Leveraging PROs Data to Inform Oncology Clinical Decision-Making: Introducing the FACE-Q Head and Neck Cancer Module

Jennifer R. Cracchiolo, MD[1], **Anne F. Klassen, DPhil**[2], **Danny A. Young-Afat, MD, MHSc**[3], **Claudia R. Albornoz, MD**[4], **Stefan J. Cano, PhD**[5], **Snehal G. Patel, MD**[1], **Andrea L. Pusic, MD, MHS**[6,7], and **Evan Matros, MD, MMSc, MPH**[8]

[1.] Department of Surgery, Head and Neck Service, Memorial Sloan Kettering Cancer Center New York, NY [2.] Department of Pediatrics, McMaster University, Hamilton, ON, Canada [3.] Department of Plastic and Reconstructive Surgery, Amsterdam University Medical Centers, VU University Medical Center, Amsterdam, the Netherlands [4.] Department of Plastic Surgery, Clínica Las Condes, Santiago, Chile [5.] Modus Outcomes, UK [6.] Department of Surgery, Plastic and Reconstructive Surgery Service, Brigham and Women's Hospital, Boston, MA [7.] Harvard University, Boston, MA [8.] Department of Surgery, Plastic and Reconstructive Surgery Service, Memorial Sloan Kettering Cancer Center, New York, NY

Abstract

Background—Existing patient-reported outcome measures (PROMs) used to assess patients with head and neck cancer have methodologic and content deficiencies. The development of a PROM that meets a range of clinical and research needs across head and neck oncology is described.

Methods—After development of the conceptual framework, which involved a literature review, semi-structured patient interviews, and expert input, patients with head and neck cancer treated at Memorial Sloan Kettering were recruited by their surgeon. The FACE-Q Head and Neck Cancer Module was completed by patients in clinic or sent by mail. Rasch measurement theory analysis was used for item selection for final scale development and to examine reliability and validity.

Scale scores for surgical defect and adjuvant therapy were compared with the cohort average to assess clinical applicability.

Results—The sample consisted of 219 patients who completed the draft scales. Fourteen independently functioning scales were analyzed. Item fit was good for all 102 items, and all items had ordered thresholds. Scale reliability was acceptable (person separation index > 0.75 for all scales; Cronbach alpha values were > 0.87 for all scales; test-retest ranged from 0.86–0.96). The scales performed well in a clinically predictable way, demonstrating functional and psychosocial differences across disease sites and with adjuvant therapy.

Conclusions—The scales forming the FACE-Q Head and Neck Cancer Module were found to be clinically relevant and scientifically sound. This new PROM is now validated and ready for use in research and clinical care.

Precis:

Current PROMs to evaluate patients with head and neck cancer are inadequate. We present a new validated PROMs module that can be used in clinical practice and trials.

Keywords

head and neck cancer; patient-reported outcomes; quality of life; Rasch; FACE-Q

INTRODUCTION

Cancers that affect the head and neck are located close to structures involved in fundamental functions, such as swallowing, speaking, eating, and socializing[1–3]. Use of multiple treatment modalities combining surgery, radiotherapy (RT), and chemotherapy have distinct short- and long-term sequelae in each of these domains.[2] Complex reconstructions using free tissue transfer are often required to restore separation between compartments and to improve facial appearance, further impacting health-related quality of life (HR-QoL)[2].

Measurement of treatment success is complex. Quantitative endpoints, such as cancer-related death and recurrence, have been used to measure treatment success in the past. Inadequate metrics, including long-term tracheostomy or gastric feeding tubes, have been used as proxy measures of patient function and HR-QoL after treatment[4, 5]. The National Cancer Institute (NCI) Symptom Management and Quality of Life Steering Community, Head and Neck subcommittee, have recommended assessment of a set of HR-QoL domains to be included in the design of all future clinical trials[6]. In this effort, use of head and neck cancer patient-reported outcome measures (PROMs) that include relevant domains, discriminate across treatment and reconstructive methods, and limiting patient burden is crucial.[1]

Although PROMs for assessing patients with head and neck cancer exist, methodologic and content deficiencies have been identified[7–9]. Many of these head and neck cancer PROMs were developed without patient input in item generation, raising concern about content validity and relevance, and using classical test theory instead of modern psychometric methods, restricting their use to population-based studies and preventing individual patient

assessment for clinical decision-making. Existing PROMs focus on the functional aspects of treatment of head and neck cancer, with little to no assessment of changes in appearance and the psychosocial sequelae. Cancer-related domains affected in head and neck cancer are site-specific; however, current PROMs require the completion of questions across broad generic domains that may not be relevant to all patients. PROMS that allow focused evaluation of specific cancer sites do not exist, which can result in imprecise assessment and contributes to patient survey burden.

