Construct validity of the anglicised FACE-Q Skin Cancer module

T D Dobbs MA, BM BCh, MRCS, PhD*

1 Reconstructive Surgery & Regenerative Medicine Research Group, Institute Of Life Sciences, Swansea University Medical School, Swansea, UK

2 Welsh Centre for Burns and Plastics, Morriston Hospital, Swansea, UK

C J Harrison BSc MBBS MRCS*

3 Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

M J Ottenhof MD

4 Patient-Reported Outcomes, Value and Experience (PROVE) Centre, Department of Surgery, Brigham and Women's Hospital, Boston, MA, USA

5 Department of Surgery, Harvard Medical School, Boston, MA, USA

J A G Gibson MBBcH, MRCS, MRes

1 Reconstructive Surgery & Regenerative Medicine Research Group, Institute Of Life Sciences, Swansea University Medical School, Swansea, UK

2 Welsh Centre for Burns and Plastics, Morriston Hospital, Swansea, UK

R N Matin PhD FRCP [Derm]

6 Department of Dermatology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

J N Rodrigues BSc(Hons), MBChB(Hons), MSc, MBA, PhD, FRCS(Plast)

7 Department of Plastic and Reconstructive Surgery, Stoke Mandeville Hospital, Aylesbury, UK

H A Hutchings BSc, PhD

8 Patient & Population Health & Informatics Research Group, Institute of Life Sciences 2, Swansea University Medical School, Swansea, UK

I S Whitaker MA (Cantab) MBBChir, PhD, FRCS(Plast)

1 Reconstructive Surgery & Regenerative Medicine Research Group, Institute

Of Life Sciences, Swansea University Medical School, Swansea, UK

2 Welsh Centre for Burns and Plastics, Morriston Hospital, Swansea

*Joint first authors

Meetings

This work has been accepted for presentation at the British Association of Plastic, Reconstructive and Aesthetic Surgeons (BAPRAS) Winter conference 2020.

Abstract

Objectives

The FACE-Q Skin Cancer module is a patient reported outcome measure (PROM) for facial skin cancer. It has been anglicised for the UK population and undergone psychometric testing using classical test theory. In this study further evaluation of construct validity using Rasch measurement theory and hypothesis testing was performed.

Methods

Patients were prospectively recruited to the Patient Reported Outcome Measures In Skin Cancer Reconstruction (PROMISCR) study and asked to complete the anglicised FACE-Q Skin Cancer module. Scalability and unidimensionality of the data were assessed with a Mokken analysis prior to Rasch analysis. Response thresholds, targeting, fit statistics, local dependency and internal consistency were examined for all items and sub-scales. Four a priori hypotheses were tested to evaluate convergent and divergent validity. We additionally hypothesised that median 'cancer worry' score would be lower in post-operative than pre-operative patients.

Results

239 patients self-completed the questionnaire between August 2017 and May 2019. Of the 10 sub-scales assessed, 5 showed relative fit to the Rasch model. Unidimensionality was present for all 5 sub-scales, with most demonstrating ordered item thresholds and appropriate fit statistics. Two items in the 'cancer worry' subscale had either disorded or very close response thresholds. Subscales of the FACE-Q Skin Cancer module demonstrated convergent and

divergent validity with relevant Skin Cancer Index comparators ($p<0.001$). Median 'cancer worry' was lower in post-operative patients (44 vs 39, $p<0.001$).

Conclusion

The anglicised FACE-Q Skin Cancer module shows psychometric validity through hypothesis testing, and both classical and modern test theory.

Key words: Patient reported outcome measures; PROM; skin cancer; FACE-Q; validation

Introduction

Skin cancer is the commonest malignancy worldwide[1], with the majority occurring on sun-exposed, cosmetically sensitive sites such as the face[2]. A diagnosis of skin cancer frequently leads to considerable psychological burden associated with anxiety relating to a cancer diagnosis[3] and concerns regarding visible scarring[4].

Measuring health-related quality of life (HRQoL) before, during and after treatment can positively influence clinical practice. Patient-reported outcome measures (PROMs) can be used to assess HRQoL, but these questionnaires need to be standardised and validated[5,6]. Previously there has been a paucity of appropriately designed and well validated PROMs for facial skin cancer[7], with the FACE-Q Skin Cancer module designed to address this need[8]. In previous published work, we have anglicised the FACE-Q skin cancer module for use in the United Kingdom (UK)[9], in accordance with guidance from the United States (US) Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for PROMs to be appropriately translated and adapted before use[10,11]. The FACE-Q Skin Cancer module subscales previously anglicised and assessed here are composed of subscales designed specifically for facial skin cancer patients along with four subscales that were designed for and validated in the BODY-Q questionnaire ('satisfaction with doctor/surgeon', 'satisfaction with clerical staff', 'satisfaction with medical/ward team' and 'satisfaction with information')[12] (*Table 1*). Initial early validation work for the anglicised FACE-Q skin cancer module, using classical test theory (CTT), supported good face and construct validity, however considerable item redundancy was noted[9]. The consensus-based standards for the selection of health measurement instruments recommend that structural validity is evaluated through modern test theory approaches[13]. In contrast to CTT, which uses correlational statistics to assess structural validity at the scale-level, modern test theory uses probabilistic modelling to assess the

performance of individual items. This can provide more granular insight into the measurement properties of a PROM, and assess an instrument's ability to deliver continuous (as opposed to ordinal) measurement[14]. Construct validity should be further assessed through hypothesis testing[13].

In this study, we aim to further evaluate the construct validity of the anglicised FACE-Q Skin Cancer module through Rasch measurement theory psychometric analysis and hypothesis testing.

