
Article:
Petersen, MA, Aaronson, NK, Arraras, JI et al. (26 more authors) (2018) The EORTC CAT Core—The computer adaptive version of the EORTC QLQ-C30 questionnaire. European Journal of Cancer, 100. pp. 8-16. ISSN 0959-8049

https://doi.org/10.1016/j.ejca.2018.04.016

The EORTC QLQ-C30-CAT – the computer adaptive version of the EORTC QLQ-C30 questionnaire

Morten Aa. Petersen[a,*], Neil K. Aaronson[b], Juan I. Arraras[c], Wei-Chu Chie[d], Thierry Conroy[e], Anna Costantini[f], Linda Dirven[g,h], Peter Fayers[i], Eva-Maria Gamper[j], Johannes M. Giesinger[j], Esther J.J. Habets[g], Eva Hammerlid[k], Jorunn Helbostad[l], Marianne J. Hjermstad[m], Bernhard Holzner[j], Colin Johnson[n], Georg Kemmler[j], Madeleine T. King[o], Stein Kaasa[p], Jon H. Loge[q], Jaap C. Reijneveld[r,s], Susanne Singer[t], Martin J.B. Taphoorn[g,h], Lise H. Thamsborg[a], Krzysztof A. Tomaszewski[u], Galina Velikova[v], Irma M. Verdonck-de Leeuw[w], Teresa Young[x] & Mogens Groenvold[a,y] on behalf of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group

[a] The Research Unit, Department of Palliative Medicine, Bispebjerg Hospital, University of Copenhagen , Copenhagen, Denmark

[b] Division of Psychosocial Research & Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[c] Medical Oncology Department, Hospital of Navarre, Pamplona, Spain

[d] Institute of Epidemiology and Preventive Medicine and Department of Public Health, College of Public Health, National Taiwan University, Taiwan

[e] Medical Oncology Department, Institut de cancérologie de Lorraine, Vandoeuvre-lès-Nancy, France

[f] Psychoncology Unit, Sant'Andrea Hospital, Faculty of Medicine and Psychology Sapienza University, Rome, Italy

[g] Department of Neurology, Haaglanden Medical Center, PO BOX 432, 2501 CK The Hague, the Netherlands

[h] Department of Neurology, Leiden University Medical Center, PO BOX 9600, 2300 RC Leiden, the Netherlands

[i] Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

[j] Department of Psychiatry, Psychotherapy and Psychosomatic Medicine, Innsbruck Medical University, Innsbruck, Austria

[k] Dept of Otolaryngology Head and Neck Surgery, Sahlgrenska University Hospital, Göteborg University, Göteborg, Sweden

[l] Department of Neuroscience, Norwegian University of Science and Technology and St. Olav University Hospital, Trondheim, Norway

[m] European Palliative Care Research Centre (PRC), Department of Oncology, Oslo University Hospital, and Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[n] Surgical Unit, Faculty of Medicine, University of Southampton, Southampton, UK

[o] School of Psychology, Faculty of Science and Sydney Medical School, Faculty of Medicine, University of Sydney, Sydney, NSW, Australia

[p] Oslo University Hospital and University of Oslo, Norway and European Palliative Care Research Centre (PRC), Norwegian University of Science and Technology, Oslo, Norway

[q] Palliative Medicine Unit, University Hospital of Trondheim, Trondheim, Norway

[r] Department of Neurology and Brain Tumor Center Amsterdam, VU University Medical Center, PO BOX 7057, 1007 MB Amsterdam, the Netherlands

[s] Department of Neurology and Brain Tumor Center Amsterdam, Academic Medical Center, PO BOX 22660, 1100 DD Amsterdam, the Netherlands

[t] Division of Epidemiology and Health Services Research, Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Centre, Mainz, Germany

[u] Health Outcomes Research Unit, Department of Gerontology, Geriatrics, and Social Work, Faculty of Education, Ignatianum Academy, Krakow, Poland

[v] Leeds Institute of Cancer and Pathology, Faculty of Medicine and Health, University of Leeds, Leeds, UK

