

FACE-Q Eye Module for Measuring Patient-Reported Outcomes Following Cosmetic Eye Treatments

Anne F. Klassen, DPhil; Stefan J. Cano, PhD; James C. Grotting, MD, FACS; Stephen B. Baker, MD, DDS; Jean Carruthers, MD, FRCS(C), FRC(Ophth); Alastair Carruthers, MA, BM, BCh, FRCPC, FRCP(Lon); Nancy Van Laeken, MD; Jonathan M. Sykes, MD; Jonathan A. Schwitzer, MD; Andrea L. Pusic, MD, MSc

IMPORTANCE Aesthetic eye treatments can dramatically change a person's appearance, but outcomes are rarely measured from the patient perspective. The patient perspective could be measured using an eye-specific patient-reported outcome measure.

OBJECTIVE To describe the development and psychometric evaluation of FACE-Q scales and an adverse effect checklist designed to measure outcomes following cosmetic eye treatments.

DESIGN, SETTING, AND PARTICIPANTS Pretreatment and posttreatment patients 18 years and older who had undergone facial aesthetic procedures were recruited from plastic surgery clinics in United States and Canada and completed FACE-Q scales between June 6, 2010, and July 14, 2014. We used Rasch Measurement Theory, a modern psychometric approach, to refine the scales and to examine psychometric properties.

MAIN OUTCOMES AND MEASURES The FACE-Q Eye Module, which has 4 scales that measure appearance of the eyes, upper and lower eyelids, and eyelashes. Scale scores range from 0 (worst) to 100 (best). The module also includes a checklist measuring postblepharoplasty adverse effects.

RESULTS Overall, 233 patients (81% response rate) 18 years and older participated. Adverse effects included being bothered by eyelid scars, dry eyes, and eye irritation. In Rasch Measurement Theory analysis, each scale's items had ordered thresholds and good item fit. Person Separation Index and Cronbach α were greater than or equal to 0.83. Higher scores on the eye scales correlated with fewer adverse effects (range, -0.26 to -0.36). In the pretreatment group, older age correlated with lower scores (range, -0.42 to -0.51) on the scales measure appearance of the eyes and upper and lower eyelids. Compared with the pretreatment group, posttreatment participants reported significantly better scores on the scales measuring appearance of eyes overall, as well as upper and lower eyelids.

CONCLUSIONS AND RELEVANCE The FACE-Q Eye Module can be used in clinical practice, research and quality improvement to collect evidence-based outcomes data.

LEVEL OF EVIDENCE NA.

JAMA Facial Plast Surg. 2017;19(1):7-14. doi:10.1001/jamafacial.2016.1018
Published online September 15, 2016.

[← Invited Commentary page 14](#)

[+ Supplemental content at jamafacialplasticsurgery.com](#)

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Anne Klassen, DPhil, McMaster University, 3N27, 1200 Main St W, Hamilton, ON L8N 3Z5, Canada (aklass@mcmaster.ca).

Blepharoplasty performed on the upper and/or lower eyelids can improve appearance, and in some cases, eyelid function, by removing excess skin from the upper eyelids and bagginess from the lower eyelids. Modern concepts of periorbital rejuvenation also include volume replacement in the aging orbit using fillers or fat grafting. In the United States, 203 934 blepharoplasty procedures were performed in 2015.¹ Eyelid surgery was the third most common cosmetic procedure for women, and the second most common procedure for men.¹ In the United Kingdom in 2014, blepharoplasty was the second most common cosmetic operation next to breast augmentation for women, and the most common operation for men, with 7752 procedures performed in total.² Despite the popularity of blepharoplasty, outcomes of the procedure and other appearance enhancing eye treatments (eg, eyelash treatment), are rarely evaluated from the patient's perspective.³

An important limitation in the ability to measure outcomes from the patient perspective has been the lack of a psychometrically sound patient-reported outcome measure (PROM). PROMs are questionnaires that measure concepts of interest important to patients, such as appearance, health-related quality-of-life, and symptoms.⁴ In 2013, a literature review funded by the UK Department of Health to identify PROMs for cosmetic surgery identified 9 cosmetic surgery-specific PROMs developed with patient input that demonstrated at least adequate psychometric properties.³ The Blepharoplasty Outcomes Evaluation^{5,6} was the only eye-specific PROM identified by the search, but this 6-item instrument was excluded because its development process did not include patient input, which is considered essential.^{3,4} The reviewers concluded that research dedicated to the evaluation of PROMs in cosmetic surgery is urgently required.

