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

Mr Thomas D Dobbs MA, BM BCh, MRCS

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

Dr Maarten Ottenhof

3 Department of Plastic, Reconstructive and Hand Surgery, Catharina Hospital, Eindhoven, Netherlands

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

Mr John A G Gibson MBBcH, MRCS

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

Professor Iain 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

Professor Hayley A Hutchings BSc, PhD

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

Corresponding Author

Mr Thomas Dobbs MA, BM BCh, MRCS

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

E: tomdobbs@doctors.org.uk

T: 07973986658

Meetings

This work has been presented at the Summer BAPRAS meeting 2019

Acknowledgements

We would like to thank the developers of the questionnaire, specifically Drs Andrea Pusic and Anne Klassen, for the permission to use an early development version of the FACE-Q Skin Cancer module.

Summary

Facial skin cancer is common, and its treatment affects patient's health-related quality of life (HRQoL), as demonstrated by patient reported outcome measures (PROMs). In this study we Anglicise and validate the novel FACE-Q Skin Cancer module for the UK population.

Anglicisation of the FACE-Q Skin Cancer module followed international guidance for cross-cultural adaptation. Cognitive interviews were performed, producing a reconciled and harmonised version for validation. Patients undergoing facial skin cancer excision were prospectively recruited and asked to complete the anglicised FACE-Q Skin Cancer module, along with the Skin Cancer Index (SCI) and European Quality of Life Five Dimensions (ED-5D) questionnaire, pre-operatively and 6-8 weeks post-operatively. Data were analysed using classical test theory. Ethical approval was received (REC: 16/WM/0445).

One hundred and ten patients were recruited between August 2017 and July 2018. Internal consistency was high (Cronbach's alpha 0.867-0.967). All subscales had a single factor solution using principal component analysis. Construct validity, as measured between the FACE-Q subscales and SCI subscales was good, with >75% of *a priori* predictions confirmed. Pearson's r for item-total correlation was >0.80 for several items and significant ceiling effects were shown in 7 of the 10 subscales, suggesting some item redundancy.

The UK version of this well-designed PROM demonstrates good face and construct validity. There is however a degree of redundancy within the scales and further work using Rasch analysis on a larger sample will help address this.

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 sites such as the face[2]. While mortality is generally low, especially for non-melanoma skin cancer (NMSC)[3,4], there is often a considerable psychological burden associated with anxiety relating to a cancer diagnosis[5] and concerns over visible scarring[6].

In order to improve global outcomes for patients with skin cancer it is important that a holistic approach to their health-related quality of life (HRQoL) is taken. This requires the assessment of HRQoL in these patients before, during and after treatment. One method for assessing HRQoL is the use of patient-reported outcome measures (PROMs). PROMs are standardised and validated questionnaires, completed by patients, that capture one or more aspects of their health and wellbeing[7,8]. They are considered by the United Kingdom (UK) Department of Health as the current best method for quantifying a patient's clinical experience, although their use clinically is still sporadic.

A recent systematic review demonstrated a paucity of appropriately designed and well validated PROMs for facial skin cancer, although evidence was found for a newly developed instrument that had considerable potential[9]. The FACE-Q Skin Cancer module has since been validated in an initial population of 209[10]. Due to the importance of robust PROM data in both clinical and research settings the United States (US) Food and Drug Administration (FDA) and the European Medicines Agency (EMA) expect a PROM to be appropriately translated and adapted before use[11,12]. This paper therefore presents the results of the UK anglicisation and initial validation of the FACE-Q Skin Cancer module.

Methods

The Patient Reported Outcomes In Skin Cancer Reconstruction (PROMISCR) study is a prospective anglicisation and validation study of the newly created FACE-Q Skin Cancer module. A study protocol was prospectively published[13] and research ethics committee approval granted (REC: 16/WM/0445). A number of international methodological guidelines for cross-cultural adaptation exist[14], with this study following those of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force[15] and the Patient Reported Outcome Consortium[16-18]. The study comprised of two phases, an anglicisation process followed by psychometric validation.