[w] Department of Otolaryngology – Head & Neck Surgery, VU University Medical Center; Amsterdam, the Netherlands

[x] East & North Hertfordshire NHS Trust incorporating Mount Vernon Cancer Centre, Northwood, Middlesex, UK

[y] Institute of Public Health, University of Copenhagen, Copenhagen, Denmark

* Corresponding author: The Research Unit, Department of Palliative Medicine, Bispebjerg Hospital, Bispebjerg bakke 23B, 2400 Copenhagen NV, Denmark. Telephone: (+45) 3863 5016. Fax: (+45) 3863 9805. Email: Morten.Aagaard.Petersen@regionh.dk

Running head:

The EORTC QLQ-C30-CAT

Abstract

Background: To optimise measurement precision, relevance to patients and flexibility, patient-reported outcome measures (PROMs) should ideally be adapted to the individual patient/study while retaining direct comparability of scores across patients/studies. This is achievable using item banks and computerized adaptive tests (CATs). The EORTC QLQ-C30 is one of the most widely used PROMs in cancer research and clinical practice. Here we provide an overview of the research program to develop CAT versions of the QLQ-C30's 14 functional and symptom domains.

Methods: The EORTC Quality of Life Group's strategy for developing CAT item banks consists of: literature search to identify potential candidate items; formulation of new items compatible with the QLQ-C30 item style; expert evaluations and patient interviews; field-testing and psychometric analyses, including factor analysis, item response theory (IRT) calibration, and simulation of measurement properties. In addition, software for setting up, running, and scoring CAT has been developed.

Results: Across 8 rounds of data collections, 9,782 patients were recruited from 12 countries for the field-testing. The four phases of development resulted in a total of 260 unique items across the 14 domains. Each item bank consists of 7-34 items. Psychometric evaluations indicated higher measurement precision and increased statistical power of the CAT measures compared to the QLQ-C30 scales. Using CAT, sample size requirements may be reduced by approximately 20-35% on average without loss of power.

Conclusions: The QLQ-C30-CAT represents a more precise, powerful and flexible measurement system than the QLQ-C30. It is currently being validated in a large independent, international sample of cancer patients.

Introduction

Patient-reported outcomes (PROs) are the primary source of information about patients' health-related quality of life (HRQOL). PRO measures (PROMs) are typically static, standardised questionnaires i.e., all patients are asked the same set of items yielding scores that are comparable across patients. To achieve precise measurements for patients at different levels of HRQOL, traditional PROMs often require a substantial number of items; more than may be feasible and/or reasonable to ask patients to complete. Therefore, such PROMs typically represent a compromise between the need to minimize patient burden, while achieving adequate measurement precision.

Item response theory (IRT) provides a family of statistical models to describe the psychometric characteristics of items in multi-item scales.[1] In recent years, there has been an increasing interest in the use of IRT when developing new PRO/HRQOL measures and for enhancing existing ones. A simple search in PubMed® using search terms 'item response theory' AND ('quality of life' OR 'patient reported outcome') resulted in 5 hits for 2000, 21 for 2005, and 69 for 2015.[2] One of the primary reasons for this increasing interest and one of the major advantages of IRT is that, when a set of items has been calibrated (estimated) to an IRT model, scores based on any subset of the items are on the same metric and hence, are directly comparable. This unique feature allows the content of questionnaires to be adapted to the individual patient without compromising the comparability of scores across patients. This is utilised by computerized adaptive tests (CATs) to present the most informative items to each patient, thereby optimising the measurement properties.[3] During a CAT assessment, item selection is tailored to the individual based on responses to prior items, i.e. in each step the choice of item is adapted to the current estimate of the patient's location on a health continuum (e.g., physical functioning or fatigue).

CAT measures have several advantages over measures based on classical test theory. These include increased measurement precision, reduced respondent burden, increased question relevance to individual

patients, and increased flexibility. The length of the CAT questionnaire can be adapted to each study or patient, and scores can be generated automatically, facilitating real time feedback of results.

