

FACE-Q Skin Cancer Module for measuring patient-reported outcomes following facial skin cancer surgery*

E.H. Lee ¹, A.F. Klassen,² S.J. Cano,³ K.S. Nehal¹ and A.L. Pusic⁴

¹Dermatology Division, Memorial Sloan Kettering Cancer Center, 16 East 60th St, New York, NY 10022, U.S.A.

²Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada

³Modus Outcomes, Letchworth Garden City, U.K.

⁴Plastic and Reconstructive Surgery Division, Brigham and Women's Hospital, Boston, MA, U.S.A.

Summary

Correspondence

Erica H. Lee.

E-mail: lee@mskcc.org

Accepted for publication

3 April 2018

Funding sources

This research is supported by The Skin Cancer Foundation and in part through the NIH/NCI Cancer Center Support Grant P30 CA008748.

Conflicts of interest

A.L.P., A.F.K. and S.J.C. are co-developers of the FACE-Q, which is owned by Memorial Sloan Kettering Cancer Center.

*Plain language summary available online

DOI 10.1111/bjd.16671

Background The patient's perspective of their facial scar after skin cancer surgery influences perception of care and quality of life (QoL). Appearance satisfaction after surgery is also an important but often overlooked treatment outcome.

Objectives To report the psychometric validation of the FACE-Q Skin Cancer Module consisting of five scales, measuring appearance satisfaction (Satisfaction with Facial Appearance, Appraisal of Scars), QoL (Cancer Worry, Appearance-related Psychosocial Distress) and the patient experience (Satisfaction with Information: Appearance).

Methods Participants underwent Mohs surgery for facial basal or squamous cell carcinoma or excision of early facial melanoma. Cohort 1 received a set of scales before and after surgery. Cohort 2 received the scales on two occasions in the post-operative period for test-retest reliability. Rasch measurement theory was used to select (item-reduce) the most clinically meaningful items for the scales. Reliability, validity, floor and ceiling effects and responsiveness were also analysed.

Results Of 334 patients, 209 (response rate 62.6%) were included. Rasch analysis reduced the total scale items from 77 to 41. All items had ordered thresholds and good psychometric fit. Reliability was high (Person separation index and Cronbach's $\alpha \geq 0.90$) and scales measuring similar constructs were correlated. High floor and ceiling effects were seen for the scales. The Cancer Worry scale demonstrated responsiveness ($P = 0.004$).

Conclusions The FACE-Q Skin Cancer Module meet the requirements of the Rasch model providing linearized measurement. Discriminating between patients with minimal appearance or worry impairment may be a limitation. The scales can be used for larger validation studies, clinical practice and research.

What's already known about this topic?

- Surgical treatment of facial skin cancers leads to scarring and changes in appearance that may negatively impact quality of life.
- The aesthetic outcome of surgery is strongly correlated to patient satisfaction.
- Current patient-reported outcome instruments are limited in their elicitation of appearance and scar satisfaction.

What does this study add?

- The FACE-Q Skin Cancer Module scales measure Facial and Scar Appearance, Appearance-related Psychosocial Distress, Cancer Worry and the patient experience.
- The FACE-Q Skin Cancer Module was validated with modern psychometric methods (i.e. Rasch measurement theory) for greater clinical applicability and individualization of scores.

What are the clinical implications of the work?

- The FACE-Q Skin Cancer Module is recommended for clinical care to improve individual patient outcomes and to inform research studies.

As the incidence of nonmelanoma skin cancer (NMSC) and melanoma has risen globally over the last few decades,^{1,2} surgical management of facial skin malignancies has also increased.³ Studies have shown that skin tumours located on the face and facial scarring are associated with significant psychological morbidity.^{4,5} As physical appearance directly influences emotions and social interactions, scarring from facial surgery can impact an individual's self-perception. The patient's perspective of their aesthetic outcome is becoming increasingly important. However, as appearance is subjective, patient-reported outcome (PRO) instruments are questionnaires designed to quantify the patient's perspective of their disease and/or treatment.