To address the limitations of the current PROMs for head and neck cancer, our research team sought to develop and psychometrically validate the FACE-Q Head and Neck Cancer Module. The qualitative phase of this effort has been previously described[8]. The purpose of this paper is to describe the psychometric properties of the field-tested scales. Specifically, the aim of this study overall was to design PROMs that meet a range of clinical and research needs across head and neck oncology.

MATERIALS AND METHODS

The development of the FACE-Q Head and Neck Cancer Module took place in two phases, which are outlined in Figure 1. First, qualitative methods identified core domains. Draft items were then developed and administered. Rasch measurement methods were used to construct item-reduced scales. Clinical trends were analyzed in the test group. Approval from the Memorial Sloan Kettering Cancer Center institutional review board, in accordance with an assurance filed with and approved by the U.S. Department of Health and Human Services, was obtained prior to starting this study. We adhered to guidelines for PROMs development, including the U.S. Food and Drug Administration guidance to industry and other recommended guidelines[10, 11].

Phase 1

We previously described in detail the development of the conceptual framework, which involved a literature review, semi-structured qualitative interviews (n = 25), and input from experts[8]. The conceptual framework incorporated important patient concerns that were categorized into the following domains: 1) altered facial *function* and the psychosocial impact; 2) altered facial *appearance* and the psychosocial impact; and 3) modifiers, such as cancer worry. Qualitative interviews were used to develop and refine a set of scales to measure aspects of the conceptual framework important to people with head and neck cancer. A set of 14 draft scales were developed.

Phase 2

Field-testing of the 14 draft scales was conducted. Patients were invited by their head and neck surgical oncologist or plastic surgeon to participate in the study by completing the FACE-Q Head and Neck Cancer Module. Participants were required to have had, or be planning to undergo, surgery for cancer of the head and neck region. All facial tumor sites and pathologic subtypes were included to ensure a heterogenous population of patients with head and neck cancer. To encourage a high response rate, surveys were completed in clinic or sent by mail to patients and re-sent once as a reminder. Data on each participant's age,

sex, and ethnicity were also collected to characterize the sample. To guide scales development, Rasch measurement theory (RMT) analysis was applied to the pool for draft items. Item reduction with psychometric analysis was used to create the final set of 14 FACE-Q Head and Neck Cancer Module scales.

Statistical Analysis

We performed RMT analysis using RUMM2030 software[12, 13]. RMT analysis examines differences between observed (patient responses) and predicted item responses to determine the extent the data "fit" the mathematical model[14]. The data are evaluated interactively using a range of statistical and graphic tests to examine each item in a scale. Using this combined evidence, a judgment about the quality of the scale is made.

The following steps were performed:

Threshold for item response options: We tested the assumption that the use of response categories scored with successive integers implied a continuum. This was done by examining the ordering of thresholds, to the point of crossover between adjacent response categories (e.g., between 'some of the time' and 'all the time'), to determine whether successive integer scores increased for the construct measured.

Item fit statistics: We examined how items of a scale worked together as a conformable set both clinically and statistically. Three indicators of fit were used: 1) log residuals (item-person interaction); 2) chi-square values (item-trait interaction); and 3) item characteristic curves. Fit statistics are often interpreted together in the context of their clinical usefulness as an item set, but ideally should be between −2.5 and +2.5, and chi-square values non-significant after Bonferroni adjustment.

Targeting and item locations: The items of the scale were targeted to the patient population for which the scale was developed. Targeting was examined by inspecting the spread of person (range of the construct as reported by the sample) and item (range of the construct as measured by the items in the scale). The items of scale were defined on a continuum. Inspecting where the items were located along the continuum showed how the items mapped out a construct. Items in a scale should be evenly spread out over reasonable range and match the range of the construct experienced by the sample.

Dependency: Residual correlations among items in a scale can artificially inflate reliability. Residual correlations between items should ideally be below 0.30.

Person separation index (PSI): The PSI measures error associated with the measurement of individuals in the sample. Higher values represent greater reliability. This measure is comparable to Cronbach alpha. In addition to the RMT analysis, we computed Cronbach alpha[15] for each scale, and we computed interclass correlation coefficients for test-reset data[16]. The Rasch logit scores were transformed from 0 to 100 and used to examine tests of construct validity. For all scales, higher scores indicated better outcomes; only the Cancer Worry scale was in opposition with higher scores indicative of more cancer worry.