Methods

Data were collected as part of the Patient Reported Outcomes In Skin Cancer Reconstruction (PROMISCR) study, a prospective anglicisation and validation study of the FACE-Q Skin Cancer module[15]. Patients with a new diagnosis of facial skin cancer were recruited from two UK centres, the Welsh Centre for Burns and Plastics, Wales and the Department of Dermatology, Oxford. Eligible patients (*Supplementary Figure 1*) were provided with study details and time to consider inclusion before obtaining written consent. Patients were provided with a study pack containing a copy of the anglicised FACE-Q Skin Cancer module (based on the US version and available at *http://qportfolio.org/*) and a copy of the Skin Cancer Index (SCI)[16]. Six to eight weeks post-operatively, the study pack was posted to all patients to be completed again. Research ethics committee approval was granted (REC: 16/WM/0445).

Data collection and statistical analysis

Questionnaires were pseudonymised using a unique patient identifier with data acquisition and storage performed in accordance with the Data Protection Act 1998 and the 2018 General Data Protection Regulation (GDPR). Demographic data were collected for each patient including diagnosis, past medical history, medication and type of reconstruction used.

In Rasch measurement analyses, sample sizes greater than 150 are considered to provide item calibrations or person measures that are stable to ± 0.5 logits with a 99% confidence. Sample sizes of 250 and more can provide definitive item calibrations with over 99% confidence[17]. Missing data were managed using listwise exclusion.

Statistical analysis was conducted in *R (v 4.0.0)* with the following packages: *foreign (v0.8-79), dplyr (v0.8.5), mirt (v 1.32.1), mirtCAT (v 1.10), mokken (v3.0.2), eRm (v 1.0-1), ltm (v1.1-1)* and *WrightMap (v1.2.)*.

Modern test theory analysis

Mokken analysis

Scalability and unidimensionality were assessed with a Mokken analysis, prior to Rasch analysis, to ensure that the item responses followed a probabilistic structure consistent with the Rasch Model. Items with a Loevinger's (H) coefficients of ≥ 0.3 were considered scalable[18].

Response threshold ordering

Item characteristic curves (ICCs) were plotted and item response thresholds were calculated using the *mirt* package[19] in *R* to determine whether response options (e.g. very

dissatisfied and somewhat dissatisfied) were scored successively. Rasch model parameters were generated using a fixed quadrature expectation-maximisation approach (*mirt* version 1.32.1)[19].

Targeting

Item person plots (IPPs) were computed to compare the position of items on the underlying scale to the frequency distribution of respondents on the same scale. This provided an indication of whether the items were targeted to relevant parts of the scale for the people in our sample.

Fit statistics

We assessed the fit of each item to the Rasch model using Chi square tests, along with infit and outfit mean squares. Infit and outfit mean squares are residual fit statistics that measure the standardised residual variance across items and persons, similar to the Residual statistic provided by the RUMM2030 platform and reported previously for FACE-Q items in an American population[8,20]. Infit and outfit mean squares that fell between 0.5 and 1.7, and non-significant ($p > 0.01$) Chi square tests were considered to suggest appropriate fit. Infit and outfit mean squares that are < 0.5 are less productive for measurement but not degrading[21].

Local Dependency

An assumption of the Rasch model is that item responses are only correlated through the latent trait which they intend to measure. Local dependency (LD) occurs where there is residual covariance between two items, suggesting that the items may be very closely related, or unintentionally measuring other latent traits[22]. LD between item pairs was assessed using Yen's Q3 statistic, with a value of > 0.2 used heuristically in the literature to indicate LD[23].

Internal consistency

Cronbach's alpha was calculated for each scale. Values of ≥ 0.7 indicate acceptable internal consistency, with values of >0.9 considered to indicate item redundancy[24].

Hypothesis testing

Five *a priori* hypotheses generated by the authors were tested to evaluate the construct validity of the instrument's subscales:

1) 'Cancer worry' subscale scores from the FACE-Q Skin Cancer module would negatively correlate with 'Emotional' subscale scores from the SCI. This is because higher scores on FACE-Q cancer worry indicate increased worry, but for the SCI a lower score indicates higher worry[16].
2) 'Cancer worry' subscale scores from the FACE-Q Skin Cancer module would negatively correlate with 'Social' subscale scores from the SCI (which measures social function).
3) 'Appearance of scars' subscale scores from the FACE-Q Skin Cancer module would positively correlate with 'Total appearance' subscale scores from the SCI (which has two items relating to scars and one relating to overall 'attractiveness').
4) 'Satisfaction with facial appearance' subscale scores from the FACE-Q Skin Cancer module would positively correlate with 'Social' subscale scores from the SCI.
5) Median pre-operative 'cancer worry' scores from the FACE-Q Skin Cancer module would be higher that median post-operative 'cancer worry' scores.

FACE-Q Skin Cancer module scores were computed in *R*, using the developer's score conversion charts, following listwise exclusion of incompete response sets. SCI scores were calculated according to the developer's guidelines[25]. Interpretation of Pearson's r values was based on guidelines by Cohen; small (|r| = 0.10 to 0.29), medium (|r| = 0.30 to 0.49) and large (|r| = 0.50 to 1.0)[26]. Shapiro-Wilk tests confirmed non-normality of pre-operative ($p = 0.003$) and post-operative ($p < 0.001$) 'cancer worry' score distributions, therefore median pre-operative and post-operative scores were compared using a two-tailed, paired Wilcoxon ranked-sum test.

Results

Demographic data

Overall, 244 patients were enrolled with five excluded due to non-completion of the pre-operative questionnaire, leaving an analysable dataset of 239 patients. There was a 67.6% response rate for the post-operative questionnaire. Demographic details are reported in *Table 2*.

