Support Care Cancer (2016) 24:3915–3924
DOI 10.1007/s00520-016-3216-0

ORIGINAL ARTICLE

International field testing of the psychometric properties of an EORTC quality of life module for oral health: the EORTC QLQ-OH15

Marianne J. Hjermstad[1,2] · Mia Bergenmar[3,4] · Kristin Bjordal[5] · Sheila E. Fisher[6] · Dirk Hofmeister[7] · Sébastien Montel[8] · Ourania Nicolatou-Galitis[9] · Monica Pinto[10] · Judith Raber-Durlacher[11] · Susanne Singer[12] · Iwona M. Tomaszewska[13] · Krzysztof A. Tomaszewski[14] · Irma Verdonck-de Leeuw[15] · Noam Yarom[16,17] · Julie B Winstanley[18,19] · Bente B. Herlofson[20] · on behalf of the EORTC QoL Group

Received: 12 January 2016 / Accepted: 5 April 2016 / Published online: 25 April 2016

Abstract
Purpose This international EORTC validation study (phase IV) is aimed at testing the psychometric properties of a quality of life (QoL) module related to oral health problems in cancer patients.
Methods The phase III module comprised 17 items with four hypothesized multi-item scales and three single items. In phase IV, patients with mixed cancers, in different treatment phases from 10 countries completed the EORTC QLQ-C30, the QLQ-OH module, and a debriefing interview. The hypothesized structure was tested using combinations of classical test theory and item response theory, following EORTC guidelines. Test–retest assessments and responsiveness to change analysis (RCA) were performed after 2 weeks.
Results Five hundred seventy-two patients (median age 60.3, 54 % females) were analyzed. Completion took <10 min for 84 %, 40 % expressed satisfaction that these issues were addressed. Analyses suggested a revision of the phase III hypothesized scale structure. Two items were deleted based on a high degree of item misfit, together with negative patient

✉ Marianne J. Hjermstad
marianne.j.hjermstad@ntnu.no

[1] Regional Centre for Excellence in Palliative Care, Department of Oncology, Oslo University Hospital, Oslo, Norway

[2] European Palliative Care Research Centre, Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University for Science and Technology, Trondheim, Norway

[3] Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden

[4] Department of Oncology, Karolinska University Hospital, Stockholm, Sweden

[5] Oslo University Hospital, Oslo, Norway

[6] University of Leeds, Leeds LS2 9JT, UK

[7] University of Leipzig, Leipzig, Germany

[8] Department of Psychology, University Paris Saint Denis, Paris, France

[9] University of Athens, Athens, Greece

[10] Rehabilitation Unit, Department of Quality of Life, Istituto Nazionale Tumori "Fondazione G. Pascale"—IRCCS, Naples, Italy

[11] Department of Oral and Maxillofacial Surgery, Academic Medical Center University of Amsterdam, Amsterdam, Netherlands

[12] Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Centre, Mainz, Germany

[13] Department of Medical Education, Jagiellonian University Medical College, Krakow, Poland

[14] Department of Anatomy, Jagiellonian University Medical College, Krakow, Poland

[15] Department of Otolaryngology—Head and Neck Surgery, VU University Medical Center, Amsterdam, Netherlands

[16] Department of Oral and Maxillofacial Surgery, Sheba Medical Center, Tel-Hashomer, Ramat Gan, Israel

[17] Department of Oral Pathology and Oral Medicine, School of Dental Medicine, Tel Aviv University, Tel Aviv, Israel

[18] Patricia Ritchie Centre, University of Sydney, Sydney, Australia

[19] Osman Consulting Pty LTD, Sydney, Australia

[20] Department of Oral Surgery and Oral Medicine, Faculty of Dentistry, University of Oslo, Oslo, Norway

feedback. The remaining 15 items formed one eight-item scale named OH-QoL score, a two-item information scale, a two-item scale regarding dentures, and three single items (*sticky saliva/mouth soreness/sensitivity to food/drink*). Face and convergent validity and internal consistency were confirmed. Test–retest reliability ($n = 60$) was demonstrated as was RCA for patients undergoing chemotherapy ($n = 117$; $p = 0.06$). The resulting QLQ-OH15 discriminated between clinically distinct patient groups, e.g., low performance status vs. higher ($p < 000.1$), and head-and-neck cancer versus other cancers ($p < 0.03$).
Conclusion The EORTC module QLQ-OH15 is a short, well-accepted assessment tool focusing on oral problems and QoL to improve clinical management.
Trial Registration ClinicalTrials.gov Identifier: NCT01724333.

Keywords Oral health · Quality of life · EORTC QLQ-C30 · QLQ-OH15 · Patient reported outcomes · Validation study

Introduction

The recognition of patient-reported outcomes (PROs) as independent outcomes in cancer represents a major shift in medicine in the last decades [1, 2]. This is consolidated by the CONSORT-PRO Extension Statement developed to improve the reporting of PROs on patients' evaluation of symptoms, functioning, and quality of life (QoL) [3].

In oncology, traditional and targeted agents represent a variety of biological mechanisms with a suppressive effect on the oral epithelium and the salivary glands [4, 5], in line with other medications for general symptom relief, e.g., pain medication, corticosteroids, antihypertensives (4,6–8). The reported prevalence of the most frequent oral complications varies by cancer diagnosis, stage, previous and ongoing antitumor treatment, comorbidities, and study designs: from 6 to 91 % for xerostomia/salivary gland hypofunction with reduced flow or altered composition [4–10], 60–86 % for taste changes [11, 12], 30 % for caries [13], up to 43 %–75 % for viral/fungal infections [5, 14–16], 23–81 % for pain [5, 16], and up to 80 % for mucositis [17, 18]. Osteonecrosis of the jaw, associated with head-and-neck radiotherapy, varies from 5 to 13 % [16, 19] and is also documented after bisphosphonates and denosumab treatments [19].