Clinical trends were examined by calculating mean Rasch scores for 1) surgical defects (glossectomy, mandibulectomy, maxillectomy, laryngectomy/pharyngectomy, soft tissue/parotidectomy) and 2) adjuvant RT (yes/no), and comparing with the cohort averages.

RESULTS

Paper questionnaire booklets were sent to 305 patients and re-sent once if not returned. Two hundred and nineteen completed the scales, a response rate of 72%. For the assessment of reliability,38 patients completed the scales twice. Table 1 shows the characteristics of the sample.

The RMT analysis provided evidence of the reliability and validity of the FACE-Q Head and Neck Cancer Module scales. In the initial RMT analysis, the 14 draft scales had 149 items, of which 71 items in 12 scales had disorder thresholds. After item reduction, response options in 11 scales (exceptions: appearance distress, speaking distress, and information) were re-scored to form either 3 or 4 response options (Table 2), after which all items had ordered thresholds. The analysis that follows was performed using the re-scored data. Item fit was good, as all 102 items fell within −2.5 to +2.5, and all items showed non-significant chi-square P-values after Bonferroni adjustment. Item residual correlations were above 0.30 for 14 pairs of items, with 2 pairs above 0.40 (range 0.31–0.45).

The item locations defined a clinical hierarchy for each of the scales. In all scales, items were spread out evenly, displaying a good match with the range of the construct experienced by the sample. Relevant information about reliability and indicators of scale performance are shown in Table 2. With extremes included, the PSI was above 0.80 for 10 of 14 scales and above 0.70 for 13 of 14 scales; 1 scale dropped slightly below 0.70, with 10 above 0.80, and the rest above 0.70. Cronbach alpha values were above 0.90 for 11 of 14 scales, 0.94 for smiling distress, and 0.93 for smiling. For all 14 scales, Rasch logit scores were used to transform scale scores into 0 (worse) to 100 (best). Interclass correlation coefficients for test-retest reliability were high and ranged from 0.86–0.98 A summary of all scales in the module are included in Table 3.

To assess function in a clinical setting, we assessed separate scale scores for each surgical defect, and compared those scores with the cohort average (Figure 2). Oral cavity defects (glossectomy, mandibulectomy, and maxillectomy) had worse (55.6) than average outcomes for eating and oral competency (cohort mean = 52.4). Better facial appearance scores were observed in glossectomy, compared with mandibulectomy (65.8 vs 44.2). For patients who were observed compared to those receiving adjuvant RT (Figure 3), scale scores for eating (50.3), oral competency (54.6), saliva (45.6), speech (46.1), swallowing (51.2), and appearance (50.4) were all below the cohort mean suggesting diminished function after RT. For eating, swallowing, appearance, appearance distress, and cancer worry, these were significant ($P = 0.006$, $P = 0.035$, $P < 0.001$, $P < 0.001$, $P = 0.014$).

DISCUSSION

Recognition of the treatment effects of head and neck cancer and improved cure rates has increased awareness about the functional impairment associated with treatment of head and

neck cancers beyond traditional endpoints. To date, PROMs developed for patients with head and neck cancer are both quantitatively and qualitatively inadequate to be used both in clinical trials and clinical practice. The FACE-Q Head and Neck Cancer Module, consisting of 14 independently functioning scales carefully developed and validated following extensive qualitative patient interviews, is a new PROM that measures impact of treatment on function, facial appearance, and psychosocial distress.

Significant advances have occurred in head and neck oncology in recent decades. Overall, treatment for head and neck cancers has intensified, resulting in greater acute and chronic toxicities. Reconstruction using free tissue transfer has revolutionized surgical management, enabling more intense extirpative procedures. Improved understanding of the biology of head and neck cancers, specifically the appreciation of the excellent outcomes in virally induced human papilloma virus-related oropharynx cancer, has highlighted the importance of HR-QoL measurement and the need for treatment de-escalation. RT techniques, such as intensity-modulated RT and proton therapy, aim to reduce treatment-related toxicity by delivering precision RT. Although maintaining excellent oncologic outcomes remains a priority, appreciation for the short- and long-term HR-QoL represents an equally important aspect of head and neck oncology, as patients have clearly indicated during the qualitative interviews.

