)	47.27 (9.80)	47.21 (8.83)	50.77 (9.53)	51.14 (9.08)	45.76 (9.33)	48.60 (11.07)	50.50 (11.08)	.040
Cognitive function	52.13 (8.71)	49.62 (9.79)	49.66 (10.31)	51.14 (9.35)	50.27 (10.34)	49.75 (9.02)	50.33 (9.38)	52.83 (8.75)	47.90 (10.02)	46.51 (8.79)	50.29 (9.38)	50.77 (9.02)	45.58 (10.05)	48.19 (11.31)	48.67 (11.61)	.028
Social function	53.05 (7.81)	49.54 (10.67)	50.00 (10.46)	51.80 (9.03)	49.93 (10.07)	51.25 (8.42)	50.35 (9.20)	52.89 (8.27)	46.63 (10.53)	47.67 (9.77)	50.77 (9.65)	52.48 (8.46)	47.62 (10.17)	47.72 (11.44)	48.22 (11.57)	.033
Fatigue	48.39 (9.35)	50.03 (9.28)	50.46 (10.57)	48.77 (10.20)	51.17 (10.54)	49.99 (8.71)	49.22 (9.45)	47.48 (9.49)	52.44 (8.20)	53.79 (8.69)	48.30 (9.49)	48.91 (8.82)	53.65 (8.79)	51.18 (10.79)	50.86 (10.93)	.035
Nausea/vomiting	47.44 (5.77)	50.30 (9.95)	51.35 (10.88)	48.87 (8.76)	49.96 (10.34)	49.43 (8.72)	50.79 (10.34)	48.13 (6.89)	51.11 (10.98)	51.42 (10.11)	49.49 (9.34)	48.83 (7.66)	54.32 (12.08)	51.10 (11.39)	52.51 (12.36)	.018
Pain	48.36 (8.98)	50.38 (9.74)	50.25 (9.97)	48.64 (9.45)	50.84 (10.91)	50.21 (8.83)	49.06 (9.20)	47.48 (8.79)	51.95 (9.19)	52.10 (9.27)	50.31 (9.35)	48.71 (9.32)	52.03 (8.83)	50.73 (11.15)	51.27 (10.96)	.016
Dyspnoea	46.91 (8.69)	49.98 (9.93)	48.46 (9.54)	48.95 (9.70)	50.72 (10.89)	48.33 (8.56)	51.59 (9.65)	47.87 (8.80)	50.08 (9.32)	52.95 (9.32)	48.66 (9.19)	50.72 (7.73)	52.54 (9.41)	50.77 (10.89)	51.25 (11.29)	.026
Sleep problems	48.70 (9.67)	51.62 (10.01)	50.47 (9.94)	49.22 (9.99)	50.96 (10.99)	48.72 (8.78)	48.25 (8.81)	48.40 (8.87)	50.57 (9.47)	51.47 (9.45)	49.23 (9.19)	48.80 (8.64)	51.82 (8.71)	52.33 (10.68)	51.46 (10.66)	.017
Appetite loss	46.87 (7.20)	50.25 (10.20)	51.09 (10.87)	49.06 (9.28)	49.92 (10.26)	49.12 (8.69)	49.51 (9.26)	47.36 (7.50)	51.78 (10.76)	51.75 (10.04)	49.45 (9.52)	49.17 (9.00)	56.34 (11.09)	52.11 (11.40)	51.84 (11.35)	.035
Constipation	47.27 (8.26)	50.35 (10.14)	49.79 (10.16)	49.24 (9.35)	48.96 (10.05)	50.74 (9.04)	50.49 (9.80)	46.45 (7.56)	52.82 (10.63)	53.16 (9.75)	51.15 (9.96)	47.77 (9.04)	55.19 (11.07)	50.65 (10.74)	51.88 (11.75)	.045
Diarrhoea	48.43 (8.52)	51.32 (10.40)	50.51 (10.15)	48.63 (8.86)	49.87 (10.33)	50.14 (9.62)	50.25 (10.02)	48.04 (8.28)	51.99 (11.05)	52.09 (10.88)	49.59 (9.41)	48.83 (8.44)	53.57 (11.09)	51.22 (10.82)	52.06 (11.57)	.016
Financial difficulties	47.12 (7.66)	50.46 (10.96)	50.41 (10.44)	47.99 (8.63)	50.04 (10.23)	51.62 (10.47)	50.15 (9.64)	46.97 (7.10)	53.38 (10.81)	54.32 (11.22)	49.73 (9.35)	47.22 (7.58)	54.93 (11.05)	51.25 (11.41)	52.60 (12.23)	.055
Overall HRQoL	54.65 (9.78)	49.89 (9.54)	50.81 (11.01)	50.74 (9.27)	50.49 (10.04)	49.82 (9.40)	49.13 (9.32)	55.79 (9.81)	47.10 (9.11)	46.91 (8.79)	50.34 (9.89)	51.74 (10.47)	47.33 (10.22)	48.50 (10.80)	49.19 (10.49)	.059