Modern test theory analysis

Results for each sub-scale are presented in *Table 3*. All sub-scales demonstrated unidimensionality and scalability with Loevinger's coefficients ≥ 0.3. Each of the 3 PROM subscales, 'satisfaction with facial appearance', 'appearance of scars' and 'cancer worry', demonstrated LD. Thresholds were generally successively ordered with appropriate coverage, however the two most positive response thresholds for item 10 in the 'cancer worry' scale were disordered, and the two most positive response thresholds for item 8 of the 'cancer worry' scale

were closely situated (*Figure 1*). The IPP for 'appearance of scars' demonstrated a ceiling effect with poor coverage towards the positive end of the scale.

Of the patient reported experience measure (PREM) subscales (*Table 1*), 'satisfaction with appearance information' functioned well with appropriate response thresholds and coverage. All other PREM subscales performed poorly, with disordered thresholds, large ceiling effects, poor coverage towards the positive end of the scale and poor or uncalculatable fit statistics. A high degree of local dependency was also noted.

The 'sun protection' checklist demonstrated successively ordered thresholds and good item coverage. Fit statistics were good with 5 pairs showing local dependency. The 'adverse effects' checklist performed less well, with item 10 demonstrating disordered thresholds and a floor effect noted. Item 2 and 10 had poor fit. 13 pairs demonstrated local dependency.

All subscales had Cronbach's alpha values ≥ 0.7, although many demonstrated very high values above 0.95.

Item characteristic curves and IPPs for 'Facial appearance', 'Appraisal of scars', 'Cancer worry' (before and after scoring modification), 'Satisfaction with appearance information' and 'Sun protection' scales are shown in *Supplementary Figures 2 and 3*. Infit and outfit statistics are presented in *Supplementary Figure 4*.

Hypothesis testing and responsiveness

Table 4 summarises the results of hypothesis testing. *A priori* hypotheses were confirmed, with 'cancer worry' correlating negatively with 'emotional' and 'social' subscales of the SCI ($r = -0.68$, $p < 0.001$, and $r = -0.53$, $p < 0.001$, respectively). A strong positive

correlation between ‘appearance of scars’ and the SCI ‘total appearance’ subscales was demonstrated ($r = 0.59$, $p < 0.001$), in addition to a positive correlation between ‘satisfaction with facial appearance’ and the SCI ‘social’ subscale ($r = 0.47$, $p < 0.001$). A significant decrease in median ‘cancer worry’ scores was observed between pre-operative and post-operative questionnaires (44 to 39, $p < 0.001$) (*Table 4*).

Discussion

The FACE-Q Skin Cancer module is a promising PROM for use in patients with facial skin cancer. In this second report from the PROMISCR study, we assessed the anglicised FACE-Q Skin Cancer module to see if it fit the Rasch model and could therefore provide interval level measurement data in a UK population. Overall, our analysis suggests acceptable Rasch model fit for a number of subscales.

A degree of local dependency was observed in all subscales, together with high Cronbach’s alpha scores, consistent with item redundancy[22]. Very high ceiling effects were observed in a number of subscales, mainly those assessing the patients’ experience (e.g. ‘satisfaction with doctor/surgeon’). There are a number of explanations for this, such as mismatch between the position of items and patients on a scale, or acquiescence bias, in which there is a tendency to respond positively to all questions[27]. This can be especially true for subscales such as ‘satisfaction with doctor/surgeon’, where patients may not want to cause offence by answering negatively. It could also be that the patient population sampled is in fact very happy with the care delivered to them.

In this sample, items 8 and 10 in the ‘cancer worry’ scale had closely situated and disordered response thresholds, respectively. This can occur, for example, when respondents

have difficulty distinguishing between two responses (e.g. “agree” and “strongly agree”) and therefore the successive response options do not reflect an increase in the underlying trait. In PROM development, close and disordered thresholds are sometimes handled by collapsing response options (e.g. combining “agree” and “strongly agree” into one “agree” option) and then repeating Rasch analysis in an independent sample to confirm improved performance[14]. Current licensing restrictions do not permit modifications to the FACE-Q or its scoring system. Future work could assess whether this observation is replicated in other samples, or if these items exhibit differential item functioning by country in a mixed UK-US sample.

Convergent and divergent validity of the instrument were confirmed through affirmation of *a priori* hypotheses. The ‘cancer worry’ subscale was shown to be responsive to change, however no other subscales were tested for responsiveness as it was felt that an expected change in post-operative scores could not be guaranteed. For example, some people may believe that their facial appearance has improved as a result of surgical removal of their skin cancer, while others may feel that reconstruction of the resulting defect with a skin graft is worse.

It is worth noting that the final subscales published for the original US FACE-Q Skin Cancer module do not include some of those assessed here. Three of the four PREM subscales were derived from the BODY-Q questionnaire and tested in this population due to good face validity in early qualitative interviewing. Unfortunately these subscale did not fit the Rasch model in this population sample and are therefore unable to provide interval level measurement. However, in our view a pragmatic approach to scale development and use is needed. The psychometric properties of even the best functioning subscales in the UK FACE-Q Skin Cancer module could potentially be improved, however a balance between content and construct validity must be struck. The removal of items may improve the psychometric results

obtained, but risks missing concepts that are important to patients. While it is vital that psychometrically valid PROMs are used in clinical practice they must also address the patient need, even if this results in some trade off in terms of construct validity. Subscales such as 'satisfaction with doctor/surgeon' and 'satisfaction with medical/ward team' could be deemed to be too psychometrically flawed to be useful (and removed such as in the original US version). However, responses to these items can provide useful qualitative information to clinicians regarding service delivery and could be delivered as checklists alongside other scales with these limitations known and understood by those using them.