Patient survey fatigue must also be considered in both routine clinical care and trial development. A modular approach to scale development was taken to provide specific scales that are relevant to the range of head and neck cancer sites, treatment plans (surgery, RT, chemotherapy) and/or reconstructive strategies. In this approach, each scale may be used separately from the other scales, thus enabling administration of only those scales that are relevant for a certain patient or clinical trial question. This allows for focused evaluation in both the clinical and research setting, while preventing survey fatigue. For example, an investigator could compare marginal mandibular nerve weakness (smile scale) between patients undergoing elective neck dissection or sentinel lymph node biopsy independent of swallowing function. To further mitigate survey fatigue, the authors developed the first version of the FACE-Q Head and Neck Cancer Module with plans to later advance to an adaptive format once larger sample sizes have been tested. Computer adaptive testing (CAT) can be used to decrease test burden and improve efficiency without compromising precision or reliability. Using CAT, the number of items faced by the patient can be decreased by 30%–50%, based upon the degree of correlation desired with the full test version.

Adoption of PROMs, specifically, the FACE-Q Head and Neck Cancer Module into clinical care and in the design of clinical trials has the potential to provide unique information and value in oncology practices, justifying cost of implementation. Literature exists supporting patient-reported symptoms and represents the most reliable modality of data collection. Physicians who discuss PROMs with patients during office visits score better on physician communication and shared decision making[17]. Furthermore, PROMs collection in oncology have been linked to improved symptom management, enhanced quality of life, and longer survival[18]. An additional catalyst will be implementation of alternative payment models, such as the Medicare Merit-based Incentive Payment System and bundled payment programs. Collection of PROMs as a measure of quality has been one strategy to measure

value. As pay-for performance initiatives advance, practices and institutions who have implemented robust PROMs programs that have included and validated clinical care measures, such as the FACE-Q Head and Neck Cancer Module, will be poised to compete in the marketplace by demonstrating their focus on the patient-centered outcomes.

Although some existing questionnaires measure the functional impact of treatment for head and neck cancer, assessment of the psychosocial impact of the functional impairment is lacking. For example, whereas speech clarity may be elicited, the effects of social isolation related to unintelligible speech are not measured. The FACE-Q Head and Neck Cancer Module covers both the functional and psychosocial impact of cancer treatment. Another important under-evaluated issue is "cancer worry," which can ameliorate or exacerbate the impact of functional and psychosocial throughout treatment[19]. Shortly after diagnosis and treatment, when cancer worry is high, functional impairment may be of less concern; however, as the disease-free interval increases and cancer worry diminishes, there is greater awareness of functional impairments. The "cancer worry" scale was incorporated into the FACE-Q Head and Neck Cancer Module to control for response shift and allow for interpretation of changes in HR-QoL measurement across time. Lastly, the FACE-Q Head and Neck Cancer Module is the first PROM to examine the impact of changes in facial appearance in this patient population. Landmark work by Funk and colleagues highlighted the broad impact of appearance on general health[1]. Because treatment negatively impacts both the way patients' see themselves as well as social integration, creation and inclusion of this scale allows for a more robust and complete assessment of HR-QoL.

The ability to quantify clinically meaningful treatment effects in clinical practice and trials depends on sound measurement by psychometrically robust and statistically validated instruments. The psychometric analysis provides evidence of the reliability and validity of the FACE-Q Head and Neck Cancer Module scales. Development using RMT has distinct advantages. In contrast to traditional psychometric methods based on the classical test theory, RMT methodology focuses on the association between a person's measurement and the probability of responding to an item, rather than the association between the person's measurement and the observed scale total score. Existing head and neck questionnaires were developed using classical test theory rather than RMT. Many advantages exist with newer approaches, most notably the ability to evaluate individual patients along the clinical care continuum. Instruments derived with classical test theory are validated to be used only for population-level research. In contrast, because surveys developed using RMT can provide information on interval as opposed to ordinal level change, scores can be integrated into clinical practice to track an individual patient's HR-QoL over time. This makes the tool ideal for both clinical practice and research needs. These properties, taken together with the extensive qualitative work, the development of independently functioning scales, and the incorporation of modifiers such as "cancer worry" will enable clinicians to leverage PROMs data to inform clinical decision-making.

Our field-test study provided evidence of reliability and validity in a sample of patients with head and neck cancer with a range of different disease types. Clinically relevant, predictable trends were observed in the study cohort. Patients reported worse eating and oral competency in oral cavity defects, compared with laryngectomy and soft tissue/parotid

resection. As clinically expected, appearance scores were better in the glossectomy group, compared with patients who required a bony resection, such as mandibulectomy. The independently functioning scales also demonstrated the impact of adjuvant RT on physical functioning and appearance. We have demonstrated that these independent functioning scales are reliable, valid, and perform as hypothesized in the clinically relevant directions. These scales can now be used both in research and in clinical care.