The patient's assessment of their aesthetic outcome after surgery can differ from the surgeon's perspective.^{6,7} Furthermore, this perception markedly influences the patient's satisfaction with the overall care provided.⁸ Patients may be dissatisfied if their scar size is greater than anticipated; others will struggle to adapt to appearance changes, leading to anxiety, social isolation and a decreased quality of life (QoL).^{9,10} Although patients express concerns related to scar appearance, it is minimally addressed by validated measures.¹¹ In a systematic review performed by the authors of PRO instruments used in NMSC, only the Skin Cancer Index addressed appearance, with limited questions about attractiveness and scar size/noticeability.¹² Attributes related to overall appearance and scarring may contribute to dissatisfaction and psychosocial distress yet are not addressed by existing PRO instruments, suggesting that a more comprehensive instrument is needed.

The vast majority of existing PRO instruments were developed and validated with the traditional classic test theory approach; however, newer or modern psychometric approaches (e.g. Rasch Measurement Theory, RMT) improve the clinical interpretation of the scale scores.¹³ In this approach, the qualitative phase is crucial as it provides data that can be used to create the content for a set of independently functioning scales, each of which functions like a 'ruler' whereby the items map out a clinical hierarchy to reflect less of the concept (e.g. cancer worry) at one end of the scale to more of the concept at the other end. The field-test data is then analysed to see if the scale's clinical hierarchy worked as hypothesized. When data collected for a scale meet the requirements of the Rasch model, the scale can be said to provide linearized measurement.

The FACE-Q¹⁴⁻¹⁹ is a multimodule PRO instrument developed for the aesthetic patient undergoing elective surgical and nonsurgical procedures. The FACE-Q Skin Cancer Module was

developed for patients undergoing surgical procedures for facial skin cancers, as additional concerns were identified by our group such as cancer worry and appearance (i.e. scarring), similar to previous research in this area.^{11,20,21} The module consists of five scales that address constructs identified to be important to patients with facial skin cancer.²⁰ In this study, the FACE-Q Skin Cancer Module scales are psychometrically analysed and validated with a modern psychometric approach for clinical use in the population with skin cancer.

Methods

Ethics review board approval was obtained from the Institutional Review Board at Memorial Sloan Kettering Cancer Center (MSKCC). The FACE-Q Skin Cancer Module was developed according to the U.S. Food and Drug Administration guidance to industry and other recommended guidelines for the development of a PRO instrument.²²⁻²⁵ To develop the conceptual framework and scales, a pool of items was generated from three sources: a systematic review of the literature,¹² qualitative interviews and expert opinion. In-depth qualitative interviews were conducted with 15 participants with NMSC and early-stage melanoma to generate the themes most important to the patients. The themes included appearance-related concerns, psychological and social function, adverse problems and the experience of care.²⁰ Participants were also shown the FACE-Q Satisfaction with Facial Appearance and Psychological Distress scales^{14,18} to identify items that might be relevant to patients with skin cancer. Examples of response options for possible inclusion were also reviewed for feedback.

The Face-Q Skin Cancer Module consists of two scales related to appearance, two QoL scales and one patient experience scale (Table 1). All the scales were developed with four response options in keeping with the best practice for scale development.²⁶ The scales are available in Appendix S1 (see Supporting Information). The scales were pilot-tested with five participants to clarify ambiguities, confirm acceptability and completion time. To determine the final number of items in each of the scales and checklists, a field-test study was performed.

Data collection for field-testing (phase II)

We enrolled patients 18 years or older undergoing Mohs surgery for a basal cell carcinoma (BCC) or squamous cell carcinoma of the head and neck region or excision for an early-stage (0 or 1A) melanoma. Participant information such as age,

Table 1 FACE-Q Skin Cancer Scales including number of items and type of response option

Name of scale	Items	Example item	Response option range
Facial Appearance	9	How symmetric your face looks?	Dissatisfied/satisfied
Appraisal of Scars	8	The color of your scar?	Extremely bothered/not at all
Cancer Worry	10	I worry my skin cancer may come back after treatment.	Agree/disagree
Appearance-related Psychosocial Distress	8	I feel self-conscious about how my face looks.	Agree/disagree
Information: Appearance	6	What your scar(s) would look like?	Dissatisfied/satisfied

sex, skin cancer site, diagnosis and type of surgical repair was obtained from the medical records. Recruitment was from July 2014 to July 2015. There were two patient cohorts:

Cohort 1 was approached by the research team in clinic and if enrolled, they received the Cancer Worry and Appearance-related Distress scales before surgery. After surgery, the same group received the same scales and in addition, received the Satisfaction with Facial Appearance, Appraisal of Scars, and Satisfaction with Information: Appearance scales.