Because of the clear advantages of CAT measurement, the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group (QLG) initiated in 2005 a project to develop a CAT PROM specifically relevant for cancer patients. Since the EORTC QLG's core questionnaire, the QLQ-C30, is one of the most widely used PROMs in cancer research and clinical practice,[4] the QLG deemed it particularly relevant to enhance the measurement properties of this instrument. Therefore, the aim of this project was to develop item banks (calibrated collections of items) for CAT measurement of the 14 functional and symptom domains out of the 15 domains assessed by the EORTC QLQ-C30 (additional items have not been developed for the overall quality of life scale).[5,6] Development of the CAT required supplementing the QLQ-C30 with additional items to more fully assess each HRQOL domain. Basing the CAT on the conceptual model of the QLQ-C30 ensures maximum backward compatibility with the original instrument. Hence, future studies using the EORTC CAT can compare results with the substantial body of literature of studies using the original QLQ-C30. In particular, a CAT assessment can be set up to ask all QLQ-C30 items (supplemented with additional items for increased precision) if a direct assessment of QLQ-C30 is desired.

We have completed the development of all 14 item banks for the CAT version of the QLQ-C30. The analyses and results of some, but not all of the individual item banks have been published in details.[6-17] Here we present for the first time an overview of the complete EORTC QLQ-C30-CAT instrument.

Methods

The development of all 14 CAT item banks followed the same general approach, which comprises four phases: 1) literature search, 2) operationalisation (selection and formulation of items), 3) pre-testing, and 4) field-testing. This approach has been described in details elsewhere.[6,12] Following is a summary of the approach.

A literature search was conducted to identify existing items used to measure the domain of focus (e.g., physical functioning or fatigue). These were not systematic reviews, but aimed at acquiring sufficient information about previous measurement of the domain to be able to formulate new, relevant items.

The retrieved items were assessed for relevance to EORTC measurement and for redundancy. As the aim was to retain the conceptual model from the QLQ-C30, items were discarded if not reflecting the HRQOL domains as measured by the QLQ-C30. For example, the QLQ-C30 items on physical functioning do not cover agility. Hence, items on agility were deemed not relevant for our measurement of physical functioning. Other items were discarded because they could not be re-worded to fit the standard QLQ-C30 timeframe and/or response options (e.g. an item asking *'how many times did you…?'* would not fit the response options *'not at all'* to *'very much'*) or because they were too similar to items already selected. The list of retained items was used as a starting point for formulating new items covering the full range of the domain of interest (e.g., from poor to excellent physical functioning) and fitting the QLQ-C30 item style. The list of items was evaluated by international HRQOL/PRO experts and revised accordingly.

In phase 3, the revised items were evaluated by a heterogeneous, international sample of cancer patients. Items from each domain were evaluated by at least 30 patients from ≥3 countries. Patient input on item relevance, clarity, coverage etc. was obtained using semi-structured interviews. Before the interviews, items were translated into the relevant languages by the Translation Office of the EORTC Quality of Life Department according to rigorous and well-established guidelines.[18] The list of items was revised based on the patient comments.

In phase 4, the resulting item banks were field-tested in heterogeneous, international samples of cancer patients. The goal was to include a total of about 1,000 patients across ≥3 countries. The patients completed the new items together with the QLQ-C30. The resulting dataset formed the basis for the final psychometric evaluations. These included: evaluating dimensionality using factor analysis for ordinal variables; calibrating the IRT model (the generalized partial credit model) and evaluating item fit; evaluating

differential item functioning (DIF) to explore whether items function similarly across different groups of patients (e.g., men and women, patients from different countries); and evaluating the measurement precision of the CATs based on the resulting item banks using both observed and simulated data.[14] Items not fitting the unidimensional IRT model or exhibiting DIF were candidates for exclusion.

Ethical approval for phases 3 and 4 was obtained in accordance with local ethical standards.

CAT measurement requires specialized software for real time selection of items, for estimating scores, etc. Therefore the EORTC QLG has developed software for conducting online CAT assessment. The group has also developed software for selecting items for short forms. Short forms are static, fixed questionnaires composed of items selected from the item banks. Scores based on such short forms are directly comparable with those based on the dynamic CAT version and are particularly useful when (online) electronic data collection is infeasible.