There are some limitations to our study. First, our sample was primarily cross-sectional. Although acceptable for instrument development, longitudinal studies are still needed to measure change over time. Also, the relatively small study sample did not allow for assessment of differential item functioning (i.e., whether items in a scale function differently across subgroups within the sample). Future research could examine differential item functioning (e.g., age, sex, disease type) for all scales. While there are limitations in our sample size, in the Rasch model of instrument development, the estimation of item parameters is independent of the sampling distribution of respondents. The emphasis is upon the degree of precision of the item and personal estimates. Therefore, our sample size of 219 patients will give a 99% confidence that the item estimates will be, "for all practical purposes free from bias."[20] Finally, the test population was heterogenous, representing different cancer sub-sites, as well as reconstructive strategies. Although this was done to incorporate a spectrum of patients and treatment options for instrument development, the sample was too small to perform subgroup analyses assessing differences in outcomes between these different surgical techniques.

In conclusion, the NCI has called for PROMs as part of routine patient-reported assessment in all future head and neck clinical trials, underscoring their importance in the management of patients with head and neck cancer. As such, we have developed and validated the FACE-Q Head and Neck Cancer Module using a modern psychometric approach. This new set of scales will allow surgeons, oncologists, and researchers to evaluate patients with head and neck cancer on an individual as well as group level throughout treatment and in clinical trials. The FACE-Q Head and Neck Cancer Module will provide high-quality, evidence-based data from the patient's perspective for de-escalation trials, novel therapeutic agent trials, and the development of new surgical techniques.

Acknowledgments

References

1. Gliklich RE, Goldsmith TA, Funk GF. Are head and neck specific quality of life measures necessary? Head Neck. 1997;19: 474–480. [PubMed: 9278754]

2. El-Deiry M, Funk GF, Nalwa S, et al. Long-term quality of life for surgical and nonsurgical treatment of head and neck cancer. Arch Otolaryngol Head Neck Surg. 2005;131: 879–885. [PubMed: 16230590]

3. Funk GF, Karnell LH, Christensen AJ. Long-term health-related quality of life in survivors of head and neck cancer. Arch Otolaryngol Head Neck Surg. 2012;138: 123–133. [PubMed: 22248560]

4. Peretti G, Piazza C, Penco S, al E. Transoral laser microsurgery as primary treatment for selected T3 glottic and supraglottic cancers. Head Neck. 2016;38: 1107–1112. [PubMed: 26970262]

5. Katsoulakis E, Riaz N, Hu M, et al. Hypopharyngeal squamous cell carcinoma: three-dimensional or intensity-modulated radiotherapy? A single institution's experience. Laryngoscope. 2016;126: 620–626. [PubMed: 26597398]

6. Chera BS, Eisbruch A, Murphy BA, et al. Recommended patient-reported core set of symptoms to measure in head and neck cancer treatment trials. J Natl Cancer Inst. 2014;106.

7. Cohen WA, Albornoz CR, Cordeiro PG, et al. Health-related quality of life following reconstruction for common head and neck surgical defects. Plast Reconstr Surg. 2016;138: 1312–1320. [PubMed: 27879602]

8. Albornoz CR, Pusic AL, Reavey P, et al. Measuring health-related quality of life outcomes in head and neck reconstruction. Clin Plast Surg. 2013;40: 341–349. [PubMed: 23506775]

9. Pusic A, Liu JC, Chen CM, et al. A systematic review of patient-reported outcome measures in head and neck cancer surgery. Otolaryngol Head Neck Surg. 2007;136: 525–535. [PubMed: 17418246]

10. The FDA's drug review process: ensuring drugs are safe and effective. Available from URL: https://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm [accessed [August 1, 2017].

11. Research ISFPaO. Available from URL: https://www.ispor.org/ [accessed August 8, 2017]].

12. Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? Med Care. 2004;42: 16–17.