Studies show that dental and oral problems are more routinely assessed in patients with head-and-neck cancer [20] than in those with general oncology [21, 22]. This may be because surgery and radiotherapy automatically direct the attention to this location [21, 22], and dental examinations are included in pretreatment procedures. Oral side effects are underreported by patients and healthcare providers, especially beyond the phases of active, curative treatment [22]. Nevertheless, depending on the number, intensity, and duration of oral adverse effects, QoL may be compromised after most cancer regimens, in the acute phase and during recovery and follow-up [11, 23–25]. Consequences may be a vicious circle with long-lasting sores, mouth pain, oral infections, and dental problems with caries and loose teeth with a negative impact on QoL dimensions like fatigue, nutritional intake, and social functioning [5].

Optimal PRO evaluation must be based on validated assessment tools that should be brief, patient-centered, and comprehensive. Most oral health assessment tools are either too long, focus on one aspect, e.g., xerostomia, and rarely evaluate impact on QoL [26]. The frequently used European Organisation for Research and Treatment of Cancer (EORTC) core questionnaire (EORTC QLQ-C30) assesses generic QoL aspects [27]; thus, the development of specific site or treatment modules is encouraged for clinical trials [28, 29].

This paper presents phase IV, an international field study of the EORTC oral health QoL module [26], intended for clinical use in cancer patients. Aims were to field-test the module in a large, international group of patients to investigate all aspects of its psychometric properties.

Patients and methods

Study design

Questionnaires The study followed the EORTC Quality of Life Group procedures for module development [28, 29] with patients completing the phase IV QLQ-OH module [26] and the EORTC QLQ-C30 [27]. The module [26] encompassed 17 questions, with four hypothesized scales: pain and discomfort (6 items), xerostomia (2 items), eating (4 items), and information (2 items) and two single items: future worries and use of and problems with dentures.

The 30-item EORTC QLQ-C30 contains five functional scales: physical, role, cognitive, emotional, and social; three symptom scales: fatigue, nausea/vomiting, and pain; and six single items [27]. All but two of the EORTC QLQ-C30 items, Global health/QoL-scale scored from 1 to 7 and the dichotomous module-item on use of dentures, are scored from 1 "not at all" to 4 "very much". Higher item scores represent better function regarding functioning and global health and more symptoms/problems on the other items and on the QLQ-OH module. All scores were linearly transformed to a 0–100 scale using EORTC guidelines [29].

Patients

Patients were recruited from 14 institutions in 10 countries between November 2012 and August 2014. Eligibility criteria were age ≥ 18 years, heterogeneous cancer diagnoses, language fluency, consent, in active treatment, or ≤3 years post-

treatment. Patients with terminal disease or obvious cognitive impairment according to standard clinical criteria; disturbed consciousness, disorientation to time/place, and attention deficits were ineligible.

A sampling matrix was used to ensure a wide distribution of diagnoses, treatment phases, and socio-demographics and included the following five patient groups: (A) in active curative treatment, (B) 2–6 months after cancer treatment, (C) 6 months–3 years after cancer treatment, (D) receiving palliative treatment, and (E) referred to hospital dentist/oral health team. The sample size chosen aimed to satisfy (1) the 'rule of thumb' of 5–10 respondents per item for efficient factor analysis; (2) sufficient to generate ample patient-group sizes to enable item response theory (IRT) methods to analyze differential item functioning (DIF) associated with groups A–D above, plus at least 50 in group E, if applicable; and (3) to ensure adequate patient groups for stability and responsiveness to change analyses (RCA).

Methods

Eligible in- and out-patients were approached and informed by the study personnel. Participants completed the QLQ-OH module [26] and the EORTC QLQ-C30 [27], prior to a set of debriefing questions regarding the module's clarity of wording, whether questions were perceived as intrusive, difficult or irrelevant, and additional comments. Study personnel completed a form on socio-demographic and medical variables including Karnofsky performance status [30].

A subset of patients (n = 177) completed the forms twice, with a 2-week time span. Test–retest reliability was assessed in 60 patients whose oral health issues were not expected to change, while RCA was evaluated in 117 patients undergoing therapy known to negatively affect oral health. A 2-week interval between the RCA assessments was deemed adequate based on clinical experience.

Ethics approval followed national/local requirements. Informed consent was obtained from all participants. The study was registered in ClinTrials.gov (Protocol 2012/1390REK).

Data analyses and item selection

The validation dataset for the QLQ-OH module was prepared in IBM SPSS Statistics v.21 for Windows (IBM Corporation, Armonk, NY) and variables screened for missing values. Preliminary descriptive analysis of responses to the 17 items was conducted and checked for severe restriction in range; that is, where only two responses accounted for more than 95 % of respondents [31]. Using a combination of techniques from classical test theory and IRT, the structure and psychometric properties of the hypothesized scales were analyzed.

Principal components analysis

The QLQ-OH module comprised 12 items scored "during the last week," three items scored "during the course of the illness," and two items related to dentures. Using the 12 items scored in the same time frame, PCA with oblimin rotation was chosen to identify potential items to form scales. Respondents with missing responses for more than 10 % of the items in the module were omitted during this stage of the analysis. This preserved a complete and unbiased dataset during the exploratory factor analysis stage; preferred to the use of imputation methods. Initial eigenvalues (>1) were inspected to assess the optimum number of factors, with a threshold value of 0.4 used for item loading coefficients in the analysis. Scale reliability was then assessed using Cronbach's alpha coefficients.