It is also important to view the psychometric results and limitations in the context of other PROMs for facial skin cancer. While many PROMs specific to this condition have been designed, all have imperfections[7]. The FACE-Q Skin Cancer module is well designed, using principles of modern test theory and demonstrates good face validity.

Limitations of this study included the representivity of the sample to the UK population and the sample size available for analysis. Recruitment from two disctinct centres in different areas of the UK resulted in patients from a range of backgrounds, however it is likely that the full spectrum of the UK population is yet to be sampled, especially those patients from ethnic minority groups. This will be important to address in those groups that are either more prone to poorer scarring or who have differing cultural and religious beliefs around facial scarring. The sample size in this study is able to provide item calibrations for the 'cancer worry' scale to within < 0.5 logits with at least 99% confidence[17]. However, sufficient sample sizes were not reached for other subscales to confidently calibrate Rasch model parameters.

Ongoing routine data collection using the anglicised FACE-Q Skin Cancer module should be performed in order to increase the sample size for further Rasch analyses. Further work in a larger cohort should focus on sub-group analysis and clinically important differences.

Conclusion

This is largest UK study of the FACE-Q Skin Cancer module, and supports the instrument's construct validity.

Conflict of Interest: We have no conflicts of interest to declare. The FACE-Q Skin Cancer module is owned by Memorial Sloan Kettering Cancer Centre (New York City, USA) and is used under license.

Funding: The work was funded by the Abertawe Bro Morgannwg University Health Board Pathway to Portfolio scheme.

References

1. Geller, A. C. & Annas, G. D. Epidemiology of melanoma and nonmelanoma skin cancer. *Seminars in Oncology Nursing* **19,** 2–11 (2003).
2. Franceschi, S., Levi, F., Randimbison, L. & La Vecchia, C. Site distribution of different types of skin cancer: New aetiological clues. *International Journal of Cancer* **67,** 24–28 (1996).
3. Körner, A., Garland, R., Czajkowska, Z., Coroiu, A. & Khanna, M. Supportive care needs and distress in patients with non-melanoma skin cancer: Nothing to worry about? *Eur J Oncol Nurs* **20,** 150–155 (2016).
4. Sobanko, J. F., Sarwer, D. B., Zvargulis, Z. & Miller, C. J. Importance of Physical Appearance in Patients With Skin Cancer. *Dermatol Surg* **41,** 183–188 (2015).
5. McGrail, K., Bryan, S. & Davis, J. Let's all go to the PROM: the case for routine patient-reported outcome measurement in Canadian healthcare. *Healthc Pap* **11,** 8–18 (2011).
6. Devlin, N. J. & Appleby, J. Getting the most out of PROMS. Putting health outcomes at the heart of NHS decision-making. *The Kings Fund* (2010).
7. Dobbs, T. D., Samarendra, H., Hughes, S., Hutchings, H. A. & Whitaker, I. Patient-reported outcome measures for facial skin cancer: a systematic review and evaluation of the quality of their measurement properties. *Br. J. Dermatol.* **180,** 1018–1029 (2019).
8. Lee, E. H., Klassen, A. F., Cano, S. J., Nehal, K. S. & Pusic, A. L. FACE-Q Skin Cancer Module for measuring patient-reported outcomes following facial skin cancer surgery. *BJD* **179,** 88–94 (2018).
9. Dobbs, T. D., Ottenhof, M., Gibson, J. A., Whitaker, I. S. & Hutchings, H. A. The Patient Reported Outcome Measures In Skin Cancer Reconstruction (PROMISCR) study – anglicisation and initial validation of the FACE-Q Skin Cancer module in a UK cohort. *Journal of Plastic, Reconstructive & Aesthetic Surgery* (2020). doi:10.1016/j.bjps.2020.09.005
10. Administration, F. A. D. *Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims*. (Fed Regist, 2009).
11. European Medicines Agency (EMA), Committee for Medicinal Products for Human Use (CHMP). Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. (2005). Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637. (Accessed: 25 May 2018)
12. Klassen, A. F. *et al.* The BODY-Q: A Patient-Reported Outcome Instrument for Weight Loss and Body Contouring Treatments. *Plast Reconstr Surg Glob Open* **4,** e679 (2016).
13. Mokkink, L. B. *et al.* The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* **19,** 539–549 (2010).
14. Andrich, D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res* **11,** 571–585 (2011).
15. Dobbs, T., Hutchings, H. A. & Whitaker, I. S. UK-based prospective cohort study to anglicise and validate the FACE-Q Skin Cancer Module in patients with facial skin cancer undergoing surgical reconstruction: the PROMISCR (Patient-Reported

Outcome Measure in Skin Cancer Reconstruction) study. *BMJ Open* **7,** e016182 (2017).
16. Rhee, J. S., Matthews, B. A., Neuburg, M., Buyzynski, M. & Nattinger, A. B. Creation of a quality of life instrument for nonmelanoma skin cancer patients. *Laryngoscope* **115,** 1178–1185 (2005).
17. Linacre, J. M., Heinemann, A. W., Wright, B. D., Granger, C. V. & Hamilton, B. B. The structure and stability of the Functional Independence Measure. *Arch Phys Med Rehabil* **75,** 127–132 (1994).
18. Stochl, J., Jones, P. B. & Croudace, T. J. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. *BMC Med Res Methodol* **12,** 1–16 (2012).
19. Chalmers, R. mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software* **48,** (2012).
20. RUMM Laboratory Pty Ltd. Test-of-fit-statistics. Available at: http://www.rummlab.com.au/faq7.html. (Accessed: 10 August 2020)
21. Wright, B. D. Reasonable mean-square fit values. *Rasch Measurement Transactions* **2,** 370 (1996).
22. Christensen, K. B., Makransky, G. & Horton, M. Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations:. *Applied Psychological Measurement* **41,** 178–194 (2016).
23. Paek, I. & Cole, K. *Using R for item response theory model applications*. (2019).
24. Mokkink, L. B. *et al.* COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res* **27,** 1171–1179 (2018).
25. Rhee, J. S. *et al.* Validation of a Quality-of-Life Instrument for Patients With Nonmelanoma Skin Cancer. *Arch Facial Plast Surg* **8,** 314–318 (2006).
26. Cohen, J. *Statistical power analysis for the behavioral sciences*. (Lawrence Earlbaum Associates, 1988).
27. Kam, C. C. S. & Zhou, M. Does Acquiescence Affect Individual Items Consistently? *Educational and Psychological Measurement* **75,** 764–784 (2015).