13. Wright B, Masters G. Rating Scale Analysis: Rasch Measurement. Chicago, IL: MESA, 1982.

14. Andrich D, Sheridan B, Luo G. RUMM 2030. Perth, Australia: RUMM Laboratory, 1997–2013.

15. Li J Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16: 297–334.

16. Steiner DL NG, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford, UK: Oxford University Press, 2008.

17. Freel J, Bellon J, Hanmer. Better Physician Ratings from Discussing PROs with Patients. NEJM Catalyst.

18. Basch E, Deal AM, Dueck AC, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA. 2017;318: 197–198. [PubMed: 28586821]

19. Van Liew JR, Christensen AJ, Howren MB, Hynds Karnell L, Funk GF. Fear of recurrence impacts health-related quality of life and continued tobacco use in head and neck cancer survivors. Health Psychol. 2014;33: 373–381. [PubMed: 23772885]

20. Sample Size and Item Calibration or Person Measure Stability. Available from URL: http://www.rasch.org/rmt/rmt74m.htm [accessed 10/3/18, 2018].

Development of the FACE-Q Head and Neck Cancer Module

Phase 1: Development of the Conceptual Model

Conceptual Model Development

Patient Interview n = 25
Literature Review
Expert Opinion

Altered Function
And
Psychosocial Impact

Altered Appearance
and
Psychosocial Impact

Modifiers
Cancer Worry

Refinement of Draft Scales

Phase 2: Field Testing

Testing of 14 Draft Scales
Participants Completed Draft Scales n = 302

Item Reduction with Psychometric Analysis

Creation of the FACE-Q Head and Neck Module
Including 14 Independently Functioning Scales

Function
1) Eating
2) Oral Competence
3) Salivation
4) Speaking
5) Swallowing
6) Smile
7) Appearance

Psychosocial
1) Eating
2) Drooling Distress
3) Appearance
4) Smiling
5) Speaking
6) Cancer Worry
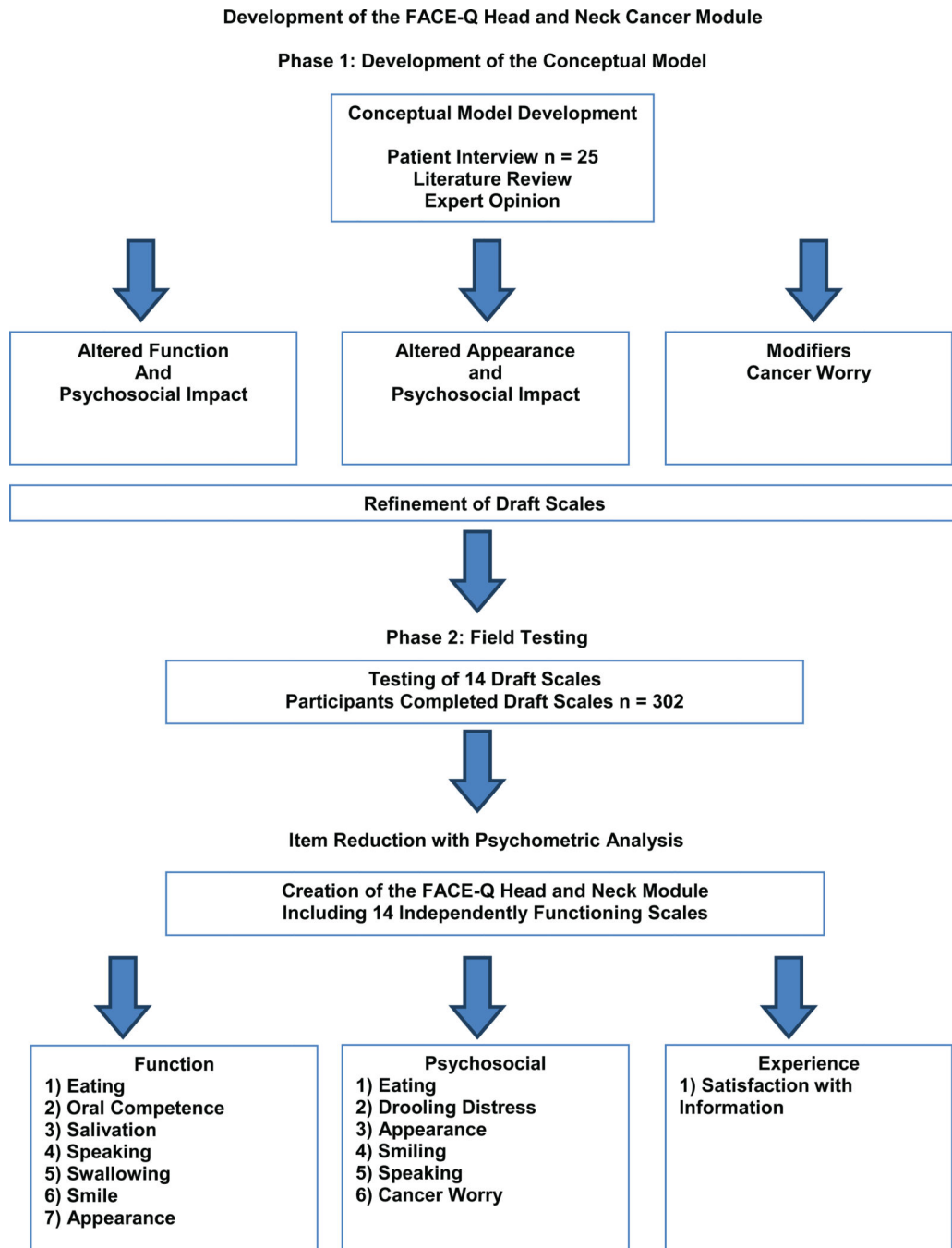
Experience
1) Satisfaction with
Information