Anglicisation

The aim of cross-cultural adaptation is to provide equivalence between the source language (the language in which the PROM was originally developed) and the target language (the new language into which it is to be translated). The degree of cross-cultural adaptation required varies depending on the situation in which the adapted PROM is being used. Guillemin et al[19] described five different scenarios where differing adaptation needs are required (*Table 1*). These range from the situation in which no adaptation is required, to full translation and cross-cultural adaptation. Anglicisation for the FACE-Q Skin Cancer module comes under scenario C, where the instrument requires cultural adaptation only (*Table 1*). *Figure 1* demonstrates the steps performed in anglicisation. The FACE-Q Skin Cancer module used in this study was an early example provided by the original instrument developers, consisting of 88 items across 10 subscales. A copy of the original source language (US English) questionnaire was given to three plastic surgeons, one dermatologist and two health outcome measure

specialists for review, with ‘Americanism’ removed and wording changed where confusing. Cognitive interviews were carried out with five patients in line with the minimum recommended number for anglicisation[18] using the anglicised questionnaire and a basic interview plan. Further changes were made to the questionnaire following this, with a final harmonised version assessed for an appropriate level of readability before being taken forward for validation.

Psychometric validation

Newly diagnosed patients were recruited from a single centre plastic surgery unit, the Welsh Centre for Burns and Plastics, Wales. Study details were provided to eligible patients and time given 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 along with a copy of the Skin Cancer Index (SCI)[20] and the generic European Quality of Life Five Dimensions (EQ5D) questionnaire[21]. A summary page of questions were also included to gain insight into their views on the questionnaire content. A post-operative questionnaire pack was sent to each patient six to eight weeks following surgery, with a reminder letter sent after two weeks to those who had not returned the second questionnaire pack. Inclusion in the PROMISCR study did not have any bearing on the treatment received by those recruited and they were free to withdraw at any point.

There are no general criteria for the required sample size when validating a PROM questionnaire[22] although a sample size of between 50 and 100 has been suggested for a validation study using classical test theory (CTT)[22,23].

Eligibility criteria

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

Data collection and psychometric analysis

All questionnaires were anonymised 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). Basic demographic data were collected on each patient along with diagnosis, past medical history, medication use and reconstruction used. Missing data were dealt with by using the mean of the completed items on a scale to replace missing values if less than 50% of the scale's items were missing as per the developer's guidelines. While missing data were reported in raw terms for the analysis of 'missing data', all other areas of data analysis used a more complete data set with mean imputation having been performed. We followed published standards on the minimization and reporting of missing data where

appropriate[24]. Scores for each subscale were calculated by summing an individual's answers for that subscale and then converting this to a Rasch transformation score using tables provided by the developer. Questionnaire and clinical data were input into Statistical Package for Social Sciences (SPSS) software V.22 (IBM Analytics, NY, USA) for analysis. Significance was taken as $p < 0.05$ unless stated otherwise.

Due to the subscale nature of the FACE-Q questionnaire the majority of data analysis was done at the subscale level. Psychometric analysis followed guidance by the Scientific Advisory Committee of the Medical Outcomes Trust[25] and methods outlined by Streiner and Norman[22]. Psychometric validation is covered elsewhere[22,26,27], but briefly the following were performed.

1. Item piloting and underlying dimensions

Missing data values were calculated for each item to assess if respondents were preferentially leaving out specific items.

Floor and ceiling effect are a measure of how skewed the data are. If > 15% of respondents score the lowest or highest score the scale is said to have a floor or ceiling effect, respectively.

Internal consistency is a measure of the homogeneity of a scale and therefore can be used to interrogate the items for their worth within a scale[22]. Cronbach's alpha is used to measure internal consistency, with a value of 0.7 used as the minimum accepted value[22].

Item-total correlations were assessed using Pearson's correlation (r), with item-total correlation of less than 0.2 or greater than 0.8 considered for removal[22,28].