Country codes: AUT = Austria; CAN=Canada; DEU = Germany; DNK = Denmark; FRA = France; HUN = Hungary; ITA = Italy; NLD = the Netherlands; POL = Poland; RUS = Russia; ESP = Spain; SWE=Sweden; TUR = Turkey; GBR = United Kingdom, USA = United Stated of America.
EORTC CAT = The computer-adaptive test of the European Organisation for Research and Treatment of Cancer; HRQoL = health-related quality of life.
[a] T-scores showed statistically significant differences ($P < .001$) between countries in each EORTC CAT Core domain.

support the notion that our data are suitable to establish the 'European CAT Norm' as well as valid norm scores for the 15 countries included in our study. Second, using online panels, we were able to collect a large database of N = 15,386 covering 15 countries and balanced by sex and age groups from 18 to 70+ years. This large sample size enabled detailed DIF analyses and precise T-score transformations. Of note, using linear transformation to transform z-scores to T-scores based on general population data, in which a substantial proportion of participants have 'perfect' scores (e.g. no pain), leads to distributional properties of the T-scores that do not follow a normal distribution. Linear transformation into T-scores is the current standard and also used by, for example, Patient-Reported Outcomes Measurement Information System (PROMIS) [34]. However, these specific distributional properties of the T-scores need to be kept in mind when interpreting the scores. Finally, for practical reasons, we were only able to collect data on item subsets from each EORTC CAT Core item bank. It was not feasible to collect data on 262 items as this would bring other problems such as respondent burden. We have to assume that selected items are representative for the full item banks. As the included items were carefully selected based on each item's psychometric properties and content validity considerations, the data presented herein are robust, state-of-the-art general population norm data for the EORTC CAT Core.

5. Conclusions

In this article, we present representative general population data for the cancer-specific computer-adaptive PRO assessment system EORTC CAT Core across 11 EU countries, Russia, Turkey, Canada and the United States. By defining the 'European CAT Norm', that is, a common European Norm for the EORTC CAT Core, scores from cancer patients obtained via this new instrument can be easily interpreted. In addition, EORTC CAT Core norm scores are provided for age-, sex- and country-specific subpopulations in 15 countries allowing for meaningful score interpretation and comparisons across countries and cultures.

Conflict of interest statement

None declared.

Ethical statement

Ethical approval was not sought as this study is solely based on panel research data. As opposed to medical research where medical professional codes of conduct apply, there is widespread agreement that health research involving volunteers from the general population is not subject to ethical approval. Both the European Pharmaceutical Market Research Association (EphMRA) and the NHS Health Research Authority specify that this type of research does not require ethical approval as long as the research conforms to ethical guidelines. Our online survey was carried out by the panel research company GfK SE, which is member of EphMRA. The multinational survey conformed to the required ethical standards by obtaining informed consent from all participants and collecting data completely anonymously. Any identification of the respondents through the authors is impossible.

Acknowledgements

This research was funded by the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group (grant number 001 2015). The work of Bernhard Holzner was partially funded by the Austrian Science Fund (FWF #P26930). The authors thank the many translators who kindly volunteered their time to embark on the task of standardising the sociodemographic variables across languages and cultures. In addition to the collaborators of this project, special thanks go to Susan Bartlett for making the variables in English and French suitable for the Canadian context; Eveline Bleiker, Jacobien Kieffer, Marieke van Leeuwen for Dutch; Thierry Conroy for French; Agnes Czimbalmos for Hungarian; Alice Iuso for Italian; TatyanaIonova for Russian; Juan Ignacio Arraras for Spanish; Eva Hammerlid, Yvonne Brandberg for Swedish; Deniz Yüce for Turkish; and Claire Snyder for English for the U.S.American context.

References

[1] Blazeby JM, Avery K, Sprangers M, Pikhart H, Fayers P, Donovan J. Health-related quality of life measurement in randomized clinical trials in surgical oncology. J Clin Oncol 2006;24: 3178–86.

[2] Boele FW, Douw L, Reijneveld JC, Robben R, Taphoorn MJB, Aaronson NK, et al. Health-related quality of life in stable, long-term survivors of low-grade glioma. J Clin Oncol 2015;33:1023–9.

[3] Cella D, Grünwald V, Nathan P, Doan J, Dastani H, Taylor F, et al. Quality of life in patients with advanced renal cell carcinoma given nivolumab versus everolimus in CheckMate 025: a randomised, open-label, phase 3 trial. Lancet Oncol 2016;17:994–1003.

[4] Fayers P, Bottomley A. Quality of life research within the EORTC—the EORTC QLQ-C30. Eur J Cancer 2002;38:125–33.

[5] 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.

[6] Luckett T, King MT, Butow PN, Oguchi M, Rankin N, Price MA, et al. Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: issues, evidence and recommendations. Ann Oncol 2011;22:2179–90.

[7] Holzner B, Bode RK, Hahn EA, Cella D, Kopp M, Sperner-Unterweger B, et al. Equating EORTC QLQ-C30 and FACT-G

scores and its use in oncological research. Eur J Cancer 2006;42: 3169–77.