Table 1 – An overview of the anglicised FACE-Q Skin Cancer module.

Subscale	Number of items	Example question	Type of outcome measure
Satisfaction with facial appearance	9	With your entire face in mind, in the past week, how satisfied or dissatisfied have you been with the shape of your face?	PROM
Appearance of scars	8	With your scars in mind, in the past week, how much have you been bothered by the length of your scar?	PROM
Cancer worry	10	With your skin cancer diagnosis in mind, in the past week, how much do you agree or disagree with each statement: I worry about my skin cancer	PROM
Satisfaction with appearance information	6	How satisfied or dissatisfied you were with the information you received in relation to the following: how your appearance would change?	PREM
Satisfaction with doctor/surgeon	10	These questions ask about the doctor or surgeon who did your most recent procedure. Did you feel that he/she: answered all your questions?	PREM
Satisfaction with clerical staff	10	These questions ask about members of the clerical staff (e.g. secretaries, receptionists) who helped you during your most recent procedure. Did you feel that they: were attentive to your needs?	PREM
Satisfaction with medical/ward team	10	These questions ask about members of the medical and ward team other than your surgeon (e.g. nurses, other doctors) who looked after you during your most recent procedure. Did you feel that they: were friendly and kind?	PREM
Satisfaction with information	10	How satisfied or dissatisfied were you with the information you received in relation to the following: how the surgery would be done?	PREM
Sun protection behaviour	5	When you spend time outdoors, how often have you: worn suncream when you were outside?	Behavioural assessment tool
Symptoms checklist	10	With the part of your face affected by the skin cancer in mind, in the past week, how much have you been bothered by: discomfort?	Checklist

The FACE-Q Skin Cancer module can be considered a hybrid measurement tool, combining sub-scales addressing PROMs (Patient Reported Outcome Measure), PREMs (Patient Reported Experience Measure), behavioural assessment and a checklist.

Table 2 – Patient demographics and characteristics of those enrolled in the PROMISCR study.

Variable	All patients (n=239)
Age	
Mean age (SD)	71.4 (12.5)
Gender	
Male	119 (49.8%)
Female	120 (50.2%)
Co-morbidities	
Cardiovascular	62 (25.9%)
Respiratory	5 (2.1%)
Cancer (other than skin cancer)	12 (5.0%
Mental health	3 (1.3%)
Musculoskeletal	5 (2.1%)
Other	25 (10.5%)
None	127 (53.1%)
Medication	
Warfarin	18 (7.5%)
Aspirin	30 (12.6%)
Clopidogrel	4 (1.7%)
Other anticoagulation	5 (2.1%)
Immunosuppression	4 (1.7%)
Other	45 (18.8%)
None	133 (55.6%)
Histology	
BCC	180 (75.3%)
SCC	25 (10.5%)

Melanoma	9 (3.8%)
Lentigo maligna	6 (2.5%)
Actinic keratosis	6 (2.5%)
Other	5 (2.1%)
Location	
Forehead	36 (15.1%)
Eyelid	40 (16.7%)
Nose	98 (41.0%)
Lips	8 (3.3.%)
Medial cheek	35 (14.6%)
Lateral face	10 (4.2%)
Ear	8 (3.3%)
Chin	1 (0.4%)
Reconstruction	
Direct closure	95 (39.7%)
Skin graft	75 (31.4%)
Local flap	46 (19.2%)
Secondary intention	19 (7.9%)
Previous facial surgery	
Yes	99 (41.4%)
No	140 (58.6%)
Previous skin cancer	
Yes	108 (45.2%)
No	131 (54.8%)

Missing data – Histology (5 patients), lesion location (3 patients), reconstruction type (3 patients). BCC – basal cell carcinoma. SCC – squamous cell carcinoma.

Table 3 – Results of Rasch analysis for each individual sub-scale for the anglicised FACE-Q Skin Cancer module.

Subscale	Analyzable sample size	Scalability	Unidimensionality	Response thresholds	Targeting	Fit	Local dependency (LD)	Cronbachs alpha
Satisfaction with facial appearance	158	Scale H = 0.82	Yes	Ordered	Small ceiling effect, good coverage	Good fit	LD of 22 pairs	0.959
Appearance of scars	89	Scale H = 0.86	Yes	Ordered	Large ceiling effect	Good fit	LD of 8 pairs	0.969
Cancer worry	214	Scale H = 0.69	Yes	Item 10 disordered Item 8 close thresholds	Excellent coverage	Item 8 χ^2 = 38.43 (p = 0.019) Item 7 χ^2 = 35.19 (p = 0.009) Item 10 χ^2 = 33.48 (p = 0.01)	LD of 20 pairs	0.942
Satisfaction with appearance information	142	Scale H = 0.89	Yes	Ordered	Good coverage with mild ceiling effect	Good fit	LD of 5 pairs	0.960
Satisfaction with doctor/surgeon	154	Scale H = 0.69	Yes	Disordered	Massive ceiling effect	χ^2 uncalculatable Infit/outfit acceptable	LD of 20 pairs	0.925