Results

The literature searches in phase 1 revealed a wide range in the number of items that have been used to measure the different QLQ-C30 domains, ranging from 122 role functioning items to 1,729 items for emotional functioning. Across all domains, deletion of redundant items and items not relevant for our measurement model or incompatible with the QLQ-C30 item style reduced the lists by 69-94% leaving, on average, 15% of the items identified in the literature.

The retained items formed the basis for formulating new, unique items that were compatible with the EORTC item style. This resulted in 14-86 candidate items across the 14 item banks. These items were then evaluated by experts from a variety of fields, reducing the lists to 12-55 items per domain.

In phase 3, we conducted 11 separate pre-tests, each comprising items from 1 to 3 domains. Each pre-test included a mixed sample of 31-52 cancer patients from 3-5 countries, resulting in a total of 433 patient

interviews. Patient characteristics are shown in Table 1. Based on feedback from patients about items being difficult to understand, confusing, intrusive or otherwise problematic, some items were reformulated or deleted. If patients judged that items on important aspects of a domain were missing, new items were formulated. These patient-based revisions resulted in item lists of 12 to 51 items per domain for the final phase 4 evaluations.

(Table 1 about here)

To reduce response burden and because of the chronology of item bank development, phase 4 consisted of 8 patient surveys, each covering 1-3 domains. Each survey included 858-1,321 patient responses, with a total of 9,782 responses. Patients were recruited from 12 countries (between 4 and 7 countries per survey). British and Danish patients participated in all surveys. Sociodemographic and clinical characteristics of the heterogeneous patient sample are shown in Table 2.

(Table 2 about here)

We evaluated the psychometric properties of the items for each domain following a thorough, stepwise procedure. Across the 14 domains, we evaluated a total of 372 candidate items. The primary reasons for deleting items were multidimensionality and misfit to the IRT model. Across the domains, approximately 20% of the candidate items were deleted to ensure that the list of items for each domain was sufficiently unidimensional for the IRT analysis. Additionally, 10% of items across the domains were deleted to obtain acceptable fit to the IRT models. For all domains, DIF analyses indicated that some items might perform slightly differently across specific patient groups, particularly across certain countries/languages. However, the potential DIF identified was relatively trivial in the sense that it was unlikely to have practical impact on the estimation of domain scores, i.e. the DIF would not bias comparisons across groups. Taken together, the psychometric evaluations resulted in the deletion of between 0% and 47% of the candidate items within each domain. The 14 resulting CAT item banks range from 7 items for lack of appetite to 34 items for

fatigue and cognitive functioning (see Table 3) and comprise in total 260 items including the 28 QLQ-C30 items (+ 2 items for overall quality of life).

(Table 3 about here)

CAT measurements asking 1, 2, 3,… items were simulated and the measurement precision/statistical power of these CATs to detect expected group differences were compared with the original QLQ-C30 scales. From this, the required sample sizes needed with the CATs to obtain the same power as with the QLQ-C30 scales were estimated. The averages of these simulations across the 14 domains are plotted in Fig. **1**. The required sample size for the CATs was, on average, substantially lower than for the QLQ-C30 scales. For example, asking 4 items in a CAT requires, on average only 75% of the sample size to match the power of the corresponding QLQ-C30 scale. As expected the more items asked in the CAT, the more power, i.e. smaller sample sizes are needed. Note that there were substantial differences in the estimated maximal sample size savings across the domains, ranging from 15% for social functioning to 55% for dyspnea.

(Fig. 1 about here)

The software developed for conducting CAT measurement consists of two parts: a front end, visible to the respondent, and a back end that performs the calculations needed to conduct a CAT assessment. The current online version of the front end, CHES, was developed by Evaluation Software Development™. A demo version of this CHES-EORTC-CAT platform is available at https://eortc.ches.pro/ and a screenshot is shown in Fig. **2**. The front end has a simple and intuitive design allowing most respondents to use it without help or instructions. The back end, termed the CAT-engine, has been developed within the EORTC CAT-project. Using a simple web-interface, required information regarding which items to ask and the item responses provided is shared between the CAT-engine and CHES. The CAT-engine may communicate with other electronic PRO data collection systems such as REDCap (https://www.project-redcap.org/), to run a CAT assessment.