RUMM 2030 software (RUMM Laboratory Pty Ltd., Australia) was then used to test the unidimensionality of subscales identified in the factor analysis. The default procedure for RUMM 2030 uses the partial credit model, which allows items to have varying numbers of response categories and does not assume the distance between response thresholds is uniform. The following summary statistics were used to assess model fit, using established guidelines [32]. A well-fitting solution would be indicated by a probability from the item-trait interaction chi-square greater than 0.05, with Bonferroni correction. Due to the sensitivity of the chi-square statistic with large sample sizes, an adjusted chi-square was adopted for a sample size of 300. Fit residual values, for both person (PFR) and item (IFR), were inspected; a mean close to zero and a SD less than 1.5 was desirable. Individual item fit residual values greater than +2.5 were taken to indicate misfit and less than −2.5 to indicate item redundancy. Internal consistency was assessed using the person separation index (PSI) with values above 0.7 considered desirable for group level analysis. Threshold maps were inspected for noteworthy disordering, which would indicate inconsistent use of the response options. Rescoring was considered if a significant improvement in model fit was seen.

Differential item functioning (DIF) was checked for possible item bias, caused by the responses by different groups in the sample: sex, age group, and treatment group. Person item threshold maps were plotted to assess whether the scales appropriately targeted the respondent group. Lastly, dimensionality was assessed using equating t tests to compare person estimates derived from the two most disparate subsets of scale items [33]. A threshold level of less than 5 % was considered acceptable. Results from the PCA and Rasch analyses were then combined to establish a solution for a set of scales which provided the best overall fit and optimal psychometric properties.

Results

Patients

Overall, 585 patients from 14 centers in 10 countries: France, Germany (2), Greece, Israel, Italy, Netherlands (2), Norway (2), Poland (2), Sweden, and UK were included, varying from 35 (Greece) to 102 (Poland). For 13 records, more than 20 % of values were missing for the items of interest for potential scales, e.g., 10 patients had a feeding tube; thus, the eating items were not applicable.

A core dataset of 572 patients (98 %), 54 % females, mean age 60.4 (SD12.9), remained for analyses, with occasional missing values acceptable for demographic and clinical variables, Table 1. The majority were married or living with partner (70 %), 53 % were outpatients. The most frequent diagnosis was head-and-neck cancer (21 %), followed by breast cancer (15 %). Forty-five percent had disseminated or metastatic disease, with metastases to the lymph nodes (10 %) or bones (8 %) being most frequent. Comorbidities were present in 51 %, with two or more in 14 %. Heart disease and/or hypertension were most prevalent (n = 109/20 %). No significant differences were found between those who were included and those who were not. The following groups were analyzed for DIF: sex, age groups (≤50, 51–60, 61–70, 71+), treatment group (in active treatment or not), and treatment intent (curative vs. not curative).

Table 1 Sociodemographic and medical characteristics

N = 572	Median	Range
Age	60.3	18–92
Karnofsky score (missing 6)	83	20–100
	N	%[a]
Gender		
Male	265	46
Female	307	54
Educational attainment (missing 12)		
Compulsory school education or less	141	25
Post compulsory school, below university level	260	45
University level	159	28
Employment situation (missing 11)		
Retired	245	43
On sick leave	140	25
Full-time/part-time paid work	103	18
Homemaker	27	4
Unemployed	24	4
Other, incl. students	22	3
Diagnosis		
Head and neck	117	21
Breast	85	15
Lymphoma/leukemia	81	15
Genitourinary, incl. prostate	67	12
Gastro-intestinal	64	11
Gynecological	52	9
Lung	51	8
Myeloma	18	3
Malignant melanoma	15	2
Other[b]	22	4
Disease stage		
Metastatic/disseminated	260	45
Treatment, ongoing or past 2 months		
No treatment	102	18
Chemotherapy (CT) only	126	22
Surgery + CT + RT	95	17
Surgery + CT	55	10
RT and CT	51	9
Surgery only	44	7
Radiotherapy (RT) only	42	7
Surgery + RT	26	4
Stem cell transplantation, incl. CT	20	3
Hormonal therapy only	4	1
Other[c]	7	2
Time of assessment(missing 2)		
In active treatment	382	67
>2 months post-treatment	188	33
Comorbidity (missing 5)		
Yes	290	51

[a] Percentages refer to overall number of respondents

[b] Includes cerebral tumors (6), sarcoma (5), thyroid (5), unknown (3), neuroendocrine (2), malignant thymoma (1)

[c] Includes symptomatic treatment (3), biological treatment only (3) expectative (1)

Acceptability

Five hundred forty-nine of 572 patients (96 %) were interviewed, varying from 94 to 100 % per country. Completion took <10 min for 58 %. Assistance was provided to 21 % (n = 114), primarily with reading and/or writing (n = 96/84 %). Forty-five patients (8 %) marked one or more items as confusing or difficult to answer; 0–20 per country. The most frequently endorsed items were *satisfaction with information*, *sensitivity to food and drink*, and *sticky saliva*. Also, 75 (14 %) patients had provided free comments on specific items. Dichotomous answer categories were suggested for the information item (n = 12), and five patients suggested dropping these items. General comments were provided by 75 patients, primarily related to satisfaction with the content and that these issues were addressed (40 %).