PROMs are currently being used in many countries to inform clinical practice, comparative effectiveness research, discussions with regulatory bodies and an evidence-based approach to treatment.⁷ The United Kingdom was the first nation to formally mandate the collection of PROM data at the health system level. Starting in 2009, PROM data were collected for all National Health Service patients treated with the following 4 procedures: hernia repair, hip and knee replacement, and varicose veins treatments.⁸

Most recently in the United Kingdom, PROM data collection has been extended to include cosmetic surgery procedures. Following the publication of Sir Bruce Keogh's Review of the Regulations of Cosmetic Interventions,⁹ the UK Royal College of Surgeons set up the Cosmetic Surgery Interspecialty Committee to address the review's recommendations. This committee recommended a minimum data set to enable collection of clinical quality and outcome indicators.¹⁰ Specifically, all cosmetic surgery providers are expected to collect and submit a minimum data set, which includes PROM data, to the Private Healthcare Information Network.¹¹ The Cosmetic Surgery Interspecialty Committee recognized satisfaction with appearance as a key outcome for people seeking aesthetic treatments from plastic surgeons. Therefore, PROMs for 6 of 7 targeted cosmetic surgery procedures—abdominoplasty, augmentation mammoplasty, blepharoplasty, liposuction, rhinoplasty and rhytidectomy—measure

Key Points

Question How were the FACE-Q scales, designed to measure patient-reported outcomes following cosmetic eye treatments, developed and validated?

Findings The psychometric analysis provided evidence of the reliability and validity of the 4 (eyes, upper and lower eyelids, and eyelashes) FACE-Q Eye Module scales. In the Rasch Measurement Theory analysis, each scale's items had ordered thresholds and good item fit, and Pearson Separation Index and Cronbach α were greater than or equal to 0.83.

Meaning The FACE-Q can be used for the collection of evidence-based information about cosmetic eye treatments from the patient perspective.

appearance from the patient perspective using a subset of scales from the BREAST-Q,^{12,13} FACE-Q¹⁴⁻¹⁶ and BODY-Q.^{17,18} In addition, 2 adverse effect checklists from the FACE-Q¹⁵ have been recommended.

The FACE-Q^{14-16,19-21} is a PROM developed to address the lack of instruments for facial aesthetic procedures. In the Oxford review of cosmetic surgery PROMs,³ the FACE-Q was singled out as one of only 3 PROMs that met international recommendations for how such tools should be developed and validated. The FACE-Q includes over 40 independently functioning scales and checklists that measure appearance (of the face, specific facial areas, and rhytides), health-related quality of life, adverse effects of treatment, and the patient experience of care. The aim of this article is to describe the development and psychometric evaluation of the set of scales and checklist that can be used to evaluate cosmetic eye treatments. Specifically, the FACE-Q Eye Module that includes 4 appearance scales (ie, eyes overall, upper eyelids, lower eyelids and lashes) and a checklist measuring adverse effects following cosmetic eye treatments.

Methods

Prior to starting the study, institutional review board approval was obtained through the New School in New York City (United States) and through the University of British Columbia behavioral research ethics board in Vancouver (Canada). The FACE-Q was developed following internationally recommended guidelines for the development of a PROM.^{4,22-25} Our mixed methods approach to develop the FACE-Q has previously been described in detail.^{14-16,19-21,26} Briefly, qualitative interviews with 50 surgical and/or nonsurgical patients who had undergone facial aesthetic procedures and input from 26 experts in the field were used to develop the FACE-Q conceptual framework, scales, and checklists. These scales and checklists were further refined through cognitive interviews with 35 patients who had undergone facial aesthetic procedures. All FACE-Q scales were developed with 4 response options in keeping with best practice for scale development.²⁷ Instructions ask respondents to answer in relation to the past week.