We have also developed a tool for selecting items for short forms. Fig. **3** shows an example of the current version of this short form generator software. For the domain of interest (e.g., physical functioning) the expected mean and standard deviation of the target population together with the score range of interest are inputted. The generator then estimates which items will be the most informative. It also provides a plot of the information function of the selected items and the score distribution of the target population (see Fig. **3**). This tool is particularly useful for selecting the most informative items for a static short form questionnaire when it is not feasible to use the CAT system.

(Fig. 2 and Fig. 3 about here)

Discussion

CAT measurement adapts in real time a PROM to the individual patient while retaining direct comparability of scores across patients. The EORTC QLQ-C30-CAT presented here measures the same HRQOL domains as the widely used EORTC QLQ-C30 questionnaire. This 'intelligent' version of the QLQ-C30 allows clinicians and researchers to assess each patient's level of symptoms and functioning using only the most informative items from the item banks that have been developed. Thereby, measurement precision and response burden/time can be adjusted to the needs of the individual study. When minimal response burden or time is essential, a short and quick CAT design can be chosen, while when high precision is paramount for key outcomes (e.g., pain when testing a new pain medication) or when sample size is limited by budget or clinical availability, more items can be asked. Even in cases where electronic data collection is not feasible, one may take advantage of the CAT item banks to construct short forms, i.e. static (paper) questionnaires customised to fit the requirements of the individual studies.

The 14 item banks developed for the QLQ-C30-CAT include 7 to 34 items each and a total of 260 items, i.e. about 9 times the number of items in the QLQ-C30. These item banks have been developed using a thorough mixed-methods approach including literature searches, expert evaluations, interviews with

cancer patients, and psychometric analyses based on large international and mixed samples of cancer patients. The development has been conducted internationally, using different languages and in different cultures which strengthens the robustness of the results and helps to ensure that the instrument is meaningful and applicable across patient groups. Currently the complete set of item banks is available in 10 languages.

A number of other research groups have developed and/or evaluated the use of CAT instruments for PRO measurement, covering a variety of patient groups and HRQOL domains.[19-29] Most prominent of these is the large scale US based PROMIS initiative.[19] PROMIS instruments have been developed to measure physical, mental, and social health for a wide range of patients groups, covering both adults and children. PROMIS measures of physical functioning, social functioning, pain, fatigue, anxiety and depression have been specifically targeted to cancer patients.[30] These measures, as PROMIS measures in general, have been developed in and for use in the US, although work is in progress to translate and validate PROMIS measures across countries.[31] Hence, in international studies, and in studies outside the US in general, the cross-national development is an advantage of the EORTC QLQ-C30-CAT. The involvement of international samples of patients in the development process is a trademark of all EORTC QLG PROM development.

The expectation that CAT measurement is more efficient and precise than standard fixed-length/fixed-format questionnaires has been confirmed for the EORTC QLQ-C30-CAT. Simulations indicated that on average, sample sizes may be reduced by typically 25-30% as compared to using the QLQ-C30, without reducing the power of a study. This means that costs and time for PRO studies can be significantly reduced without affecting quality, simply by using this new instrument. The increased measurement precision also means that the CAT instrument will be more appropriate for measurement at the individual patient level, e.g. for monitoring changes over time, which generally requires more precise measures than studies at group level.

It should be noted that the simulations were based on the same data as used for the development of the QLQ-C30-CAT. Therefore, a large international validation study has been initiated to assess the psychometric properties of the QLQ-C30-CAT in an independent sample. When this validation is completed, the QLQ-C30-CAT will be released as a fully validated EORTC instrument. However, the current version is available for use, with the understanding that some minor refinements in the instrument may be required based on the findings of the validation study. The EORTC QLG website provides more information on the current use of the QLQ-C30-CAT and short forms at http://groups.eortc.be/qol/eortc-cat.