Scale structure and reliability

Assessment of the item responses

The dataset was screened for missing responses to the scoring items of the QLQ-OH module. Four items: *sensitivity to food/drink*, *taste change*, *eating solid food*, *satisfaction with information* had a high proportion of missing values. Two items showed significant restriction in range with low endorsement by patients in this dataset; over 70 % of patients did not report either bleeding gums or having lip sores. However, due to their obvious clinical significance in certain groups of patients, these were retained for further examination in the psychometric analyses.

Principal components analysis

In phase III, the three hypothesized scales (12 items) using a four-point scale, "during the last week" exhibited good internal consistency and reliability [26]. The first principal components analysis (PCA) on the phase IV dataset suggested a two-factor structure, accounting for 55.2 % of the variance. The first factor had a comparatively high eigenvalue (5.26) with the second (1.36) and subsequent factors having small eigenvalues. Inspection of the pattern matrix showed a fair degree of cross-loading for three items between the two factors (Table 2), supporting yet another hypothesis, that the QLQ-OH module could be unidimensional. The information scale, using the timeframe "during the course of your illness" and the dichotomous item *use of dentures* were analyzed separately.

Rasch analysis

First, the two factors identified in Table 2 were tested for goodness of fit (GOF) to the Rasch model. Summary statistics indicated that for factor 1, two items needed to be removed to improve fit whereas for factor 2, one item needed to be removed. Second, all 12 items were tested together to test for unidimensionality. At each step, the item with the greatest misfit or greatest redundancy was removed. Items were removed in the following order, following standard methodology: (1) *sensitivity to food and drink*, fit residual (FR) = −4.191, (2) *problems enjoying meals*, FR = −3.510, (3) *soreness in mouth*, FR = −3.301, and lastly (4) *sticky saliva*, FR = −2.779.

Eight items formed a clinically useful scale (named OH-QoL) with good fit to the Rasch model (overall chi-square—69.6, df—64, p = 0.295/8 = 0.037). These items, all scored on the conventional EORTC four-point scale were *pain in gums*, *bleeding gums*, *lip sores*, *problems teeth*, *sore in mouth corners*, *dry mouth*, *taste change*, and *problems eating solid food*, with an acceptable Cronbach's alpha coefficient of 0.786. Inspection of the threshold map revealed slight disordering of thresholds for three items; these could be explained by small frequencies in some categories.

Table 2 Principal components analysis; factor loading coefficients[a]

	Component	
Items	1	2
Problems enjoying meals	0.882	
Taste change	0.871	
Problems eating solid food	0.738	
Sticky saliva	0.722	
Sensitivity to food/drink	0.719	
Dry mouth	0.680	
Soreness mouth	0.506	
Pain in gums		0.741
Bleeding gums		0.727
Problems with teeth		0.639
Sore in mouth corners		0.528
Lip sores		0.387

[a] Table 2 shows the factor loading coefficients for each item for the first two components in the solution with eigenvalues greater than 1

Optimum solution

The statistical analyses conducted in this study served to complement one another. Taking into account the cross-loading of items across the two factors in the PCA and the subsequent results of the Rasch analyses (both on the individual factors and the combined items), the optimum solution adopted for the module was the QLQ-OH15 questionnaire, with an OH-QoL score (8 items), information scale (2 items), scale regarding dentures (2 items), and three single items (*sticky saliva*/*mouth soreness*/*sensitivity to food*/*drink*). In line with other EORTC QOL modules, the overall total score of the 8 items was standardized to a scale from 0-100, 100 meaning highest QOL (lowest symptom burden). Table 3 displays the correlation matrix showing the item-by-item correlation for the overall eight-item OH-QoL score. Fit statistics to the Rasch model were all within accepted limits: person fit residual (PFR) (SD = 0.950, mean = −0.307), item fit residual (IFR) (SD = 1.462, mean = −0.893), and a PSI of 0.600. The percentage of equating t tests was below the 5 % threshold [1.57 %] and no DIF for sex, age, or curative vs. non-curative treatment (p > 0.003, Bonferroni correction). Only one item *taste change* showed slight uniform DIF for treatment group (active treatment vs. not).

Known group comparisons

There were no significant differences in the OH-QoL score for sex, age group, treatment group (curative or not), or whether satisfied with information given. However, there were highly significant differences in the overall OH-QoL score as to the extent of the patients' sore mouth, problems with dentures, and problems with sticky saliva (p < 0.003). The head-and-neck patients scored lower on the OH-QoL score (more problems) than the patient group with other cancers (Fig. 1). The OH-QoL score also varied significantly according to patient performance status (worse Karnofsky score = lower QoL) (Fig. 2).

Table 3 Item-by-item correlation for the eight-item OH-QoL score

		Pain in gums	Bleeding gums	Lip sores	Problems with teeth	Sore in mouth corners	Dry mouth	Taste change	Problems eating solid food
Bleeding gums	Pearson correlation	.446**	1						
Lip sores	Pearson correlation	.381**	.247**	1					
Problems with teeth	Pearson correlation	.395**	.240**	.195**	1				
Sore in mouth corners	Pearson correlation	.457**	.294**	.475**	.264**	1			
Dry mouth	Pearson correlation	.288**	.154**	.351**	.178**	.361**	1		
Taste change	Pearson correlation	.307**	.156**	.311**	.165**	.353**	.458**	1	
Problems eating solid food	Pearson correlation	.361**	.210**	.365**	.241**	.414**	.407**	.506**	1

**Correlation is significant at the 0.01 level (2-tailed)

* *Boxplot with medians (interquartile ranges). Lower OH-QoL scores indicate more oral health related problems Test–retest and responsiveness*

Test–retest validity collected 2 weeks apart (n = 60) revealed no significant differences in responses over time (Wilcoxon matched paired signed ranks test, z = 0.229, p = 0.82). Responsiveness was tested in 117 patients with varying diagnoses undergoing therapy with potential oral adverse effects. Consistently higher levels of oral problems were reported at the second assessment, albeit not statistically significant (z = 1.904, p = 0.056).