Principal component analysis (PCA) was applied to the pre-operative data to identify the underlying 'components' or 'factors' that make up individual subscales. PCA was

performed using the Direct Oblimin rotation technique to determine which items associate with one another into factors[29]. Kaiser's rule where only factors with an eigenvalue of ≥ 1.0 are retained was used[30]. Items were considered for removal if their loading onto a factor was < 0.4. Finally, Horn's parallel analysis was performed to confirm the number of factors present[31].

2. Validity

Construct validity is a measure of the correlation of the scale being tested to another instrument that is believed to assess the same or similar attributes. Pearson's correlation coefficient (r) was used to assess the correlation between subscales of the FACE-Q Skin Cancer module and subscales on the SCI, along with correlations between one FACE-Q subscale and another. A number of *a priori* hypotheses were made with construct validity assumed if greater than 75% of these *a priori* hypotheses proved to be correct[23]. Interpretation of Pearson's r values were based on guidelines by Cohen[32]; small (r = 0.10 to 0.29), medium (r = 0.30 to 0.49) and large (r = 0.50 to 1.0)

3. Responsiveness

Responsiveness in the instrument is its ability to detect change in a patient's condition when a change has occurred[33]. This was assessed by looking at group level change between pre-operative and post-operative questionnaires on subscales that were predicted to be influenced by the process of having surgery and interacting with the hospital environment.

Results

Anglicisation

The five patient participants had all been diagnosed and received treatment for a facial skin cancer within the last year, a sufficient length of time to have reflected on the process and no longer still be alarmed by the diagnosis, but not so long as to have forgotten the details of their treatment and how they felt. Words such as *'color'* and *'behavior'* were identified as US English spelling and corrected. Other words such a *'sunscreen'* were deemed by many to be an American term and UK residents would be more likely to use *'suncream'*. Similarly the word *'crooked'*, while used in UK English it was felt that in the context of the assessment of a scar, few UK English speakers would use that term. A number of terms used in US medical settings were also unfamiliar to UK patients, such as the term *'office staff'*, which was converted to *'clerical staff'* to encompass those members of the team such as the clinic receptionist and consultant secretaries. Face validity was also deemed to be good for the FACE-Q Skin Cancer module by all those that reviewed it.

All results were combined and a reconciled version of the anglicised FACE-Q Skin Cancer module was created (*Supplementary Figure 1*). Readability of this finalised version was good, with an approximate reading age of US grade five or UK school age 8-9 years old across a number of readability scores (*Supplementary Figure 2*).

Psychometric analysis

Demographic data

A total of 113 patients were recruited. Three patients withdrew from the study after consenting to inclusion, stating the length of the questionnaire as their reason,

resulting in a cohort of 110 patients completing the questionnaire pre-operatively (*Table 2*). Post-operative follow-up questionnaires were sent to all 110 patients. Seventy-three were returned, representing a 66% response rate. The mean length of time between operation (time point 1) and completion of a post-operative questionnaire (time point 2) was 8.6 weeks (SD = 3.8 weeks).

Missing Data

Missing data were calculated from the raw scores obtained from each questionnaire. *Table 3* summarises the range of missing data for each subscale of the FACE-Q Skin Cancer module and the SCI, with a number of the FACE-Q Skin Cancer subscales having greater than 20% missing data.

Floor and ceiling effect

Floor and ceiling effects for transformed scores were calculated for each subscale (*Table 4*). Significant ceiling effects above the recommended 15% maximum can be seen for a number of the subscales. Skewedness was calculated, showing that all subscales apart from subscale 10 (symptom checklist), were skewed towards the higher end of the spectrum. Normal values for skewedness are between -1 and 1, therefore five of the subscales are skewed outside of this normal range.

Internal consistency

Cronbach's alpha was calculated for each individual subscale and ranged between 0.867 and 0.967. Only four subscales (cancer worry, satisfaction with information, sun protection behaviour and the symptom checklist) had Cronbach's alphas of < 0.95.

Item-item and item-total correlation

A large number of items (41/88, 46.6%) had an item-total correlation of > 0.80. There were, however, no items that had a Pearson's r of < 0.20.