Among both the qualitative and cognitive interview samples, patients having cosmetic eye treatments were well represented (qualitative = 25, cognitive interviews = 19). The 4 eye scales each contain 7 items that ask about appearance. The scales measuring eyes overall and eyelashes evaluate satisfaction (Very Dissatisfied, Somewhat Dissatisfied, Somewhat Satisfied, Very Satisfied). The scales measuring upper and lower eyelids and the adverse effect checklist measure the extent which someone is bothered by their appearance or an adverse effect (Not At All, A Little, Moderately, Extremely). Each item in each scale includes a unique descriptor that is used to measure a different aspect of appearance. Together, the items of a scale map out a clinical hierarchy for each concept of interest. For example, the eyelash scale measures appearance using the following 7 descriptors: nice, feminine, dark, long, attractive, thick, and full.

For validation purposes, all participants were also asked to complete the FACE-Q 10-item Satisfaction with Facial Appearance scale. In addition, some clinics included the FACE-Q 10-item Psychological Function scale and 8-item Social Function scale. All 3 of these FACE-Q scales previously demonstrated reliability, validity, and the ability to detect change.^{14,20} Participants were also asked questions that would allow us to characterize the sample, including sex and race/ethnicity.

Inclusion criteria were any patient aged 18 years or older who was pretreatment or posttreatment for 1 or more facial aesthetic treatments. Patients were recruited from 11 plastic surgery clinics and 3 dermatology clinics in the United States (n = 11) and Canada (n = 3). In 11 clinics, patients were recruited when they checked in for an appointment. In 3 clinics, patients were invited to participate via a postal survey. The survey included a personalized letter from the relevant physician along with the FACE-Q booklet, with up to 3 reminders mailed as necessary. All potential participants were provided a \$5 gift certificate to a coffee shop to thank them for their time. As this was a questionnaire survey study, completion of the FACE-Q booklet implied consent. Recruitment took place between June 6, 2010, and July 14, 2014.

Analysis

For the checklist, we computed the proportion of postoperative blepharoplasty patients that endorsed each of the 4 available response options for the 6 adverse effects.

For the appearance scales, we used Rasch Measurement Theory,²⁸ a modern psychometric approach, within RUMM2030 software.²⁹ Rasch Measurement Theory analysis examines the difference between observed and predicted responses for each item in a scale to determine if data for a scale fits a mathematical model.³⁰⁻³² A set of graphical and statistical tests were examined, with the results considered together to make decisions about the overall quality of each scale.³² The following tests, which are described more fully elsewhere,³¹ were conducted:

1. *Thresholds for item response options*: We examined thresholds between response options (eg, between Very Satisfied and Somewhat Satisfied) to determine if response categories scored with successive integers increased for the construct measured.
2. *Item fit statistics*: Three indicators of fit were examined to determine if each scale's items measured an unidimensional construct in the form of a clinical hierarchy: (1) log residuals (item-person interaction); (2) χ^2 values (item-trait interaction); and (3) item characteristic curves (ICC). Fit residuals should be between -2.5 and +2.5, and χ^2 values should be nonsignificant after Bonferroni adjustment.
3. *Targeting*: Person and item locations were examined to determine if the scales' items were evenly spread over a reasonable range that matched the range of the construct experienced by the sample.
4. *Dependency*: We examined residual correlations between items to identify correlations above 0.30 as high correlations can artificially inflate a scale's reliability.³⁰ A substest can be performed to determine how much correlated items affect scale reliability.
5. *Person separation index (PSI)*: We computed the PSI for each scale, which is a measure of the error associated with the measurement of people in a sample. The PSI is comparable to the Cronbach α .³³ Higher values indicate greater reliability.

In addition to the Rasch Measurement Theory outlined above, we computed Cronbach α ,³³ missing data, floor and ceiling effects, and the grade reading level for each scale.

The Rasch logit score for each participant's pattern of responses to a scale were transformed into scores from 0 (worst) to 100 (best). The scoring algorithm is available by contacting the corresponding author. To aid in the interpretation of the meaning of scores, we computed the implied range of scores for each response option based directly on the threshold plots produced through the Rasch analysis.^{29,34}

Using the 0 to 100 scores, we computed Pearson correlations to examine associations between scores and independent samples *t* tests to test for differences between means. We computed the number of posttreatment adverse effects experienced following a blepharoplasty and predicted that a higher number of adverse effects would correlate with lower scores on the 4 eye scales. We also predicted that fewer adverse effects and higher scores on the 4 eye scales would correlate with higher scores on the FACE-Q Satisfaction with Facial Appearance, Psychological Function, and Social Function scales. For patients in the pretreatment group, we expected that older participants would report lower scores on the 4 eye scales compared with younger patients. Finally, we predicted that pretreatment participants would report lower scores on all FACE-Q scales compared with posttreatment participants. In these analyses, *P* values less than .05 were considered statistically significant.