Figure 1.
- Development of the FACE-Q Head and Neck Cancer Module.

	Cohort Mean	Glossectomy	Mandibulectomy	Maxillectomy	Laryngectomy/ Pharyngectomy	Soft Tissue/ Parotidectomy
Function Scales						
Eating	52.4	50.5	47.7	51.8	56.6	58.7
Oral Competency	55.7	54.3	45.7	51.1	74.6	65.4
Saliva	46.9	44.0	38.9	44.0	52.3	57.9
Speech	47.7	39.6	48.9	56.5	39.3	57.9
Swallowing	53.1	55.5	49.2	29.4	45.8	64.2
Appearance	64.7	65.8	44.2	47.3	70.0	61.6
Smile	58.8	71.1	51.0		75.3	50.3
Distress Scales						
Eating	61.4	54.6	59.7	65.3	69.2	65.2
Drooling	73.2	70.3	68.9	76.9	78.3	77.6
Speaking	62.8	56.3	57.6	67.0	57.6	67.2
Appearance	64.74	64.0	56.1	54.1	83.6	69.3
Smile	71.2	72.9	69.7		66.0	72.9
Modifiers						
Cancer Worry	31.5	40.0	33.5	28.8	27.2	27.9

Below the mean   Worse
Above the mean   Better
Scale not done

* For all function and distress scales higher score represents better outcome, or function. For the modifier scale, "cancer worry," a higher score represents more worry. n = 165 (glossectomy n = 32, mandibulectomy n = 35, maxillectomy n = 18, laryngectomy/pharyngectomy n = 16, skin/soft tissue/parotidectomy n = 64)

Figure 2.
- Heat map of scale scores for each surgical defect compared to the cohort average postoperatively.

	Cohort Mean	Observation	Adjuvant Radiation	p-value
Function Scales				
Eating	58.1	55.8	50.3	0.006*
Oral Competency	58.9	63.1	54.6	0.105
Saliva	51.7	57.9	45.6	0.139
Speech	53.0	60.0	46.1	0.133
Swallowing	60.4	69.7	51.2	0.035*
Appearance	64.3	78.2	50.4	<0.001*
Smile	57.9	56.6	59.3	0.765
Distress Scales				
Eating	66.3	73.2	59.4	0.094
Drooling	72.4	71.5	73.4	0.830
Speaking	65.3	68.6	62.1	0.391
Appearance	69.2	78.7	59.7	<0.001**
Smile	70.4	69.2	71.7	0.794
Modifiers				
Cancer Worry	29.5	25.2	33.7	0.014*

Below the mean   Worse
Above the mean   Better
Scale not done

* For all function and distress scales, higher score represents better outcome, or function. For the modifier scale, "cancer worry," a higher score represents more worry. **= Independent-samples Mann-Whitney U Test and p values <0.05, *p values <0.05

Figure 3.

- Heat map of scale scores for observation vs. adjuvant RT compared to the cohort average postoperatively.

Table 1.

Patient Characteristics

Characteristics	n (%)
Age	
21–30	5 (2)
31–45	5 (2)
46–60	70 (32)
> 60	139 (64)
Sex	
Male	144 (66)
Female	75 (34)
Race	
Caucasian	190 (87)
Black/African American	11 (5)
Asian or Pacific Islander	11 (5)
South Asian or East Indian	3 (1)
Other	2 (1)
Unknown	2 (1)
Ethnicity	
Non-Hispanic	202 (92)
Hispanic	14 (7)
Unknown	3 (1)
Surgical Treatment	
Mandibulectomy	49 (22)
Maxillectomy	23 (11)
Glossectomy/Floor of mouth	42 (19)
Laryngectomy/Pharyngectomy	22 (10)
Skin/Soft tissue/Parotidectomy	83 (38)