Principal Component Analysis

Table 5 demonstrates the Kaiser-Meyer-Olkin (KMO), Bartlett's test of sphericity, number of factors, eigenvalue and variance explained by this for each subscale. Single factor solutions were present for all subscales following Monte Carlo analysis. Strong loading is seen for all items in each scale, with values above 0.4, suggesting that none are candidates for removal.

Construct validity

Construct validity between FACE-Q Skin Cancer subscales and subscales of the SCI are summarised in *Table 6*. Correlation of individual FACE-Q subscales with each other also confirmed the *a priori* hypotheses. 'Satisfaction with facial appearance' showed a strong positive correlation with 'appearance of scars' ($r = 0.619$, $p < 0.001$). Higher 'cancer worry' correlated negatively with 'satisfaction with facial appearance' ($r = - 0.292$, $p = 0.005$), with this correlation present for both pre-operative and post-operative questionnaires. Interestingly the hypothesis that those who had a greater number of post-operative symptoms would score worse on 'appearance of scars' was also confirmed to be true ($p = - 0.448$, $p < 0.001$).

Responsiveness

Data were skewed with a significant Kolmogorov-Smirnov test, therefore Wilcoxon signed rank tests were used to assess the data. Median scores were assessed

between pre-operative and post-operative patients as all of these had undergone a change in their condition (i.e. surgery). 'Satisfaction with facial appearance' was non-significantly reduced between pre- and post-operative assessment. A significant decrease in 'cancer worry' was seen. 'Satisfaction with appearance information' increased in the post-operative cohort although this was non-significant (*Table 7*). 'Appearance of scars' could not be assessed as patients would not have had a pre-operative scar.

Discussion

There has been an identified need for a well-designed and validated PROM for those undergoing surgical treatment of a facial skin cancer. To develop a new PROM from the beginning is expensive, time consuming and potentially unnecessary[34]. If a PROM exists that can be adapted, either with the addition or removal of items and psychometric validation in the target population, this can have significant advantages. The PROMISCR study aimed to do this for the newly created FACE-Q Skin Cancer module.

The anglicisation process followed international consensus guidelines[15,18], with a small number of changes required to convert it to UK English spelling and remove language that was not understandable to a different cultural population.

A number of interesting results were found during psychometric validation. A significant amount of missing data was seen, with up to 47% of patients not completing some items. There are a number of reasons why this could be the case, such as those questions being too difficult for people to answer or the feeling that they are repetitive of others. Internal consistency supports the view that the scales are reliable and homogenous with Cronbach's alpha above 0.7 in all subscales. However, in some cases

Cronbach's alpha of ≥0.95 were seen, suggesting item redundancy. This was also the case for item-total correlation with many items having a Pearson's r of >0.8, again suggesting item redundancy[22,28]. In combination with the high levels of missing data there is considerable evidence that the number of items in the FACE-Q Skin Cancer module should be reduced.

A significant ceiling effect was seen in all subscales apart from 'cancer worry' and 'sun protection behaviour'. This means that a significant number of people are scoring the highest obtainable Rasch transformed score on these subscales, reducing responsiveness and interpretability of the scale. For example, if someone is to score the highest obtainable score and their condition changes, the instrument will only be able to detect this in one direction (i.e. a fall in scores). If the condition of these patients improves further however, it cannot be detected by the instrument. A floor or ceiling effect of greater than 15% is considered to be too high and may suggest that a scale is not functioning as intended[23]. One reason for the high scores seen could be that the patient population is generally very happy, however a range of EQ-5D-5L scores suggest that some people had lower levels of general HRQoL despite still scoring highly on the FACE-Q subscales. Acquiescence bias, in which there is a tendency to respond positively to all questions, may also be the cause of the high ceiling effects[35]. This is especially true for subscales that ask about feelings towards the staff treating the patient, where patients do not want to cause offence by answering negatively.

Principal component analysis (PCA) is a powerful analytical process for identifying factors within a group of items and those items that do not fit the model. All subscales were shown to be assessing a single underlying factor, with factor loading of greater than 0.4 for all items providing counter evidence to the assumption that the total items should be reduced.