Satisfaction with clerical staff	176	Scale H = 0.79	Yes	Items 5, 9 and 10 disordered	Massive ceiling effect	Item 3 χ^2 = 10.09 (p = 0.006) Item 7 χ^2 = 14.61 (p = 0.001)	LD of 21 pairs	0.962
Satisfaction with medical/ward team	162	Scale H = 0.83	Yes	Item 5 disordered	Massive ceiling effect	χ^2 uncalculatable Infit/outfit acceptable	LD of 17 pairs	0.953
Satisfaction with information	122	Scale H = 0.70	Yes	Item 3 disordered	Large ceiling effect	Item 4 χ^2 = 15.29 (p = 0.000)	LD of 17 pairs	0.929
Sun protection behaviour	193	Scale H = 0.55	Yes	Ordered	Excellent coverage	Good fit	LD of 5 pairs	0.835
Symptoms checklist	126	Scale H = 0.61	Yes	Item 10 disordered	Floor effect	Item 2 χ^2 = 17.58 (p = 0.004) Item 10 χ^2 = 15.91 (p = 0.007)	LD of 13 pairs	0.917

Table 4 – Hypothesis testing of the FACE-Q Skin Cancer module

Subscales	**Pearson's r**	**p value**	**Explanation**
'cancer worry' AND SCI subscale 1 'emotional'	-0.68	< 0.001	A large negative correlation – as predicted due to the scoring of items in each scale (i.e. as FACE-Q cancer worry increases (higher score) SCI cancer worry also increases (but higher worry is represented by a lower score)
'cancer worry' AND SCI subscale 2 'social'	-0.53	< 0.001	A large negative correlation – as predicted those people that are more worried by their skin cancer on the FACE-Q cancer subscale have more social worry on the SCI
'appearance of scars' AND SCI subscale 3 'total appearance'	0.59	< 0.001	A large positive correlation – as predicted better scores on FACE-Q 'appearance of scars' subscale indicate greater satisfaction with scars, which mirror increasing scores on SCI 'appearance' subscale
'satisfaction with facial appearance' AND SCI subscale 2 'social'	0.47	< 0.001	A medium positive correlation – as predicted with increasing happiness with facial appearance on FACE-Q correlating with increasing happiness with appearance on the SCI
Subscales	**Pre-operative median score** (n = 214)	**Post-operative median score** (n = 156)	**p value**	**Explanation**
Cancer worry	44	39	< 0.001	Cancer worry significantly decreased post-operatively

SCI – Skin Cancer Index

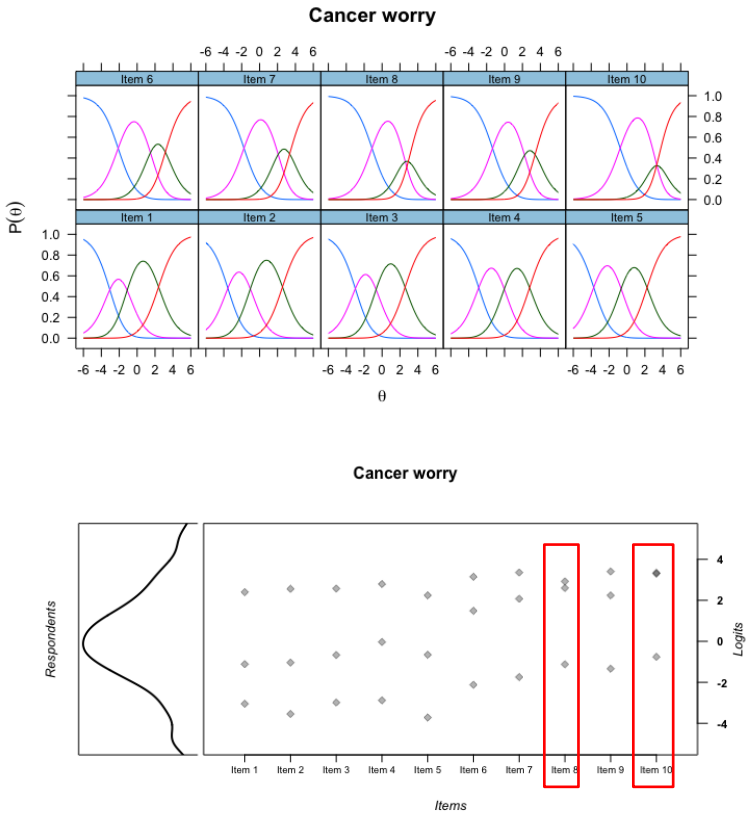
Figure Legends:

1 –

Upper plots – Item Characteristic Curves, Lower plots – Item Person Plots

Item Person Plots (IPP): y-axis demonstrates a Rasch scale between +4 and -4 Logits. The left panel of the IPP represents the score distribution and the right panel represents the item distributions. Points in the right panel illustrate item response thresholds.

Figure 1 – Item Characteristic Curves and Item Person Plots for the 'cancer worry' subscale showing results for pre- and post- response option collapsing



Upper plots – Item Characteristic Curves, Lower plots – Item Person Plots

Item Person Plots (IPP): y-axis demonstrates a Rasch scale between +4 and -4 Logits. The left panel of the IPP represents the score distribution and the right panel represents the item distributions. Points in the right panel illustrate item response thresholds.