The correlation between the overall OH-QoL score and the EORTC QLQ-C30 scores showed mild to moderate correlations; 0.3 to 0.4, $p < 0.05$, Table 4. Scores on the overall OH-QoL scale showed a lower QoL score (higher symptom burden) for head and neck (mean: 71), compared to other cancers (mean: 78), Table 4, and the QLQ-OH15 single items showed a higher symptom burden in the head-and-neck group.

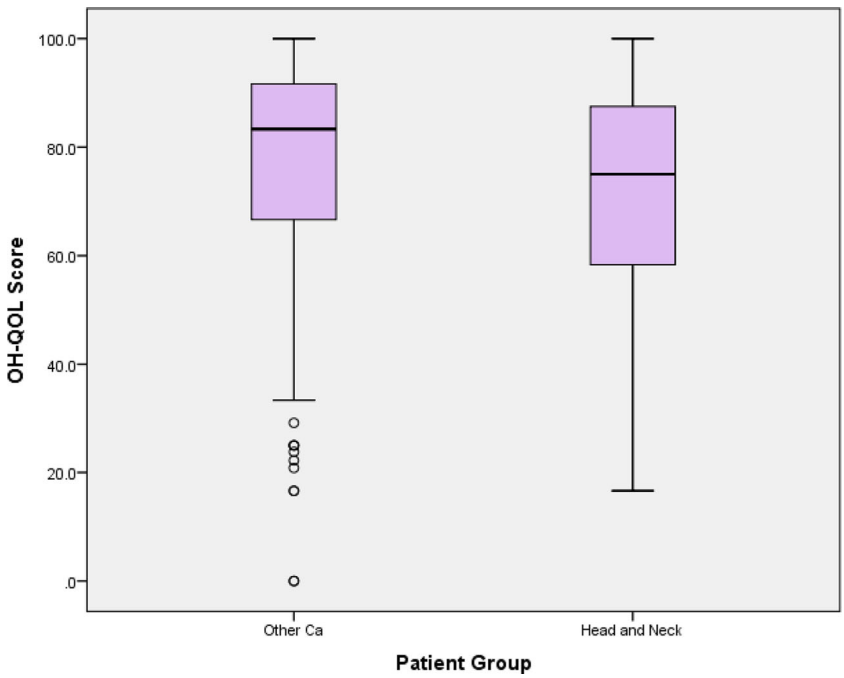

Fig. 1 Median OH-QOL scores between patients with Head-and-Neck cancer (n = 117) vs. other cancers (n = 455). Boxplot with medians (interquartile ranges). Lower OH-QoL scores (0-100) indicate more oral health related problems

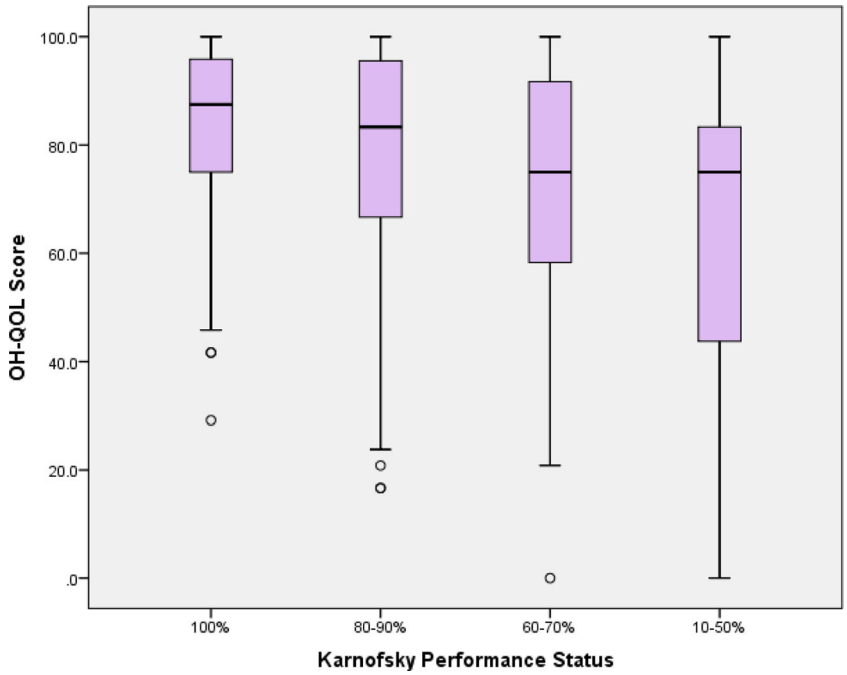

Fig. 2 Median OH-QOL scores split by Karnofsky Performance Status Score. Boxplot with medians (interquartile ranges). Lower OH-QoL scores (0-100) indicate more oral health related problems

Table 4 Correlation between transformed scores on the QLQ-C30 and the QLQ-OH15, all diagnoses versus head-and-neck