The anglicised questionnaire demonstrated good construct validity (both convergent and divergent) with greater than 75% of the *a priori* hypotheses confirmed[23]. Responsiveness was identified in three subscales (satisfaction with facial appearance, cancer worry and satisfaction with appearance information). This suggests that these subscales are able to detect change in a patients' condition, with further research and greater numbers required to confirm these results and identify if other subscales are also responsive. These results are similar to those described in the developers' initial validation study, with significant floor and ceiling effects, good construct validity and responsiveness in the 'cancer worry' subscale also seen[10].

It is acknowledged that using a single centre plastic surgery cohort could introduce bias, however the demographics of this patient group were representative of those patients with facial skin cancer across the UK. The population studied was varied, but drawn from a South Wales centre with many people from rural and deprived backgrounds. It is possible that many of these patients were more content with their treatment and outcomes than a more highly educated and less deprived population in a larger city in the southeast of England would be. The merits of classical test theory (CTT) versus modern test theory (MTT) have been discussed at length in the literature[36,37]. CTT was chosen in this validation work for a number of reasons. In early validation of an instrument (such as when a new instrument is designed or translation occurs), CTT can be very useful in identifying items for removal and exploring the underlying dimensions of a scale. The importance of using CTT in PROM validation (in conjunction with MTT) is borne out in the continued presence of CTT in guidelines such as the COSMIN checklist[38,39] and those by Terwee et al[23] and Prinsen et al[40]. In order to address the limitations of this study a second phase of validation work is

underway, with a larger cohort of patients being recruited from a second site in England and planned Rasch analysis in line with the original instrument validation study.

Conclusion

The anglicised FACE-Q Skin Cancer module appears to be a well designed and valid PROM with good construct validity and responsiveness in some subscales. With further refinement and validation, the anglicised FACE-Q Skin Cancer module will play an important role in collecting and analysing patient reported data on facial skin cancer treatment outcomes in years to come.

Table 1 – Scenarios in which different degrees of cross-cultural adaptation are required. Adapted from Guillemin et al(19) and Beaton et al(41).

	Results in a change in			Adaptation required	
	Culture	Language	Country of use	Translation	Cultural adaptation
A) Use in same population. No change in culture, language or country	--	--	--	--	--
B) Use in established immigrants in source country	Yes	--	--	--	Yes
C) Use in another country, but same language	Yes	--	--	--	Yes
D) Use in new immigrants, not source language speaking but in the source country	Yes	Yes	--	Yes	Yes
E) Use in another country and another language	Yes	Yes	Yes	Yes	Yes

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

Variable	All patients (n=110)
Age	
Mean age (SD)	72 (12)
< 65 years of age	25 (22.7%)
> 65 years of age	85 (77.3%)
Gender	
Male	66 (60%)
Female	44 (40%)
Co-morbidities	
Cardiovascular	41 (37.3%)
Respiratory	2 (1.8%)
Cancer (other than skin cancer)	9 (8.2%)
Mental health	1 (0.9%)
Musculoskeletal	3 (2.7%)
Other	14 (12.7%)
None	40 (36.4%)
Medication	
Warfarin	11 (10%)
Aspirin	15 (13.6%)
Clopidogrel	4 (3.6%)
Other anticoagulation	4 (3.6%)
Immunosuppression	4 (3.6%)
Other	25 (22.7%)
None	47 (42.7%)
Histology	
BCC	61 (55.5%)
SCC	22 (20%)
Melanoma	4 (3.6%)
Lentigo maligna	5 (4.5%)

Other	5 (4.5%)
Actinic keratosis	6 (5.5%)
Location	
Forehead	29 (26.4%)
Eyelid	10 (9.1%)
Nose	36 (32.7%)
Lips	2 (1.8%)
Medial cheek	25 (22.7%)
Lateral face	4 (3.6%)
Ear	2 (1.8%)
Reconstruction	
Direct closure	41 (37.3%)
Skin graft	49 (44.5%)
Local flap	18 (16.4%)
Previous facial surgery	
Yes	47 (42.7%)
No	63 (57.3%)
Previous skin cancer	
Yes	52 (47.3%)
No	58 (52.7%)

Table 3 – Range of missing data for each subscale of the FACE-Q Skin Cancer module and the Skin Cancer Index (SCI).