Cohort 2 had completed dermatological surgery at MSKCC within 3 years. In the Mohs and Dermatologic Surgery Unit, procedure notes from surgical cases are stored prospectively. Participants for possible inclusion were selected from this database. To reflect clinical practice, a wide range of ages and surgical repair types were included. Participants unable to speak or read English and locations not on facial skin were excluded. This group was mailed the same set of scales (Cancer Worry, Appearance-related Psychosocial Distress, Satisfaction with Facial Appearance, Appraisal of Scars, Satisfaction with Information: Appearance) with instructions to complete one set upon receipt of the package and the second 2 weeks later. The participants were asked to answer the questions in reference to their most recent skin cancer treatment. To increase participation, a gift card incentive was included. The participants also received a reminder letter and telephone call.

The scales are scored separately. For each scale, if missing data comprised less than 50% of the scale's items, the responses were summed. Using a conversion table specific to each scale, the Rasch logit scores were transformed into 0 to 100. Higher scores for the two Appearance scales and the Information scale indicated a better outcome, whereas higher scores for Cancer Worry and Psychosocial Distress scales indicated worse outcomes.

Statistical analysis

RMT analysis^{27,28} was used to select the items for the final versions of the scales with RUMM2030 statistical software.²⁹ The analysis uses a number of statistical and graphic tests to examine each item in a scale^{30–32} and considers the results together when making decisions about the overall scale quality. For the five scales we performed the following:

- 1 **Threshold for item response options:** For each scale, we examined thresholds between response options (e.g. very dissatisfied

and somewhat dissatisfied) to determine if a scale's response categories scored with successive integer scores as intended.

- 2 **Item fit statistics:** The items of a scale must work together (fit) as a set both clinically and statistically. When items do not fit (misfit), it is inappropriate to sum item responses to reach a total score and the validity of the scale is questioned. As there are no absolute criteria for interpreting fit statistics, it is more meaningful to interpret them together and in the context of their clinical usefulness.

Three indicators of fit were assessed: (1) log residuals (item–person interaction), (2) Chi-square (χ^2) values (item–trait interaction) and (3) item characteristic curves. Fit residual should fall between -2.5 and $+2.5$ with associated nonsignificant χ^2 values after Bonferroni adjustment.³⁰

- 1 **Dependency:** Residual correlations between pairs of items were examined to identify any that were 0.30 or higher, as high residual correlations can artificially inflate reliability.
- 2 **Targeting:** Informs about the suitability of the sample for evaluating the scales and how suitable the scale is for measuring the sample. Better targeting allows for an improved ability to interpret the psychometric data with confidence. We examined person and item locations to determine if items were evenly spread across a reasonable range that matched the range of the construct experienced by the sample.
- 3 **Person separation index (PSI):** We examined internal reliability using the PSI, a statistic comparable with Cronbach's α . The PSI measures error associated with the measurement of people in a sample. Higher values indicate greater reliability.

Cronbach's α was computed for each scale, which provides a measure of how closely related a set of items are as a group (internal consistency).³³ For test–retest reliability, interclass correlation (ICC) was computed.³⁴ Floor and ceiling effects for each scale were also calculated.

For responsiveness, we computed group-level change for the two scales (Cancer Worry and Appearance-related Psychosocial Distress) completed by a subgroup before and after surgery. Cancer Worry was anticipated to decrease after surgery, whereas Appearance-related Distress was expected to remain the same or increase after surgery. We compared preoperative and postoperative Rasch transformed scores using

the paired *t*-test and then calculated an effect size³³ [i.e. the mean time 1 (pre-op) minus the mean time 2 (post-op) divided by the SD at time 1]. Cohen criteria were used to interpret the results (0.2, small; 0.5, moderate; 0.8, large).^{35–38}

Construct validity

Pearson correlations were performed to examine associations among scores and two-tailed independent sample *t*-tests to assess for differences among means to test the hypotheses:

- 1 Scales measuring similar constructs (e.g. appearance) would correlate more with each other.
- 2 Higher Cancer Worry scores (more worry) would correlate with higher Appearance-related Psychosocial Distress (more distress). Degree of Cancer Worry may also correlate with Appearance scores with more worry being associated with lower facial satisfaction and more scar bother.