	All diagnoses				Head and neck				
	Mean	SD	Median	Valid *N*	Mean	SD	Median	Valid *N*	Correlation coefficient[a]
EORTC QLQ-C30									
Physical[b]	65	27.0	67	455	75	22.8	80	117	.371*
Role[b]	54	35.4	50	455	68	35.2	67	115	.416*
Emotional[b]	71	24.3	75	452	66	26.8	67	117	.344*
Cognitive[b]	76	25.0	83	451	78	23.1	83	117	.334*
Social[b]	60	33.2	67	452	65	33.9	67	117	.442*
Fatigue[c]	51	28.7	44	454	38	30.5	33	117	−.452*
Nausea/vomiting[c]	18	25.7	0	455	11	24.3	0	117	−.369*
Pain[c]	35	32.7	33	455	30	31.7	17	117	−.399*
Dyspnea[c]	28	30.1	33	455	26	28.5	33	117	−.224*
Insomnia[c]	35	33.4	33	454	33	33.6	33	117	−.296*
Appetite loss[c]	29	35.2	0	455	27	36.0	0	117	−.467*
Constipation[c]	27	33.4	0	450	17	29.0	0	116	−.268*
Diarrhea[c]	19	29.1	0	449	11	22.0	0	117	−.233*
Financial[c]	22	32.0	0	451	33	36.8	33	114	−.238*
Global health status[b]	55	23.0	58	451	54	25.6	58	117	.385*
EORTC QLQ-OH15									
OH-QoL score[d]	78	18.9	83	455	71	22.4	75	71	
Sticky saliva[c]	25	32.1	0	449	38	39.4	33	38	
Sensitivity to food and drink[e]	27	33.4	0	453	35	38.2	33	35	
Sore mouth[e]	21	31.1	0	449	34	36.0	33	34	
Information[e]	40	37.4	33	438	53	37.8	67	53	

The negative direction denotes lower QoL scores with higher symptom burden on the QLQ-OH15

*p values <0.001

[a] Spearman rank correlation coefficients for subscales of the QLQ-C30 with the QLQ-OH QoL score

[b] EORTC QLQ-C30: higher scores on the function scales and global QoL denote better function/global QoL [27]

[c] EORTC QLQ-C30: higher scores on the symptom scales and single items denote higher symptom burden [27]

[d] OH-QoL score, eight-item scale of the EORTC QLQ-OH15; higher score denotes better QoL (lower symptom burden) [27]

[e] Single items of the EORTC QLQ-OH15; higher score denotes higher symptom burden [27]

Discussion

This study represents the final phase of the EORTC module development process and investigates the reliability, validity, and psychometric properties of an EORTC QLQ-OH module in an international heterogeneous sample of cancer patients. Two items were removed from the phase III module, due to statistical misfit and patient feedback, yielding a 15-item questionnaire: the QLQ-OH15, containing one eight-item OH-QoL scale, three single items (*sticky saliva*/*mouth soreness*/*sensitivity to food*/*drink*), and two two-item contingency scales regarding use (yes/no) and problems with dentures and reception of (yes/no) and satisfaction with information. Patients' appreciative comments on the debriefing forms indicate that the instrument was well understood and perceived relevant.

The standardized cross-cultural development of questionnaires under the EORTC umbrella ensures the identification of issues perceived as relevant by patients and the necessary psychometric properties for international use. No apparent cross-cultural differences were observed in this study. One item, *sticky saliva*, was reported as difficult by patients, particularly among Swedish patients. This item was taken from the EORTC item-bank, used in the head-and-neck module [34] since 1994 with no reported problems, and no mistakes were identified in the Swedish translation. Because of the high clinical importance, this item was retained, as was the item, *sensitivity to food and drink*, also from the item bank. Patient feedback resulted in a change of answer categories from 1 to 4 to yes/no on the first information item, *received information*, offering a skip option for the subsequent, *satisfaction with information*.

Although the hypothesized scale structure of the QLQ-OH module during phase III needed some refinement, the items developed remained robust during phase IV. This study demonstrates the powerful combination of classical test theory and IRT in the development of new scales. The eight-item scale represents an overall OH-QoL scale that is influenced by the oral health status. Thus, the final OH15 module has three multi-item scales: the OH-QoL, the information scale, and one regarding use and problems with dentures, supplemented by three single items on symptoms, all perceived relevant by patients and clinicians during the stepwise development. When used in conjunction with EORTC QLQ-C30 as by convention for EORTC modules, the multidimensional concept of QoL is well addressed, e.g., how oral problems may influence social activities and functioning. Thus, we regard the QLQ-OH15 as an overall screening instrument for QoL related to oral health. It should be noted that the measurement properties of the eight items constituting the OH-QoL score are not maintained if split into subscales. The items may be used to assess the frequency or severity of these issues, if solely based on the 1–4 raw scores, although we do not recommend this. As opposed to some of the other QLQ modules, e.g., the elderly and social well-being modules [35, 36], the QLQ-OH15 may be viewed as having a predominantly physical focus. In our opinion, this is no drawback, as the initial idea originated from clinical practice in an oncology oral health team. A brief, easy-to-use assessment tool may improve the awareness of oral problems in all cancers among healthcare providers and patients. Thus, preventative and supportive care actions can be taken during treatment and follow-up, e.g., alleviation of mucositis and dry mouth; early detection and treatment of oral mucosal infections, periodontal diseases, and caries; adjustment of ill-fitting prostheses; dietary counseling; etc.

However, assessment tools have little value if they are not perceived as relevant by the users, are unable to discriminate between groups that are perceived as different with respect to symptom intensity, or are insensitive to change over time. All these requirements were met in the present study. No significant differences in scores on the oral health issues were found with demographic- or treatment-related variables, supporting the discriminant and criterion validity. Internal consistency was acceptable, test–retest results showed no significant differences whereas responsiveness was shown in patients whose oral health was expected to change over time. As the primary intention of the QLQ-OH15 development was to produce a clinically useful tool, a number of clinical hypotheses were also investigated, showing that patients with head-and-neck cancer, those with lower performance status, sore mouth and sticky saliva, and problems with dentures had significantly lower OH-QoL scores compared to others (p values <0.001). Although most people today are dentate, ill-fitting dentures that may lead to nutritional problems should be acknowledged, especially among the elderly.