Results

Overall, 233 of 287 patients (81% response rate) participated. The response rate for face-to-face recruitment (n = 169 of 172 [98%]) was higher than that of mail-out with reminders (n = 64 of 115 [56%]). **Table 1** shows sample characteristics. Female participants (n = 192) composed 82% of the sample.

Table 1. Patient Characteristics

Characteristic	No. (%)
No. of participants	233
Age, mean (SD) [range], y	53.6 (12.4) [19-80]
Sex	
Female	192 (82.4)
Male	23 (9.9)
Missing	18 (7.7)
Race/ethnicity	
White	163 (70.0)
Other	46 (19.7)
Missing	24 (10.3)
Country	
United States	139 (59.7)
Canada	94 (40.3)
Assessments, No.	256
Times completed	
One	229 (89.5)
Two	12 (9.4)
Three	1 (1.2)
Timing of assessment	
Pretreatment	85 (33.2)
Posttreatment	171 (66.8)
Procedures	
Blepharoplasty	86 (33.6)
Blepharoplasty and other facial surgery	55 (21.5)
Other facial surgery	10 (3.9)
Minimally invasive procedure	105 (41.0)
Scales and/or checklists completed	
Satisfaction with eyes overall	173
Satisfaction with lashes	71
Upper eyelids	155
Lower eyelids	231
Adverse effects	96
Satisfaction with facial appearance	252
Psychological function	88
Social function	87

The adverse effects checklist was completed by 96 patients a mean (range) of 11 (0.5-54.0) months after blepharoplasty surgery. The most common adverse effects reported on the checklist included being bothered by eyelid scars (38%), dry eyes (35%), eye irritation (33%), excessive tearing (25%), eyes looking hollowed out (10%), and difficulty closing eyes (4%). eTable 1 in the Supplement shows the proportion of patients to report each adverse effect for each response option.

The Rasch Measurement Theory analysis supported the reliability and validity of the 4 eye scales. Each of the scales' 7 items had ordered thresholds. This finding provides evidence that each scale's 4 response options worked as a continuum for the construct measured. Table 2 shows the items for each scale sorted by the item locations; 25 of the 28 items had fit residuals within -2.5 to $+2.5$, which is the recommended range. All 28 were not significant in terms of Bonferroni adjusted χ^2 *P* values. Item residual correlations were above

Table 2. Rasch Measurement Theory Statistical Indicators of Fit

Scales, Items ^a	Item Location (SE)	FitResid	χ^2	<i>P</i> Value
Eyes overall				
Shape	-0.99 (0.17)	3.24	4.48	.11
Attractive	-0.31 (0.18)	-2.73	2.18	.34
Alert	0 (0.17)	-0.85	0.61	.74
Open	0.02 (0.17)	-2.41	2.67	.26
Bright-eyed	0.08 (0.18)	-1.19	0.34	.85
Nice	0.25 (0.18)	0.06	0.43	.81
Youthful	0.96 (0.18)	-0.33	0.51	.78
Eyelashes				
Nice	-2.74 (0.33)	-0.21	1.48	.48
Feminine	-0.06 (0.33)	-1.32	0.28	.87
Dark	0.16 (0.28)	1.7	7.29	.03
Long	0.43 (0.3)	-0.81	1.11	.57
Attractive	0.44 (0.35)	-1.62	1.3	.52
Thick	0.77 (0.34)	-0.8	3.38	.18
Full	1.01 (0.33)	-1.02	3.05	.22
Upper eyelid				
Skin on lashes	-0.78 (0.19)	2.81	5.19	.08
Saggy	-0.14 (0.19)	-1.98	5.27	.07
Droopy	0 (0.18)	-1.66	4.05	.13
Eyelid folds	0.15 (0.19)	2.32	6.74	.03
Heavy	0.2 (0.18)	0.61	0.38	.83
Look tired	0.27 (0.18)	-1.58	4.12	.13
Look old	0.3 (0.18)	-1.95	2.46	.29
Lower eyelids				
Excess fat	-0.82 (0.13)	-0.22	0.51	.78
Excess skin	-0.27 (0.12)	-1.65	3.75	.15
Puffiness	-0.08 (0.12)	1.31	0.55	.76
Noticeable lines	-0.03 (0.12)	1.36	0.71	.7
Crepey skin	0.28 (0.12)	0.71	0.23	.89
Look tired	0.42 (0.12)	-0.98	4.77	.09
Look old	0.51 (0.12)	-0.46	2.85	.24