Results

Response rate and sample characteristics

A total of 334 participants were approached, of whom 89 participants were enrolled in cohort 1 and 120 participants in cohort 2 (already had surgery), for an overall response rate of 62.6%. In cohort 1, the participants completed the postoperative scales at a median of 16 days after surgery. In cohort 2, 52 participants filled out the scales a second time for the test-retest (43.3%) at an average of 26 days (median 17 days, range 12–105 days). There was one outlier of 105 days and without this outlier, the maximum number of days was 62. Patient characteristics of the sample are summarized in Table 2.

Rasch measurement theory analysis

The RMT analysis supported the reliability and validity of the five independent scales. The Satisfaction with Facial Appearance scale was reduced from 14 to nine items, Appraisal of Scars from 13 to eight items, Cancer Worry from 15 to 10 items, Appearance-related Psychosocial Distress from 15 to eight items, and Satisfaction with Information: Appearance from 20 to six items. All 41 items had ordered thresholds, which provide evidence that the response options for each scale worked as a continuum that increased for the construct being measured. Only one item was minimally outside the -2.5 to $+2.5$ range (Table S1; see Supporting Information). This item (Information scale: options scar) was retained, given that the other fit statistics were satisfied. The χ^2 *P*-values for all 41 items were nonsignificant, also indicating good item fit. For dependency, the item residuals were above 0.30 for four pairs of items; however, subtests performed revealed only a marginal effect on the scale reliability (0 to 0.01 differences in PSI value). In Figure 1, this histogram shows how well the

Table 2 Patient characteristics of sample (*n* = 209)

Variable	<i>n</i> (%)
Age (years), mean (range)	64 (25–92)
Sex	
Female	113 (54.1)
Male	96 (45.9)
Skin cancer	
Basal cell carcinoma	143 (68.4)
Squamous cell carcinoma	40 (19.1)
Melanoma	25 (12.0)
Other	1 (0.5)
Repair	
Second intention	31 (14.8)
Primary repair	101 (48.3)
Flap repair	54 (25.8)
Skin graft	20 (9.6)
Other (nonsurgical)	3 (1.4)
Scales completed (<i>n</i> = 326) ^a	
Cancer Worry	321 (98.5)
Facial Appearance	229 (70.2)
Scar Appearance	234 (71.8)
Appearance Distress	326 (100.0)
Information	232 (71.2)

^aIncludes participants who completed the scales/checklists more than once.

participants are measured by the items of the Cancer Worry scale. The top histogram shows the frequency of person estimates (or measurements) and the bottom histogram shows the frequency of item thresholds (for each response option for each item). Less Cancer Worry is represented towards the left of the histograms, more to the right. For 88% of the sample there is good coverage of item thresholds, meaning there are items to verbalize the different levels of Cancer Worry for each of these participants. Conversely, 12% of the sample scored at the floor of the scale and, therefore, are not covered by the scale content.

The five scales demonstrate high reliability. The PSI ranged from 0.80 to 0.91 (with extremes) and 0.85 to 0.90 (with no extremes), indicating good internal reliability (Table 3). Cronbach's α values were 0.93 and higher. The ICC value was 0.93 and higher for the Appearance, Scar, Psychosocial Distress and Information scales and 0.76 for Cancer Worry. The Appearance-related Psychosocial Distress scale shows a floor effect of 40% and for the Cancer Worry scale, 15%. The ceiling effects were $> 30\%$ for the two Appearance scales and the Satisfaction with Information scale.