One limitation of the present study may be that inspections of the threshold map revealed slight disordering of thresholds for two items in the OH-QoL score. This could be explained, for example, by low incidence of bleeding gums in these patients, despite the large sample size. On the other hand, this may also occur with even larger samples, unless a strict stratification or a very detailed inclusion matrix was applied. In hindsight, responsiveness related to one particular cancer treatment or diagnosis could have been investigated in more detail. However, our experience with international, multicenter studies shows that researchers' access to patients and diagnostic groups varies, and that may influence patient recruitment.

Statistical study strengths relate to the large sample size and the utilization of the combination of CTT and IRT methods, as best practice in scale development. Overall study strengths are

the cross-cultural validation and systematic development according to established EORTC guidelines, the apparent clinical validity and applicability across treatment phases and cancer diagnoses, and the positive patient feedback. The EORTC Quality of Life Group supports the development of symptom-based questionnaires (SBQs) focusing on side effects related to new treatment regimens. The QLQ-OH15 module fits well with this, as a review reported substantial differences in oral mucositis and stomatitis in cancer patients treated with different tyrosine kinase inhibitors [37]. Also, a recent randomized trial demonstrated more QoL improvements in cancer patients undergoing systematic monitoring of PROs compared to those being monitored at the discretion of the clinicians [38], thereby demonstrating a beneficial effect on clinical outcomes.

Conclusion

The results from this large-scale international study support the psychometric properties of the QLQ-OH15 as a clinical instrument for evaluation of oral health issues that may impact on QoL. Its use in conjunction with EORTC QLQ-C30 makes it feasible to assess, treat, or prevent oral problems before, during, and after cancer treatment.

Acknowledgments This project received financial support from the EORTC Quality of Life Group (QLG) Grant No. 001/2012. We express our gratitude to the EORTC QLG Executive Committee for reviewing and approving the phase IV module development report and manuscript.

Special thanks are also given to the study coordinators at the participating centers and the patients who took part in the study, and to Cand. odont. Petter Wilberg for valuable advice in the preparation of this manuscript.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

Financial support This study was supported by Grant No. 001/2012 from the EORTC Quality of Life Group.

References

1. NIH State-of-the-Science Statement on symptom management in cancer: Pain, depression, and fatigue (2002). NIH Consens State Science Statements 19(4): 1–29
2. U.S.Department of Health and Human services NIoH. National Cancer Institute: The NCI strategic plan for leading the nation. To eliminate the suffering and death due to cancer (2006). In: U.S.Department of Health and Human services NIoH, ed., Bethesda, MD, US, Free-eBooks.net
3. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD (2013) Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 309(8): 814–822
4. Jensen SB, Pedersen AM, Vissink A, Andersen E, Brown CG, Davies AN, et al. (2010) A systematic review of salivary gland hypofunction and xerostomia induced by cancer therapies: prevalence, severity and impact on quality of life. Support Care Cancer 18(8):1039–1060
5. Epstein JB, Thariat J, Bensadoun RJ, Barasch A, Murphy BA, Kolnick L, et al. (2012) Oral complications of cancer and cancer therapy: from cancer treatment to survivorship. CA Cancer J Clin 62(6):400–422
6. Villa A, Wolff A, Narayana N, Dawes C, Aframian DJ, Lynge Pedersen AM, Vissink A, et al. (2015) World workshop on oral medicine VI: a systematic review of medication-induced salivary gland dysfunction. Oral Dis. doi:10.1111/odi.12402
7. Epstein JB, Jensen SB (2015) Management of hyposalivation and xerostomia: criteria for treatment strategies. Compend Contin Educ Dent 36(8):600–603
8. Furness S, Worthington HV, Bryan G, Birchenough S, McMillan R (2011) Interventions for the management of dry mouth: topical therapies. Cochrane Database Syst Rev 12:CD008934
9. Furness S, Bryan G, McMillan R, Birchenough S, Worthington HV (2013) Interventions for the management of dry mouth: non-pharmacological interventions. Cochrane Database Syst Rev 9: CD009603
10. Jensen SB, Pedersen AM, Vissink A, Andersen E, Brown CG, Davies AN, et al. (2010) A systematic review of salivary gland hypofunction and xerostomia induced by cancer therapies: management strategies and economic impact. Support Care Cancer 18(8): 1061–1079
11. Heckel M, Stiel S, Ostgathe C (2015) Smell and taste in palliative care: a systematic analysis of literature. Eur Arch Otorhinolaryngol 272(2):279–288
12. Epstein JB, Smutzer G, Doty RL (2016) Understanding the impact of taste changes in oncology care. Supp Care Cancer. doi:10.1007/s00520-016-3083-3088
13. Hong CH, Napenas JJ, Hodgson BD, Stokman MA, Mathers-Stauffer V, Elting LS, et al. (2010) A systematic review of dental disease in patients undergoing cancer therapy. Support Care Cancer 18(8):1007–1021
14. Elad S, Epstein JB, Raber-Durlacher J, Donnelly P, Strahilevitz J (2012) The antimicrobial effect of Iseganan HCl oral solution in patients receiving stomatotoxic chemotherapy: analysis from a multicenter, double-blind, placebo-controlled, randomized, phase III clinical trial. J Oral Pathol Med 41(3):229–234
15. Lalla RV, Latortue MC, Hong CH, Ariyawardana A, D'Amato-Palumbo S, Fischer DJ, et al. (2010) A systematic review of oral fungal infections in patients receiving cancer therapy. Support Care Cancer 18(8):985–992
16. Multinational Association for Supportive Care in Cancer (MASCC). Evidence-based management strategies for oral complication from cancer treatment. http://www.mascc.org/assets/documents/Oral_Care-Summary-Oral_Complications_Systematic_Reviews.pdf. Accessed November 26, 2015
17. Al-Dasooqi N, Sonis ST, Bowen JM, Bateman E, Blijlevens N, Gibson RJ, et al. (2013) Emerging evidence on the pathobiology of mucositis. Support Care Cancer 21(11):3233–3234
18. Riley P, Glenny AM, Worthington HV, Littlewood A, Clarkson JE, McCabe MG (2015) Interventions for preventing oral mucositis in patients with cancer receiving treatment: oral cryotherapy. Cochrane Database Syst Rev 12:CD0115
19. Ruggiero SL, Dodson TB, Fantasia J, Goodday R, Aghaloo TMB, O'Ryan F (2014) American Association of Oral and Maxillofacial Surgeons position paper on medication-related osteonecrosis of the jaw—2014 update. J Oral Maxillofac Surg 72(10):1938–1956