Scale	Subscale	Range of missing data (%)
FACE-Q	Satisfaction with facial appearance	11.8 – 16.4
	Appearance of scars	41.8 – 47.3
	Cancer worry	3.6 – 7.4
	Satisfaction with information: appearance	25.5 – 30.9
	Satisfaction with doctor/surgeon	26.4 – 32.7
	Satisfaction with clerical staff	11.8 – 21.8
	Satisfaction with medical/ward team	26.4 – 30.9
	Satisfaction with information	25.5 – 40
	Sun protection behaviour	5.5 – 28.2
	Symptoms checklist	29.1 – 32.7
SCI	Emotional	6.4 – 10.9
	Social	9.1 – 10.9
	Appearance	7.3 – 10.9

Table 4 – Floor and ceiling effects calculated for each subscale in the FACE-Q skin cancer module.

	Pre-operative questionnaires						
Subscale	**Mean**	**SD**	**Range**	**Median**	**Skewness**	**Worst score – floor effect (% achieving this)**	**Best score – ceiling effect (% achieving this)**
Satisfaction with facial appearance	73.3	21.9	0 – 100	74	-0.58	0.9%	22.7%
Appearance of scars	80.9	23.2	0 – 100	91	-1.18	0.9%	24.5%
Cancer worry	49.4	21.4	0 – 100	50	-0.31	2.7%	1.8%
Satisfaction with information: appearance	79.5	21.6	0 – 100	80	-0.88	0.9%	30%
Satisfaction with doctor/surgeon	92.8	15.0	0 – 100	100	-3.63	0.9%	49.1%
Satisfaction with clerical staff	91.6	14.9	0 – 100	100	-1.63	0.9%	58.2%

Satisfaction with medical/ward team	95.7	11.0	44 – 100	100	-3.11	0%	57.3%
Satisfaction with information	82.5	18.7	40 – 100	90	-0.59	0%	29.1%
Sun protection behavior*	14.4	4.0	5 – 20	15	-0.31	0.9%	12.7%
Symptoms checklist*	15.6	6.8	10 – 40	13	1.49	19.1%	0.9%

* No transformed score available with original scale development therefore sum score of sub-scale used as per the developers' recommendations

Table 5 – Principal component analysis (PCA) for individual subscales of the FACE-Q Skin Cancer module.

Subscale	**KMO**	**Bartlett's test of sphericity**	**Number of factors**	**Eigenvalue**	**Variance explained by Eigenvalue (cumulative)**
Satisfaction with facial appearance	0.886	<0.001	1	6.591	73.2%
Appearance of scars	0.910	<0.001	1	6.267	78.3%
Cancer worry	0.906	<0.001	2 *1 following Monte Carlo	6.419 1.180	64.2% 76.0%
Satisfaction with appearance information	0.888	<0.001	1	5.169	86.2%
Satisfaction with doctor/surgeon	--	--	1	7.286	72.9%
Satisfaction with clerical staff	0.808	<0.001	1	7.269	72.7%
Satisfaction with medical/ward team	0.773	<0.001	1	7.316	73.2%

Satisfaction with information	0.875	<0.001	2 *1 following Monte Carlo	5.786 1.004	57.9% 67.9%
Sun protection behaviour	0.838	<0.001	1	3.350	67.0%
Symptoms checklist	0.887	<0.001	2 *1 following Monte Carlo	6.323 1.140	63.2% 74.6%

* Monte Carlo PCA for Parallel Analysis was used to confirm the number of factors after Oblimin rotation for all sub-scales as per the methods. Three sub-scales initially had two factors, although this was reduced to one following parallel analysis.

-- KMO/Bartlett's could not be calculated as the matrix showed linear dependency with an Eigenvalue of 0 for one item.

Table 6 – Summary of correlations between FACE-Q Skin Cancer module subscales and Skin Cancer Index subscales in order to assess construct validity.