[8] Knobel H, Loge JH, Brenne E, Fayers P, Hjermstad MJ, Kaasa S. The validity of EORTC QLQ-C30 fatigue scale in advanced cancer patients and cancer survivors. Palliat Med 2003;17: 664–72.

[9] Cella D, Gershon R, Lai J-S, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res 2007;16:133–41.

[10] Bjorner JB, Chang C-H, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. Qual Life Res 2007;16:95–108.

[11] Liegl G, Gandek B, Fischer HF, Bjorner JB, Ware JE, Rose M, et al. Varying the item format improved the range of measurement in patient-reported outcome measures assessing physical function. Arthritis Res Ther 2017;19:66.

[12] Petersen MA, Aaronson NK, Arraras JI, Chie WC, Conroy T, Costantini A. The EORTC computer-adaptive tests measuring physical functioning and fatigue exhibited high levels of measurement precision and efficiency. J Clin Epidemiol 2013;66.

[13] Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. Arthritis Res Ther 2011;13:R147.

[14] Petersen MA, Groenvold M, Aaronson NK, Chie W-C, Conroy T, Costantini A, et al. Development of computerised adaptive testing (CAT) for the EORTC QLQ-C30 dimensions–general approach and initial results for physical functioning. Eur J Cancer 2010;46:1352–8.

[15] Petersen MA, Groenvold M, Aaronson NK, Chie W-C, Conroy T, Costantini A, et al. Development of computerized adaptive testing (CAT) for the EORTC QLQ-C30 physical functioning dimension. Qual Life Res 2011;20:479–90.

[16] Gamper EM, Petersen MA, Aaronson N, Costantini A, Giesinger JM, Holzner B, et al. Development of an item bank for the EORTC role functioning computer adaptive test (EORTC RF-CAT). Health Qual Life Outcomes 2016;14:72.

[17] Gamper E-M, Groenvold M, Petersen MA, Young T, Costantini A, Aaronson N, et al. The EORTC emotional functioning computerized adaptive test: phases I–III of a cross-cultural item bank development. Psychooncology 2014;23: 397–403.

[18] Giesinger JM, Petersen MA, Groenvold M, Aaronson NK, Arraras JI, Conroy T, et al. Cross-cultural development of an item list for computer-adaptive testing of fatigue in oncological patients. Health Qual Life Outcomes 2011;9:19.

[19] Petersen MA, Aaronson NK, Chie WC, Conroy T, Costantini A, Hammerlid E, et al. Development of an item bank for computerized adaptive test (CAT) measurement of pain. Qual Life Res 2016;25:1–11.

[20] Petersen MA, Giesinger JM, Holzner B, Arraras JI, Conroy T, Gamper E-M, et al. Psychometric evaluation of the EORTC computerized adaptive test (CAT) fatigue item pool. Qual Life Res 2013;22:2443–54.

[21] Thamsborg LH, Petersen MA, Aaronson NK, Chie WC, Costantini A, Holzner B, et al. Development of a lack of appetite item bank for computer-adaptive testing (CAT). Support Care Cancer 2015;23:1541–8.

[22] Petersen MA, Aaronson NK, Arraras JI, Chie W-C, Conroy T, Costantini A, et al. The EORTC CAT Core—the computer adaptive version of the EORTC QLQ-C30 questionnaire. Eur J Cancer 2018;100:8–16.

[23] Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, et al. A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. J Clin Epidemiol 2009;62:288–95.

[24] Holland PW, Wainer H. Differential item functioning. Routledge; 2012.

[25] Choi SW, Gibbons LE, Crane PK. Lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. J Stat Software 2011;39:1.

[26] Jodoin MG, Gierl MJ. Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. Appl Meas Educ 2001;14:329–49.

[27] Cohen J. Statistical power analysis for the behavioral sciences. 1988.

[28] United Nations. World population prospects: the 2017 revision. 2017.

[29] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012, ISBN 3-900051-07-0. 2014.

[30] IMB Corp. IBM SPSS statistics for windows, version 22.0. Armonk, NY: IBM Corp; 2013.

[31] Hinz A, Singer S, Brahler E. European reference values for the quality of life questionnaire EORTC QLQ-C30: results of a German investigation and a summarizing analysis of six European general population normative studies. Acta Oncol 2014;53:958–65.

[32] Jensen RE, Potosky AL, Moinpour CM, Lobo T, Cella D, Hahn EA, et al. United States population-based estimates of patient-reported outcomes measurement information system symptom and functional status reference values for individuals with cancer. J Clin Oncol 2017;35:1913–20.

[33] Fischer HF, Wahl I, Nolte S, Liegl G, Brahler E, Lowe B, et al. Language-related differential item functioning between English and German PROMIS depression items is negligible. Int J Methods Psychiatr Res 2017;26.

[34] http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis [Accessed 30 August 2018].

[35] Nolte S, Liegl G, Petersen MA, Aaronson NK, Costantini A, Fayers PM, et al. General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the Unites States. Eur J Cancer 2019;107:153–63.