The endorsement frequencies and missing values for the five scales are provided in Table S2 (see Supporting Information). The endorsement frequencies show the clinical hierarchy for each scale. For example, for the Satisfaction with Facial Appearance scale, the clinical hierarchy ranges from the item 'The shape of your face', which had 64% of participants indicate they were very satisfied, to 'How your face looks up close', which had 40% of participants indicate they were very satisfied. Between the two ends of the 'ruler', the remaining

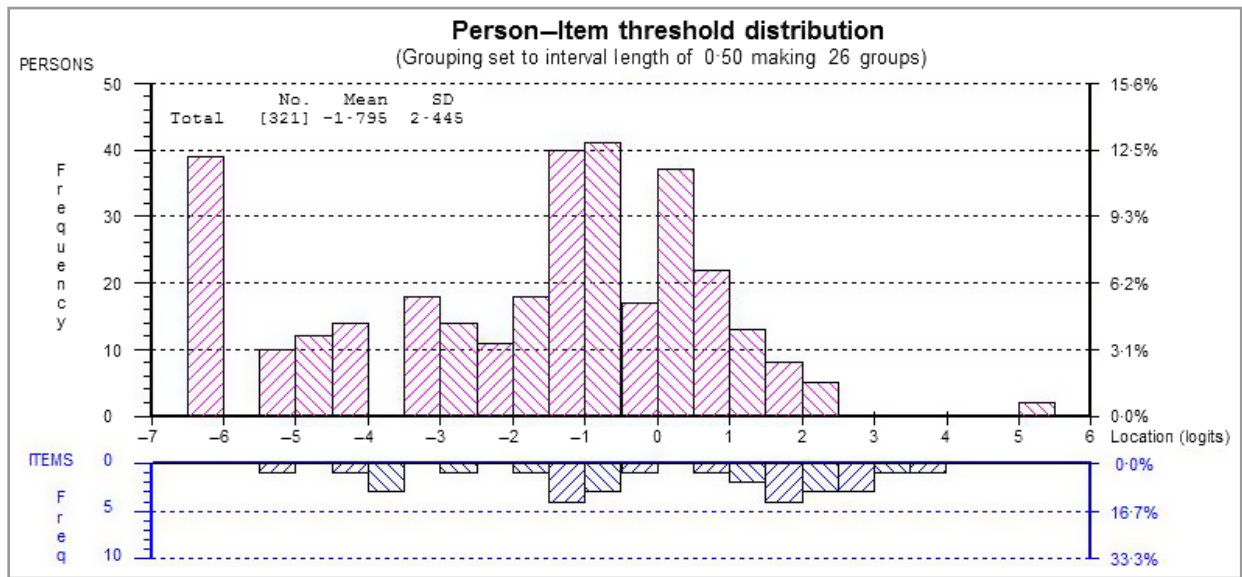


Fig 1. Person–Item threshold distribution for Cancer Worry. The x-axis represents Cancer Worry. The y-axis shows the frequency of person locations (top histogram) and item locations (bottom histogram).

Table 3 Reliability statistics and floor and ceiling effects

Scale	PSI with extremes	PSI with no extremes	Cronbach’s α	ICC	% floor score 0	% ceiling score 100
Appearance	0.90	0.90	0.97	0.95	1.7	32.8
Scar	0.81	0.86	0.94	0.97	0.4	40.6
Worry	0.91	0.87	0.93	0.76	15.3	0.6
Distress	0.80	0.87	0.93	0.98	39.9	0
Information	0.81	0.85	0.95	0.93	1.3	47.6

ICC, interclass correlation coefficients; PSI, person separation index.

eight items mapped out the scales concept in terms of decreasing satisfaction with appearance of the face.

Responsiveness

Sixty-three participants completed the Cancer Worry and Appearance-related Psychosocial Distress scales before and after surgery. Cancer Worry scores changed from a mean of 40.2 (SD 17.6) before surgery to 32.2 (SD 20.8) after surgery ($P = 0.004$, effect size 0.46), which shows a significant improvement after surgery, and a moderate effect size. The Psychosocial Distress scale scores did not have a significant change before [mean score 18.6 (SD 16.1)] and after [19.9 (SD 20.3)] surgery ($P = 0.61$, effect size -0.1).

Construct validity

Pearson correlations between the Appearance [Facial Appearance and Scar (0.55, $P = 0.01$)] and QoL scales [Cancer Worry and Appearance-related Psychosocial Distress (0.39, $P = 0.01$)] correlated and were significant. Higher Cancer Worry scores correlated with more Appearance-related Psychosocial Distress (0.39, $P = 0.01$) and less satisfaction with

Facial Appearance (-0.18 , $P = 0.01$) and more Scar bother after surgery (-0.27 , $P = 0.01$).