Concurrent with the validation study, large samples of general population data from several countries across Europe have been collected.[32] These will be used to norm the QLQ-C30-CAT so that the European general population has a mean of 50 and a standard deviation of 10 (a so-called T-scoring) for each HRQOL domain. This scoring allows for simple interpretation of patients' HRQOL in comparison with the general population (stratified by age and gender, when desired). Norm data for the US, Canada, Russia and Turkey have also been collected. These reference data and the norming of the QLQ-C30-CAT will be ready and implemented before the release of the validated instrument.

With this psychometrically sound CAT version of the widely used EORTC QLQ-C30, the measurement precision and flexibility of EORTC PRO assessment in cancer patients will be significantly improved. This means more reliable and relevant PRO assessment in clinical research and trials, but also that the QLQ-C30-CAT has greater potential for use in clinical practice for monitoring individual changes and enhancing patient-clinician communication. Although cancer patients are the primary target group of the QLQ-C30-CAT, it may, as is the case with the QLQ-C30, be applicable and relevant in many other patient groups as well as in the general population.

Conflict of interest statement

None declared.

Funding

The study was funded by the EORTC Quality of Life Group. The work of Johannes M. Giesinger was funded by the Austrian Science Fund (FWF #P26930).

Role of the funding source

The funder of the study had no role in the study design, data collection, analysis, or interpretation, or writing of the article. The executive committee of the EORTC Quality of Life Group read the article and approved it as being 'on behalf of EORTC Quality of Life Group'. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

The EORTC Quality of Life Group's PROMs can be used free of charge in academic settings, but the group receives compensation fees when its PROMs are used in commercial settings. The CHES system for electronic data capture is developed by the commercial company Evaluation Software Development™ (ESD).

References

1. Reise SP, Revicki DA .*Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*. New York: Routledge; 2015.

2. MEDLINE search. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/ January 5th 2017. 2017;

3. Wainer H .*Computerized Adaptive testing: A Primer*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.; 2000.

4. Fayers P, on behalf of the EORTC Quality of Life Group and of the Quality of Life Unit. Quality of life research within the EORTC-the EORTC QLQ - C30. *Eur J Cancer* 2002;38 Suppl 4:S125-S133.

5. Aaronson NK, Ahmedzai S, Bergman B 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(5):365-76.

6. Petersen MAa, Groenvold M, Aaronson NK 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.

7. Dirven L, Groenvold M, Taphoorn MJB et al. Psychometric evaluation of an item bank for computerized adaptive testing of the EORTC QLQ-C30 cognitive functioning dimension in cancer patients. *Qual Life Res* 2017;26:2919-29.

8. Dirven L, Taphoorn MJ, Groenvold M et al. Development of an item bank for computerized adaptive testing of self-reported cognitive difficulty in cancer patients. *Neuro-Oncology Practice* 2017;4(3):189-96.

9. Gamper EM, Groenvold M, Petersen MAa et al. The EORTC Emotional Functioning computer adaptive test (CAT): phase I-III of a cross-cultural item bank development. *Psycho-Oncology* 2014;23:397-403.

10. Gamper EM, Petersen MAa, Aaronson N 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)

11. Giesinger JM, Petersen MAa, Groenvold M 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)

12. Petersen MAa, Groenvold M, Aaronson NK et al. Development of computerized adaptive testing (CAT) for the EORTC QLQ-C30 physical functioning dimension. *Qual Life Res* 2011;20(4):479-90.

13. Petersen MAa, Giesinger JM, Holzner B et al. Psychometric evaluation of the EORTC computerized adaptive test (CAT) fatigue item pool. *Qual Life Res* 2013;22(9):2443-54.

14. Petersen MAa, Aaronson NK, Arraras JI et al. The EORTC computer-adaptive tests measuring physical functioning and fatigue exhibited high levels of measurement precision and efficiency. *J Clin Epidemiol* 2013;66(3):330-9.

15. Petersen MAa, Aaronson NK, Chie W-C et al. Development of an item bank for computerized adaptive test (CAT) measurement of pain. *Qual Life Res* 2016;25(1):1-11.