Table 2

Reliability Statistics and Other Indicators of Scale Performance

Scale	N full	N analysis	Items	High scores	Chi-Sq	DF	P-value	Residuals	PSI with extremes		PSI without extremes		Cronbach alpha	TRT	Disordered Thresholds original item set before rescoring	Disordered Thresholds on final item set before rescoring
Appearance	180	122	10	better	28.054	20	0.108	4	0.845	0.787	0.670	0.772	0.97 scored 1,234 / 0.96 rescored 1,233	0.93 n = 23	13/13	10/10
Eating and drinking	138	128	8	better	16.573	16	0.414	1	0.746	0.715	0.654	0.695	0.84 scored 1,234 / 0.80 rescored 1,123	0.96 n = 19	6/9	5/8
Oral competence	152	132	5	better	15.293	10	0.122	0	0.688			0.611	0.81 scored 1,234 / 0.80 rescored 1,123	0.91 n = 22	3/5	3/5
Salivation	156	131	8	better	27.75	16	0.034	0	0.791			0.727	0.92 scored 1,234 / 0.90 rescored 1,123	0.95 n = 23	9/9	8/8
Smiling	110	83	7	better	9.818	14	0.775	0	0.838			0.790	0.93 scored 1,234 / 0.91 rescored 1,123	0.86 n = 16	1/8	1/7
Speaking	147	114	7	better	15.170	14	0.367	3	0.874	0.852	0.816	0.824	0.95 scored 1,234 / 0.94 rescored 1,123	0.92 n = 22	3/10	2/7
Swallowing	126	105	8	better	17.708	16	0.341	1	0.824	0.797	0.668	0.704	0.89 scored 1,234 / 0.89 rescored 1,123	0.98 n = 18	6/11	3/8
Appearance distress	218	141	6	better	10.538	12	0.569	1	0.858	0.844	0.810	0.832	0.94 scored 1,234	0.97 n = 32	5/17	0/6
Eating distress	140	111	7	better	9.902	14	0.769	0	0.846			0.824	0.94 scored 12,345 / 0.92 rescored 12,223	0.96 n = 20	6/8	3/7
Drooling distress	151	78	6	better	23.241	12	0.024	1	0.804	0.729	0.751	0.777	0.96 scored 12,345 / 0.95 rescored 12,223	0.91 n = 22	6/8	5/6
Smiling distress	115	64	5	better	15.872	10	0.103	0	0.765			0.741	0.94 scored 12,345 / 0.93 rescored 12,223	0.87 n = 16	4/5	4/5
Speaking distress	152	117	7	better	9.189	14	0.819	1	0.878	0.860	0.863	0.878	0.95 scored 12,345	0.95 n = 23	3/12	0/7
Cancer worry	221	195	8	worse	39.665	24	0.024	1	0.857	0.834	0.815	0.824	0.91 scored 12,345 / 0.90 rescored 12,334	0.90 n = 36	6/11	4/8
Satisfaction with information	222	122	10	better	29.335	20	0.081	1	0.839	0.829	0.873	0.881	0.96 scored 1,234	0.96 n = 36	0/14	0/10

Abbreviations: DF, degrees of freedom; PSI, person separation index; TRT, test-retest

NOTE: For Cronbach alpha, use the version based on the re-scored data set.

Table 3.

FACE-Q Head and Neck Cancer Module Scales, Including Number of Items, Response Options, FK Grade Level

Name of scale	Name of scale	Number items	Example item	Response options	FK
Function	Eating	8	It takes a lot of effort to eat.	3	7.8
	Oral competence	5	Liquids spill out of my mouth.	3	8.2
	Salivation	8	I am unable to eat without fluids.	3	8.0
	Speaking	7	I have a problem making certain sounds.	3	7.2
	Swallowing	8	Food can get stuck in my throat.	3	8.1
	Smile	7	My smile looks crooked.	3	8.2
	Appearance	10	My face looks damaged.	4	8.4
Psychosocial	Drooling distress	6	I feel insecure because of drooling.	3	8.7
	Eating	7	I avoid eating in public.	3	6.6
	Appearance	6	I worry that my face does not look normal.	4	9.2
	Smiling	5	I avoid smiling.	3	8.2
	Speaking	7	I get frustrated when I have to repeat myself.	5	8.1
	Cancer worry	8	I worry my cancer may come back.	4	9.2
Experience	Satisfaction with information	10	The worst-case scenario?	4	9.6

Abbreviations: FK, - Flesch-Kincaid grade reading level