Discussion

Facial appearance defines an individual’s identity, and directly influences emotions and social interactions.³⁹ Therefore, visible scarring can lead to anxiety, poor self-esteem and a decreased QoL.⁴⁰ While satisfaction with facial appearance after surgery is an important outcome, it is often overlooked in dermatological surgery. The FACE-Q Skin Cancer Module provides the dermatological and reconstructive surgery communities with a comprehensive set of meaningful and scientifically sound set of scales for the population with facial skin cancer.

The psychometric analyses provide evidence of reliability and validity of the five scales that comprise the FACE-Q Skin Cancer Module. The use of RMT methods has fundamental advantages. The RMT methods differ from traditional psychometric methods (based on classic test theory) because their focus is on the association between a person’s measurement and the probability of responding to an item, rather than the association between a person’s measurement and the observed

scale total score.³⁰ Advantages of using RMT to develop PRO instruments include: (1) RMT provides measurements of people that are independent of the sampling distribution of the people in whom they are developed; (2) RMT improves the potential to diagnose item-level psychometric issues and (3) RMT allows for a more accurate picture of individual person measurements.³⁰ These qualities, together with the qualitative work performed to create the FACE-Q Skin Cancer Module, are what set it apart from other PRO instruments in the same clinical area. A recent Dutch questionnaire also utilized modern psychometrics (item-response theory); however, the appearance questions are limited to scar worry and attractiveness and not specific to scar attributes or overall facial appearance.⁴¹

The two appearance scales ask about satisfaction with facial appearance and the degree of bother for different scar attributes, respectively. The majority of studies assessing QoL in the NMSC population use the Skindex-16,¹² a rigorously developed instrument that focuses on symptoms, emotions and physical/social limitations.⁴² A PRO questionnaire specific to advanced BCC and BCC naevus syndrome showed that scarring was an important patient concern.⁴³ Studies have also demonstrated diminished QoL with scarring and the value of reconstruction of facial defects.^{6,39} Therefore, an instrument that measures facial appearance satisfaction and scar outcome is especially applicable to the skin cancer population.

The Satisfaction with Information: Appearance scale inquires about scar and healing expectations. As patients' perception of their scars influences service perception,⁸ this scale could be used to identify patient education gaps for an individual clinician and for larger quality-improvement efforts. The correlation between scales also showed more worry and appearance-related distress were negatively correlated with appearance satisfaction, suggesting a potential relationship between these constructs for future studies. The scales could explore the driving factors for patient satisfaction; for example, addressing specific scar attributes may impact overall facial satisfaction and in turn, appearance-related distress. An advantage of using Rasch analysis as the statistical model is that the scores are interpreted for the individual person and not for group comparisons only.⁴⁴ The scale scores are interpreted at the individual level to offer tangible and unique clinical benefits for the clinician.

We acknowledge that there could be some bias as the sample is from one institution and participants were not recruited consecutively. However, the responses were varied and a range of patients participated. The number of questionnaires filled out for each scale varied, reflecting the challenges of obtaining follow-up and mail surveys despite including a monetary incentive card. The presence of floor and ceiling effects in the current scales indicate an inability to differentiate between patients with low levels of impairment. It could be argued that this could be due to a limitation in the content validity of these scales. However, our previous research,⁴⁵ and extensive qualitative research developing the current scales, suggests that this is more likely a reflection of the clinical

picture related to low morbidity. Thus, the floor effect for the Cancer Worry scale may reflect the perceived less serious nature of NMSC, for at least a proportion of patients. The observed greater ceiling effects (Facial Appearance, Scar Bother scales) and floor effect (Appearance-related Distress Scale) may reflect minimal impact of surgical scarring and successful healing experienced by many patients. Nevertheless, it is important to bear in mind that our scales are better targeted to those patients with greater impairments related to NMSC.