16. Petersen MAa, Gamper EM, Costantini A et al. An emotional functioning item bank of 24 items for computer adaptive testing (CAT) was established. *J Clin Epidemiol* 2016;70:90-100.

17. Thamsborg LH, Petersen MAa, Aaronson NK et al. Development of a lack of appetite item bank for computer-adaptive testing (CAT). *Support Care Cancer* 2014;23(6):1541-8.

18. Kulis D, Bottomley A, Velikova G, Greimel E, Koller M .*EORTC Quality of Life Group Translation Procedure - 4th edition*. Brussels: European Organization for Research and Treatment of Cancer; 2016.

19. The Patient-Reported Outcomes Measurement Information System (PROMIS). http://www.healthmeasures.net/explore-measurement-systems/promis. Accessed August 2017.

20. Anatchkova MD, Saris-Baglama RN, Kosinski M, Bjorner JB. Development and preliminary testing of a computerized adaptive assessment of chronic pain. *J Pain* 2009;10(9):932-43.

21. Bjorner JB, Kosinski M, Ware JE, Jr. Calibration of an item pool for assessing the burden of headaches: an application of item response theory to the headache impact test (HIT). *Qual Life Res* 2003;12(8):913-33.

22. Devine J, Otto C, Rose M et al. A new computerized adaptive test advancing the measurement of health-related quality of life (HRQoL) in children: the Kids-CAT. *Qual Life Res* 2015;24(4):871-84.

23. Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res* 2005;14(10):2277-91.

24. Gershon RC, Lai JS, Bode R et al. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. *Qual.Life Res.* 2012;21(3):475-86.

25. Gibbons RD, Weiss DJ, Pilkonis PA et al. Development of a computerized adaptive test for depression. *Arch.Gen.Psychiatry* 2012;69(11):1104-12.

26. Haley SM, Ni P, Fragala-Pinkham MA, Skrinar AM, Corzo D. A computer adaptive testing approach for assessing physical functioning in children and adolescents. *Dev.Med Child Neurol.* 2005;47(2):113-20.

27. Hart DL, Wang YC, Stratford PW, Mioduski JE. Computerized adaptive test for patients with foot or ankle impairments produced valid and responsive measures of function. *Qual Life Res* 2008;17(8):1081-91.

28. Jette AM, Haley SM, Ni P, Olarsch S, Moed R. Creating a computer adaptive test version of the late-life function and disability instrument. *J Gerontol.A Biol.Sci.Med.Sci.* 2008;63(11):1246-56.

29. Lai JS, Cella D, Dineen K et al. An item bank was created to improve the measurement of cancer-related fatigue. *J Clin Epidemiol* 2005;58(2):190-7.

30. Garcia SF, Cella D, Clauser SB et al. Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative. *J Clin.Oncol.* 2007;25(32):5106-12.

31. The Patient-Reported Outcomes Measurement Information System (PROMIS) International. http://www.healthmeasures.net/explore-measurement-systems/promis/measure-development-research/promis-international. Accessed August 2017. 2017;

32. Nolte S, Liegl G, Petersen MAa et al. (202.1) First results of a 15,000 person population survey to establish European norm data for the EORTC Computer Adaptive Test QLQ-CAT. *Qual.Life Res.* 2017;26(Suppl. 1):28-9.

Legends

Table 1. Sociodemographic and clinical characteristics of the N=433 patients included across the 11 phase 3 pre-testings*

Table 2. Sociodemographic and clinical characteristics of the N=9,782 patients included across the 8 phase 4 data collections*

Table 3. Number of items per domain in the EORTC QLQ-C30 and the EORTC QLQ-C30-CAT (including the QLQ-C30 items), respectively.

Fig. 1. The simulated relative required sample size (%) using CAT measurement compared to using the QLQ-C30 sum scales. Values less than 100% indicate that smaller samples may be used with CAT without reducing the power.

Fig. 2. Example of the CHES-EORTC-CAT software.

Fig. 3. Example of the current version of the EORTC short form generator software.