Supplementary Figure 1 – Inclusion and exclusion criteria for the Patient Reported Outcome Measures In Skin Cancer Reconstruction (PROMISCR) study.

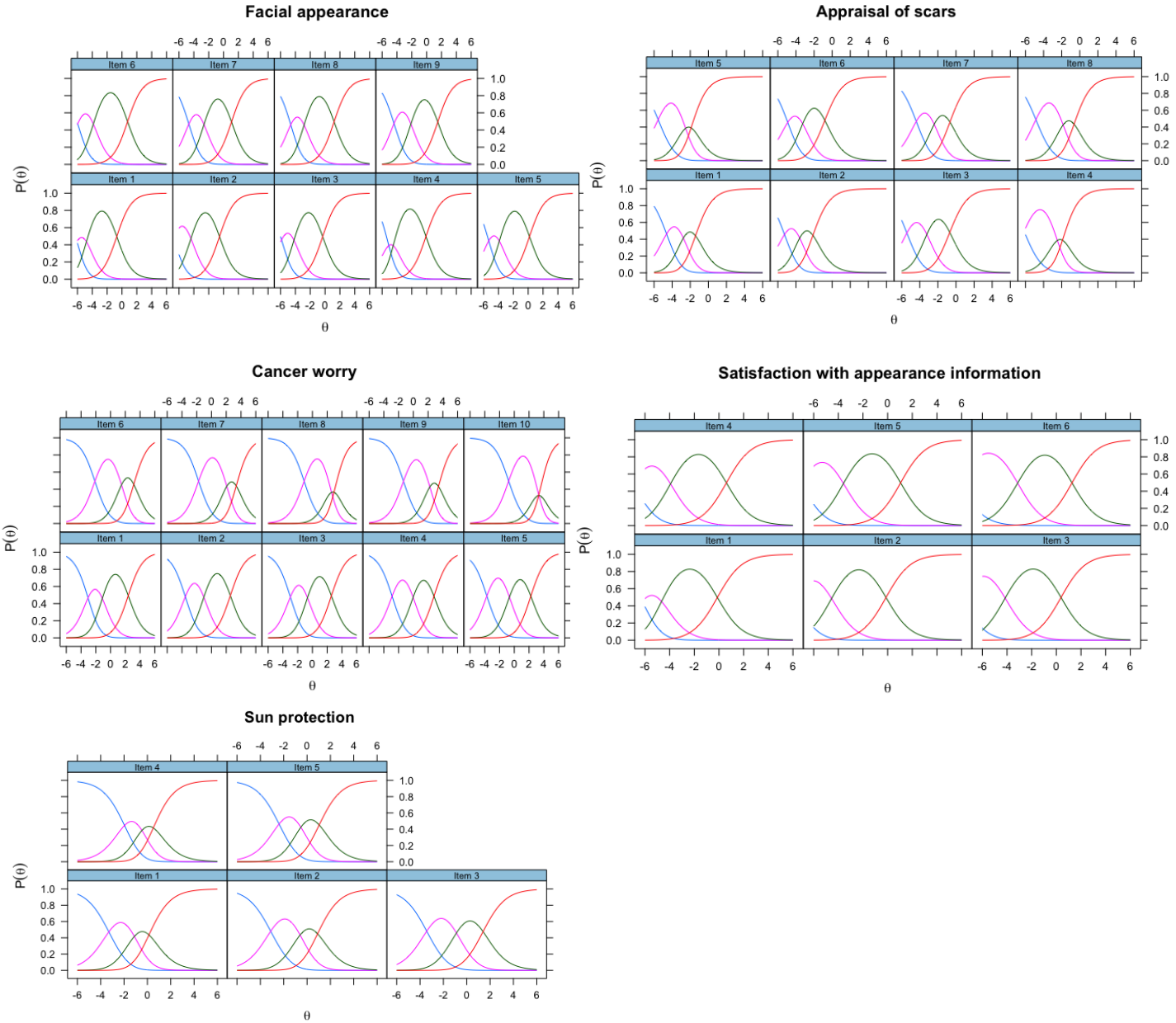
Inclusion criteria

- Skin cancer (all types included) of the face
- Over 18 years of age
- Active treatment with wide local excision of the lesion