The FACE-Q Skin Cancer Module is a promising new instrument that can be used by dermatological and plastic surgeons to evaluate their outcomes at the individual patient or larger practice level. The unique inclusion of Facial Appearance and Scar Appraisal and QoL allows for future studies comparing the long-term success of different surgical techniques from the patient's perspective. It could also aid in identifying patients at risk for poor outcomes and dissatisfaction and allow for rigorous studies in different age groups and anatomical locations to further validate the instrument. The FACE-Q Skin Cancer Module can be used for clinical practice, research or quality improvement and may complement existing clinician-based outcomes.

References

- 1 World Health Organization. Ultraviolet radiation: skin cancers. Available at: <http://www.who.int/uv/faq/skincancer/en/index1.html> (last accessed 16 May 2018).
- 2 Rogers HW, Weinstock MA, Feldman SR, Coldiron BM. Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the U.S. population, 2012. *JAMA Dermatol* 2015; **151**:1081–6.
- 3 Donaldson MR, Coldiron BM. Dermatologists perform the majority of cutaneous reconstructions in the Medicare population: numbers and trends from 2004 to 2009. *J Am Acad Dermatol* 2013; **68**:803–8.
- 4 Steinbauer J, Koller M, Kohl E *et al.* Quality of life in health care of non-melanoma skin cancer – results of a pilot study. *J Dtsch Dermatol Ges* 2011; **9**:129–35.
- 5 Dropkin MJ. Body image and quality of life after head and neck cancer surgery. *Cancer Pract* 1999; **7**:309–13.
- 6 Brown BC, McKenna SP, Siddhi K *et al.* The hidden cost of skin scars: quality of life after skin scarring. *J Plast Reconstr Aesthet Surg* 2008; **61**:1049–58.
- 7 Young VL, Hutchison J. Insights into patient and clinician concerns about scar appearance: semiquantitative structured surveys. *Plast Reconstr Surg* 2009; **124**:256–65.
- 8 Dixon AJ, Dixon MP, Dixon JB. Prospective study of long-term patient perceptions of their skin cancer surgery. *J Am Acad Dermatol* 2007; **57**:445–53.
- 9 Cassileth BR, Lusk EJ, Tenaglia AN. Patients' perceptions of the cosmetic impact of melanoma resection. *Plast Reconstr Surg* 1983; **71**:73–5.
- 10 Dropkin MJ. Coping with disfigurement and dysfunction after head and neck cancer surgery: a conceptual framework. *Semin Oncol Nurs* 1989; **5**:213–19.
- 11 Radiotis G, Roberts N, Czajkowska Z *et al.* Nonmelanoma skin cancer: disease-specific quality-of-life concerns and distress. *Oncol Nurs Forum* 2014; **41**:57–65.
- 12 Lee EH, Klassen AF, Nehal KS *et al.* A systematic review of patient-reported outcome instruments of nonmelanoma skin cancer in the dermatologic population. *J Am Acad Dermatol* 2013; **69**:e59–67.