Correlation	Pearson's r	Variance explained	p value	Explanation
'cancer worry' AND SCI subscale 1	- 0.756	57.2%	< 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	- 0.560	31.36%	< 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
'satisfaction with appearance information: appearance' AND SCI subscale 3	0.358	12.8%	0.001	A medium positive correlation – as predicted
'appearance of scars' AND SCI subscale 3	0.570	32.49	< 0.001	A large positive correlation – as predicted as better scores on FACE-Q appearance of scars indicate greater happiness with scars, along with increasing scores on SCI appearance subscale
'satisfaction with facial	0.439	19.3%	< 0.001	A medium positive correlation – as predicted with increasing

appearance' AND SCI subscale 2				happiness with facial appearance on FACE-Q correlating with increasing happiness with appearance on the SCI

Table 7 – Responsiveness of the FACE-Q Skin Cancer module to change in clinical condition.

FACE-Q subscale	**Pre-operative median score**	**Post-operative median score**	**Wilcoxon signed rank z value and level of significance (p)**	**Effect size**
Satisfaction with facial appearance	91.0	78.0	z = - 1.177 p = 0.239	0.104
Cancer worry	50	43	z = - 2.907 p = 0.004	0.220
Satisfaction with appearance information	80	92	z = - 0.299 p = 0.765	0.024
Satisfaction with doctor/surgeon	100	100	z = - 0.597 p = 0.550	0.048
Satisfaction with clerical staff	100	100	z = - 0.691 p = 0.489	0.054
Satisfaction with ward team	100	100	z = - 0.625 p = 0.532	0.051
Satisfaction with information	90	90	z = - 0.049 p = 0.961	0.004

Figure 1 – Anglicisation process applied to the FACE-Q Skin Cancer module following international guidelines.

Supplementary Figure 1 – Changes made to the FACE-Q Skin Cancer module during the anglicisation process.

Supplementary Figure 2 – Summary of readability scores for the anglicised FACE Q Skin Cancer module.

References

1. Geller AC, Annas GD. Epidemiology of melanoma and nonmelanoma skin cancer. Seminars in Oncology Nursing. W.B. Saunders; 2003 Feb 1;19(1):2–11.

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. Wiley Subscription Services, Inc., A Wiley Company; 1996 Jul 3;67(1):24–8.

3. Cancer Research UK statistics. 2016 Apr 7;:1–1. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer/incidence

4. Weinstock MA, Bogaars HA, Ashley M, Litle V, Bilodeau E, Kimmel S. Nonmelanoma skin cancer mortality. A population-based study. Arch Dermatol. 1991 Aug;127(8):1194–7.

5. 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. Elsevier; 2016 Feb 1;20:150–5.

6. Sobanko JF, Sarwer DB, Zvargulis Z, Miller CJ. Importance of Physical Appearance in Patients With Skin Cancer. Dermatol Surg. 2015 Feb 1;41(2):183–8.

7. 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. 2011;11(4):8–18.

8. Devlin NJ, Appleby J. Getting the most out of PROMS. Putting health outcomes at the heart of NHS decision-making. The Kings Fund. London; 2010.

9. Dobbs TD, Samarendra H, Hughes S, Hutchings HA, 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. John Wiley & Sons, Ltd (10.1111); 2019 May;180(5):1018–29.

10. Lee EH, Klassen AF, Cano SJ, Nehal KS, Pusic AL. FACE-Q Skin Cancer Module for measuring patient-reported outcomes following facial skin cancer surgery. BJD. 2018 Jul;179(1):88–94.

11. Administration FAD. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Fed Regist; 2009.

12. 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. [Internet]. 2005 [cited 2018 May 25]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637

13. Dobbs T, Hutchings HA, Whitaker IS. 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. British Medical Journal Publishing Group; 2017 Sep 1;7(9):e016182.

14. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. Journal of Clinical Epidemiology. Elsevier; 2015 Apr 1;68(4):435–41.

15. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. Value in Health. 2005 Mar;8(2):94–104.