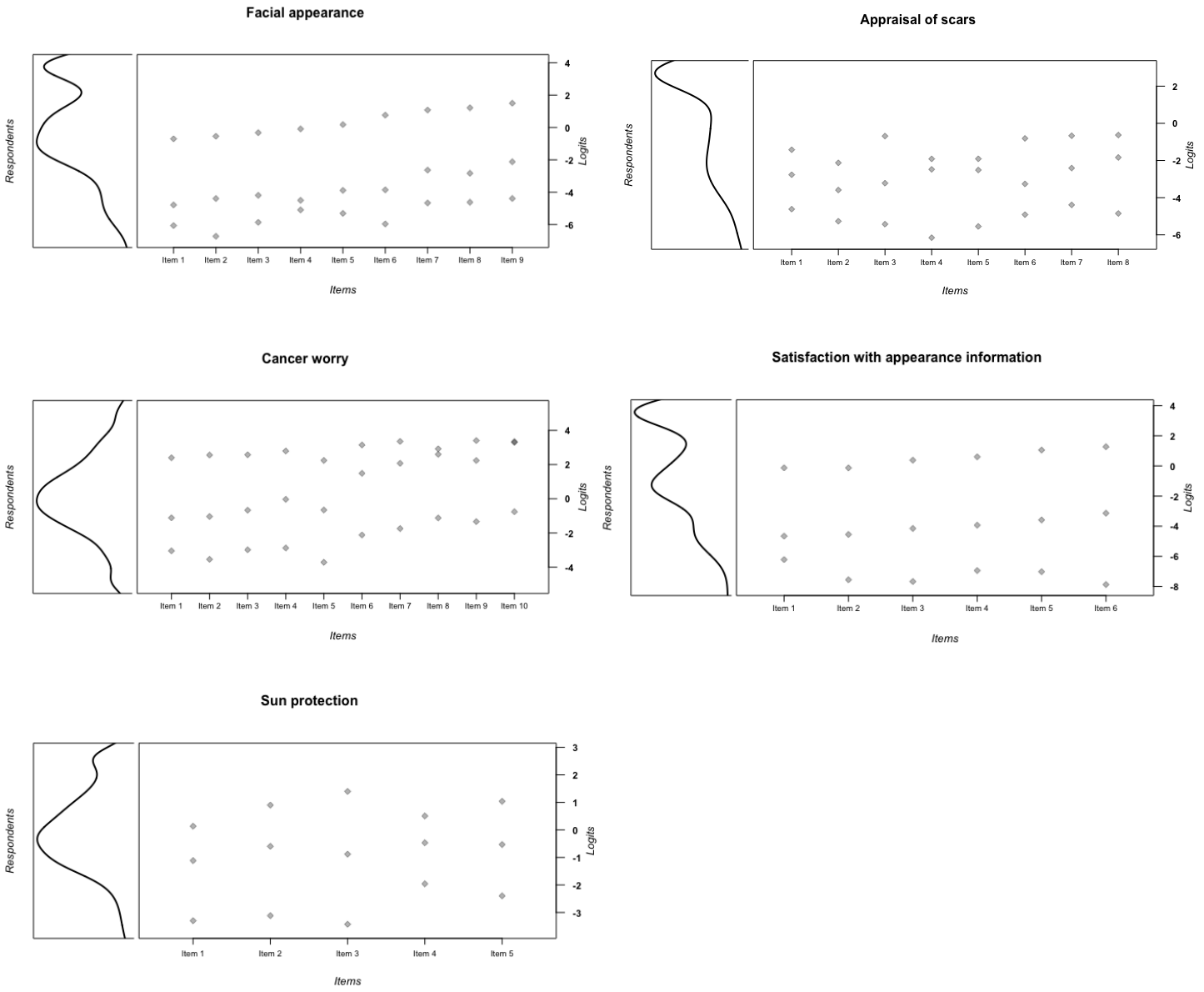
Exclusion criteria

- Inability to consent to participation in the study
- Known learning difficulties or dementia
- English language not of a standard to understand and complete the questionnaire
- Treatment of lesion with topical chemotherapy/laser or other methods that are not excisional
- Free tissue reconstruction

Supplementary Figure 2 – Item Characteristic Curves for FACE-Q Skin Cancer module subscales.



Supplementary Figure 3 – Item Person Plots for the FACE-Q Skin Cancer module subscales.



Supplementary Figure 4 – Infit and Outfit statistics for the FACE-Q Skin Cancer module subscales.

Subscale	Item	Infit	Outfit
Satisfaction with facial appearance	1	0.72	0.575
	2	0.641	0.462
	3	0.579	0.413
	4	0.695	0.568
	5	0.733	0.574
	6	1.104	0.863
	7	0.897	0.716
	8	0.509	0.419
	9	0.653	0.561
Appearance of scars	1	0.857	0.401
	2	0.708	0.316
	3	0.569	0.366
	4	0.722	0.427
	5	0.404	0.26
	6	0.602	0.389
	7	0.533	0.311
	8	0.973	0.605
Cancer worry	1	0.816	0.783
	2	0.721	0.666
	3	0.697	0.66
	4	0.877	0.856
	5	0.941	0.937
	6	1.073	1.04
	7	0.786	0.714
	8	0.871	1.025
	9	0.844	0.716
	10	1.185	0.979

Satisfaction with appearance information	1	0.776	0.582
	2	0.505	0.327
	3	0.48	0.312
	4	0.653	0.437
	5	0.377	0.265
	6	0.62	0.497
Satisfaction with doctor/surgeon	1	0.557	0.043
	2	0.857	0.226
	3	1.236	0.287
	4	0.52	0.032
	5	0.784	0.227
	6	0.894	0.265
	7	0.74	0.166
	8	0.923	0.217
	9	1.193	0.461
	10	1.185	0.412
Satisfaction with clerical staff	1	0.817	0.18
	2	0.668	0.155
	3	0.89	0.198
	4	0.592	0.119
	5	0.507	0.098
	6	0.478	0.099
	7	1.462	0.568
	8	0.537	0.103
	9	0.591	0.139
	10	0.741	0.241
Satisfaction with medical/ward team	1	1.633	0.417
	2	0.463	0.039
	3	0.726	0.068
	4	0.749	0.175
	5	0.525	0.049
	6	0.521	0.043
	7	0.512	0.05

	8	0.656	0.104
	9	0.656	0.104
	10	1.129	0.286
Satisfaction with information	1	0.84	0.566
	2	1.303	1.04
	3	1.307	0.814
	4	0.722	0.173
	5	0.614	0.273
	6	0.876	0.338
	7	0.7	0.534
	8	0.858	0.574
	9	0.735	0.432
	10	0.697	0.42
Sun protection behaviour	1	0.671	0.604
	2	1.158	1.084
	3	0.715	0.695
	4	0.974	0.885
	5	0.553	0.529
Symptoms checklist	1	0.868	0.613
	2	0.595	0.443
	3	0.693	0.514
	4	1.296	1.293
	5	0.87	0.702
	6	0.831	0.719
	7	0.959	0.825
	8	1.134	1.06
	9	1.228	0.929
	10	0.858	0.532