- 13 Cano SJ, Hobart JC. The problem with health measurement. *Patient Prefer Adherence* 2011; **5**:279–90.
- 14 Pusic AL, Klassen AF, Scott AM, Cano SJ. Development and psychometric evaluation of the FACE-Q satisfaction with appearance scale: a new patient-reported outcome instrument for facial aesthetics patients. *Clin Plast Surg* 2013; **40**:249–60.
- 15 Klassen AF, Cano SJ, Scott A *et al.* Measuring patient-reported outcomes in facial aesthetic patients: development of the FACE-Q. *Facial Plast Surg* 2010; **26**:303–9.
- 16 Panchapakesan V, Klassen AF, Cano SJ *et al.* Development and psychometric evaluation of the FACE-Q Aging Appraisal scale and Patient-Perceived Age Visual Analog scale. *Aesthet Surg J* 2013; **33**:1099–109.
- 17 Klassen AF, Cano SJ, Scott AM, Pusic AL. Measuring outcomes that matter to face-lift patients: development and validation of FACE-Q appearance appraisal scales and adverse effects checklist for the lower face and neck. *Plast Reconstr Surg* 2014; **133**:21–30.
- 18 Klassen AF, Cano SJ, Schwitzer JA *et al.* FACE-Q scales for health-related quality of life, early life impact, satisfaction with outcomes, and decision to have treatment: development and validation. *Plast Reconstr Surg* 2015; **135**:375–86.
- 19 Klassen AF, Cano SJ, Schwitzer JA *et al.* Development and psychometric validation of the FACE-Q Skin, Lips, and Facial Rhytids Appearance scales and adverse effects checklists for cosmetic procedures. *JAMA Dermatol* 2016; **152**:443–51.
- 20 Lee EH, Klassen AF, Lawson JL *et al.* Patient experiences and outcomes following facial skin cancer surgery: a qualitative study. *Australasian J Dermatol* 2016; **57**:e100–4.
- 21 Burdon-Jones D, Thomas P, Baker R. Quality of life issues in non-metastatic skin cancer. *Br J Dermatol* 2010; **162**:147–51.
- 22 U.S. Department of Health and Human Services Food and Drug Administration, Centers for: Drug Evaluation and Research; Biologics Evaluation and Research; Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. FDA: 2009. Available at: <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf> (last accessed 16 May 2018).
- 23 Aaronson N, Alonso J, Burnam A *et al.* Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002; **11**:193–205.
- 24 Patrick DL, Burke LB, Gwaltney CJ *et al.* Content validity – establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 1 – eliciting concepts for a new PRO instrument. *Value Health* 2011; **14**:967–77.
- 25 Patrick DL, Burke LB, Gwaltney CJ *et al.* Content validity – establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2 – assessing respondent understanding. *Value Health* 2011; **14**:978–88.
- 26 Khadka J, Gothwal VK, McAlinden C *et al.* The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes* 2012; **10**:80.
- 27 Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? *Med Care* 2004; **42** (Suppl. 1):17–16.
- 28 Wright BD, Masters GN. *Rating Scale Analysis*. Chicago: Mesa Press, 1982.
- 29 Andrich D, Sheridan B. RUMM2030 (software program). Perth, Australia: RUMM Laboratory, 1997–2011.
- 30 Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009; **13**:iii, ix–x, 1–177.
- 31 Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmark: Danish Institute for Education Research, 1960.
- 32 Andrich D. *Rasch Models for Measurement*. Newbury Park, CA: Sage, 1988.
- 33 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; **16**:297–334.
- 34 Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A practical guide to their development and use*. Oxford: Oxford University Press, 2008.
- 35 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; **27**(Suppl. 3):S178–89.
- 36 Cohen J. A power primer. *Psychol Bull* 1992; **112**:155–9.
- 37 Katz JN, Larson MG, Phillips CB *et al.* Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992; **30**:917–25.
- 38 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
- 39 Dey JK, Ishii LE, Joseph AW *et al.* The cost of facial deformity: a health utility and valuation study. *JAMA Facial Plast Surg* 2016; **18**:241–9.
- 40 Sobanko JF, Sarwer DB, Zvargulis Z, Miller CJ. Importance of physical appearance in patients with skin cancer. *Dermatol Surg* 2015; **41**:183–8.
- 41 Waalboer-Spuij R, Hollestein LM, Timman R *et al.* Development and validation of the Basal and Squamous Cell Carcinoma Quality of Life (BaSQoL) questionnaire. *Acta Derm Venereol* 2018; **98**:234–9.
- 42 Chren MM, Lasek RJ, Sahay AP, Sands LP. Measurement properties of Skindex-16: a brief quality-of-life measure for patients with skin diseases. *J Cutan Med Surg* 2001; **5**:105–10.
- 43 Mathias SD, Chren MM, Colwell HH *et al.* Assessing health-related quality of life for advanced basal cell carcinoma and basal cell carcinoma nevus syndrome: development of the first disease-specific patient-reported outcome questionnaires. *JAMA Dermatol* 2014; **150**:169–76.
- 44 Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res* 2011; **11**:571–85.
- 45 Cano SJ, Browne JP, Lamping DL *et al.* The Patient Outcomes of Surgery-Head/Neck (POS-head/neck): a new patient-based outcome measure. *J Plast Reconstr Aesthet Surg* 2006; **59**:65–73.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1 Content of the FACE-Q Skin Cancer Module scales.

Table S1 Rasch measurement theory statistical indicators of fit.

Table S2 Endorsement frequencies and missing values for the five scales.