Table 1. Sociodemographic and clinical characteristics of the N=433 patients included across the 11 phase 3 pre-testings*

		N (per cent)**
Age, mean (response rate)		58·8 years (96·5%)
Gender	Female	233 (53·8%)
	Male	193 (44·6%)
Country	Austria	20 (4·6%)
	Denmark	117 (27·0%)
	France	42 (9·7%)
	Germany	20 (4·6%)
	Italy	45 (10·4%)
	Poland	20 (4·6%)
	Spain	30 (6·9%)
	Taiwan	20 (4·6%)
	The Netherlands	3 (0·7%)
	UK	116 (26·8%)
Cancer stage	I-II	147 (33·9%)
	III-IV	252 (58·2%)
Cancer site	Breast	78 (18·0%)
	Gastrointestinal	109 (25·2%)
	Gynaecological	50 (11·5%)
	Head and neck	41 (9·5%)
	Lung	32 (7·4%)
	Urogenital	40 (9·2%)
	Other	69 (15·9%)
Current treatment	Chemotherapy	222 (51·3%)
	Other treatment	73 (16·9%)
	No current treatment	119 (27·5%)
Education	0-10 years	88 (20·3%)
	11-13 years	124 (28·6%)
	14-16 years	95 (21·9%)
	>16 years	98 (22·6%)
Work	Retired	187 (43·2%)
	Working	139 (32·1%)
	Other	75 (17·3%)
Cohabitation	Living with a partner	297 (68·6%)
	Living alone	110 (25·4%)

*: A minor subsample of patients has participated in more than one pretesting and hence, contributes more than once to this count.

**: Some per cents sum to less than 100% because of missing data.

Table 2. Sociodemographic and clinical characteristics of the N=9,782 patients included across the 8 phase 4 data collections*

		N (per cent)**
Age, mean		60·7 years (99·4%)
Gender	Female	5,300 (54·2%)
	Male	4,443 (45·4%)
Country	Australia	122 (1·2%)
	Austria	476 (4·9%)
	Denmark	2,094 (21·4%)
	France	964 (9·9%)
	Germany	263 (2·7%)
	Italy	262 (2·7%)
	Poland	532 (5·4%)
	Spain	422 (4·3%)
	Sweden	326 (3·3%)
	Taiwan	307 (3·1%)
	The Netherlands	126 (1·3%)
	UK	3,888 (39·7%)
Cancer stage	I-II	4,734 (48·4%)
	III-IV	4,386 (44·8%)
Cancer site	Breast	1,916 (19·6%)
	Gastrointestinal	1,370 (14·0%)
	Gynaecological	1,142 (11·7%)
	Head and neck	1,086 (11·1%)
	Lung	519 (5·3%)
	Urogenital	1,481 (15·1%)
	Other	1,918 (19·6%)
Current treatment	Chemotherapy	3,571 (36·5%)
	Other treatment	2,277 (23·3%)
	No current treatment	3,903 (39·9%)
Education	0-10 years	2,989 (30·6%)
	11-13 years	2,268 (23·2%)
	14-16 years	2,310 (23·6%)
	>16 years	2,010 (20·5%)
Work	Retired	4,852 (49·6%)
	Working	3,220 (32·9%)
	Other	1,551 (15·9%)
Cohabitation	Living with a partner	7,039 (72·0%)
	Living alone	2,573 (26·3%)

*: A minor subsample of patients has participated in more than one data collection and hence, contributes more than once to this count.

**: Some per cents sum to less than 100% because of missing data.

Table 3. Number of items per domain in the EORTC QLQ-C30 and the EORTC QLQ-C30-CAT (including the QLQ-C30 items), respectively.

Domain	# items in QLQ-C30	# items in CAT item bank
Physical functioning	5	31
Emotional functioning	4	24
Cognitive functioning	2	34
Social functioning	2	13
Role functioning	2	10
Fatigue	3	34
Nausea & vomiting	2	19
Pain	2	16
Dyspnoea	1	32
Diarrhoea	1	13
Constipation	1	10
Financial difficulties	1	9
Insomnia	1	8
Lack of appetite	1	7
Overall quality of life*	2	2
Total across 15 domains	**30**	**262**

*: additional items for CAT have not been developed for overall quality of life.