20. Schuurhuis JM, Stokman MA, Witjes MJ, Dijkstra PU, Vissink A, Spijkervet FK (2015) Evidence supporting pre-radiation elimination of oral foci of infection in head and neck cancer patients to prevent oral sequelae. A systematic review. Oral Oncol 51(3):212–220
21. Epstein JB, Parker IR, Epstein MS, Gupta A, Kutis S, Witkowski DM (2007) A survey of National Cancer Institute-designated comprehensive cancer centers' oral health supportive care practices and resources in the USA. Support Care Cancer 15(4):357–362
22. Ohrn KE, Wahlin YB, Sjoden PO, Wahlin AC (1996) Indications for and referrals to oral care for cancer patients in a county hospital. Acta Oncol 35(6):743–748
23. Wilberg P, Hjermstad MJ, Ottesen S, Herlofson BB (2014) Chemotherapy-associated oral sequelae in patients with cancers outside the head and neck region. J Pain Symptom Manag 48(6): 1060–1069
24. Lalla RV, Bowen J, Barasch A, Elting L, Epstein J, Keefe DM, et al. (2014) MASCC/ISOO clinical practice guidelines for the management of mucositis secondary to cancer therapy. Cancer 120(10): 1453–1461
25. Gamper EM, Zabernigg A, Wintner LM, Giesinger JM, Oberguggenberger A, Kemmler G, et al. (2012) Coming to your senses: detecting taste and smell alterations in chemotherapy patients. A systematic review. J Pain Symptom Manag 44(6):880–895
26. Hjermstad MJ, Bergenmar M, Fisher SE, Montel S, Nicolatou-Galitis O, Raber-Durlacher J, et al. (2012) The EORTC QLQ-OH17: a supplementary module to the EORTC QLQ-C30 for assessment of oral health and quality of life in cancer patients. Eur J Cancer 48(14):2203–2211
27. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al (1993). 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 85(5):365–376 (http://groups.eortc.be/QoL/questionnaires_qlqc30.htm)
28. DeWolf L, Koller M, Velikova G, Jokovic A, Scott N, Bottomley A (2009) EORTC quality of life group translation procedure, 3rd edn. EORTC Study Group on Quality of Life, Brussels ISBN 978–2–930064-38-3
29. Johnson C, Aaronson N, Blazeby J, Bottomley A, Fayers P, Koller M, et al. (2011) Guidelines for developing questionnaire modules, 4th edn. EORTC Quality of Life Group, Brussels ISBN 978–2–930064-413
30. Karnofsky D, Abelmann W, Craver L, Burchenal J (1948) The use of nitrogen mustard in the palliative treatment of cancer. Cancer 1(4):634–656
31. Streiner DL, Norman GR (2003) Health measurement scales: a practical guide to their development and use, 3rd edn. Oxford University Press, Oxford
32. Pallant JF, Tennant A (2007) An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 46(Pt 1):1–18
33. Smith EV Jr (2002) Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 3(2):205–231
34. Bjordal K, Ahlner-Elmqvist M, Tollesson E, Jenssen AB, Razavi D, Maher EJ, et al. (1994) Development of a European Organization for Research and Treatment of Cancer (EORTC) questionnaire module to be used in quality of life assessments in head and neck patients. Acta Oncol 33:879–885
35. Vivat B, Young T, Efficace F, Sigurdardottir V, Arraras JI, Asgeirsdottir GH, et al. (2013) Cross-cultural development of the EORTC QLQ-SWB36: a stand-alone measure of spiritual wellbeing for palliative care patients with cancer. Palliat Med 5:457–469
36. Sodergren SC, White A, Efficace F, Sprangers M, Fitzsimmons D, Bottomley A, et al. (2014) Systematic review of the side effects associated with tyrosine kinase inhibitors used in the treatment of gastrointestinal stromal tumours on behalf of the EORTC Quality of Life Group. Crit Rev Oncol Hematol 91(1):35–46
37. Wheelwright S, Darlington AS, Fitzsimmons D, Fayers P, Arraras JI, Bonnetain F, et al. (2013) International validation of the EORTC QLQ-ELD14 questionnaire for assessment of health-related quality of life elderly patients with cancer. Br J Cancer 109(4):852–858
38. Basch E, Deal AM, Kris MG, Scher HI, Hulis CA, Sabbatini P, et al. (2015) Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 34(6):557–565