16. Coons SJ, Kothari S, Monz BU, Burke LB. The Patient-Reported Outcome (PRO) Consortium: Filling Measurement Gaps for PRO End Points to Support Labeling Claims. Clin Pharmacol Ther. Wiley-Blackwell; 2011 Nov 1;90(5):743–8.

17. Hayes RP, Blum SI, Gordon MF, Piault E, Burke LB, Slagle AF, et al. The Patient-Reported Outcome (PRO) Consortium: Lessons Learned Along the Path to PRO Instrument Qualification. Therapeutic Innovation & Regulatory Science. SAGE PublicationsSage CA: Los Angeles, CA; 2014 Sep 14;49(1):132–8.

18. Eremenco S, Pease S, Mann S, Berry P, on behalf of PRO Consortium's Process Subcommittee. Patient-Reported Outcome (PRO) Consortium translation process: consensus development of updated best practices. J Patient Rep Outcomes. Nature Publishing Group; 2017;2(12):1–11.

19. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. Journal of Clinical Epidemiology. Elsevier; 1993 Dec 1;46(12):1417–32.

20. Rhee JS, Matthews BA, Neuburg M, Buyzynski M, Nattinger AB. Creation of a quality of life instrument for nonmelanoma skin cancer patients. Laryngoscope. Wiley-Blackwell; 2005 Jul 1;115(7):1178–85.

21. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy. 1990 Dec;16(3):199–208.

22. Streiner DL, Norman GR, Cairney J. Health Measurement Scales. A practical guide to their development and use. Fifth Edition. Oxford University Press; 2015.

23. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology. 2007 Jan;60(1):34–42.

24. Li T, Hutfless S, Scharfstein DO, Daniels MJ, Hogan JW, Little RJA, et al. Standards should be applied in the prevention and handling of missing data for patient-centered outcomes research: a systematic review and expert consensus. Journal of Clinical Epidemiology. 2014 Jan;67(1):15–32.

25. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002 May;11(3):193–205.

26. Dobbs T, Hughes S, Mowbray N, Hutchings HA, Whitaker IS. How to decide which patient-reported outcome measure to use? A practical guide for plastic surgeons. Journal of Plastic, Reconstructive & Aesthetic Surgery. 2018 Mar;71(7):957–66.

27. Wormald JCR, Rodrigues JN. Outcome measurement in plastic surgery. J Plast Reconstr Aesthet Surg. 2018 Mar;71(3):283–9.

28. Kline P. A Handbook of Test Construction. Introduction to Psychometric Design. Routledge; 2015.

29. Tabachnick BG, Fidell LS. Using multivariate statistics. 5 ed. Boston, MA: Allyn & Bacon/Pearson Education; 2007.

30. Kaiser HF. An index of factorial simplicity. Psychometrika. Springer-Verlag; 1974;39(1):31–6.

31. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika. Springer-Verlag; 1965 Jun;30(2):179–85.

32. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.

33. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: A clarification. Journal of Clinical Epidemiology. 1989 Jan;42(5):403–8.

34. Snyder CF, Watson ME, Jackson JD, Cella D, Halyard MY, the Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcome instrument selection: designing a measurement strategy. Value Health. 2007 Nov;10 (2):S76–S85.

35. Kam CCS, Zhou M. Does Acquiescence Affect Individual Items Consistently? Educational and Psychological Measurement. SAGE PublicationsSage CA: Los Angeles, CA; 2015 Oct 1;75(5):764–84.

36. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures_ A Comparison of Worked Examples. Value in Health. Elsevier; 2015 Jan 1;18(1):25–34.

37. Baylor C, Hula W, Donovan NJ, Doyle PJ, Kendall D, Yorkston K. An Introduction to Item Response Theory and Rasch Models for Speech-Language Pathologists. Am J Speech Lang Pathol. American Speech-Language-Hearing Association; 2011 Aug 1;20(3):243–59.

38. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, 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. Springer Netherlands; 2010 May;19(4):539–49.

39. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. Springer International Publishing; 2018 May 1;27(5):1171–9.

40. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. Springer International Publishing; 2018 Feb 12;27(5):1147–57.

41. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. Spine. 2000 Dec 15;25(24):3186–91.