Abbreviations: FitResid, Item Fit Residual; SE, Standard Error.

^a Items are in serial order for each scale.

0.30 for 6 pairs of items from 3 of the scales. We performed subtests on the pairs of items, which revealed marginal effect on scale reliability for 2 scales (Satisfaction with Eyes and Eyelashes; 0.01 and 0.03 difference in PSI) but a larger effect for the scale measuring appearance of the upper eyelids (ie, 0.08 decrease in PSI to 0.80).

The *P* values for fit to the Rasch model were not significant for 3 scales (eyes overall, lower eyelids, eyelashes), providing support for the data satisfying the requirements of the Rasch model. For the remaining scale (upper eyelids) the *P* value was 0.01. Pearson Separation Index and Cronbach α were 0.83 or greater. Other scale level findings are shown in Table 3. The 4 scales were easy for participants to comprehend and had minimal missing data. eTable 2 in the Supplement provides a FACE-Q interpretation table as an example. This table shows the implied range of scores for each of the possible responses for the 7-item overall eyes scale.

Table 3. Reliability Statistics

Scale	PSI Without Extremes	Cronbach α	Participant Scale Completion, %			FK Grade
			Floor Pre/Post	Ceiling Pre/Post	Missing Data	
Eyes overall	0.90	0.97	2/4	7/40	4	2.1 (0.5-5.2)
Eyelashes	0.90	0.97	6/8	22/15	6	6.2 (5.2-8.1)
Upper eyelids	0.88	0.97	9/0	13/59	7	5.7 (3.7-13.1)
Lower eyelids	0.83	0.93	3/1	12/28	7	4.8 (3.6-6.7)

Abbreviations: FK, Flesch-Kincaid grade level; Pre/Post, pretreatment vs posttreatment; PSI, Person Separation Index.

Table 4. Pearson Correlations Between Adverse Effects and Appearance Scales

	Adverse Effects	Eyes	Eyelashes	Upper Eyelids	Lower Eyelids
Adverse effects					
R	1	-.255 ^a	-.280 ^a	-.300 ^b	-.362 ^b
P Value015	.049	.005	.001
Participants, No.	93	90	50	85	80
Eyes overall					
R	-.255 ^a	1	.440 ^b	.672 ^b	.588 ^b
P Value	.015		.000	.000	.000
Participants, No.	90	173	71	152	151
Eyelashes					
R	-.280 ^a	.440 ^b	1	.285 ^a	.426 ^b
P Value	.049	.000024	.001
Participants, No.	50	71	71	63	55
Upper eyelids					
R	-.300 ^b	.672 ^b	.285 ^a	1	.645 ^b
P Value	.005	.000	.024000
Participants, No.	85	152	63	155	137
Lower eyelids					
R	-.362 ^b	.588 ^b	.426 ^b	.645 ^b	1
P Value	.001	.000	.001	.000	...
Participants, No.	80	151	55	137	231
Facial appearance					
R	-.236 ^a	.608 ^b	.274 ^a	.612 ^b	.624 ^b
P Value	.024	.000	.021	.000	.000
Participants, No.	92	172	70	153	228
Psychological function					
R	-.416 ^a	.431 ^b		.480 ^b	.589 ^b
P Value	.014	.000		.000	.000
Participants, No.	34	80		74	82
Social function					
R	-.357 ^a	.251 ^a		.378 ^b	.369 ^b
P Value	.038	.024		.001	.001
Participants, No.	34	81		75	83

Abbreviations: R, correlation between scales; ellipses, not applicable or no data available.

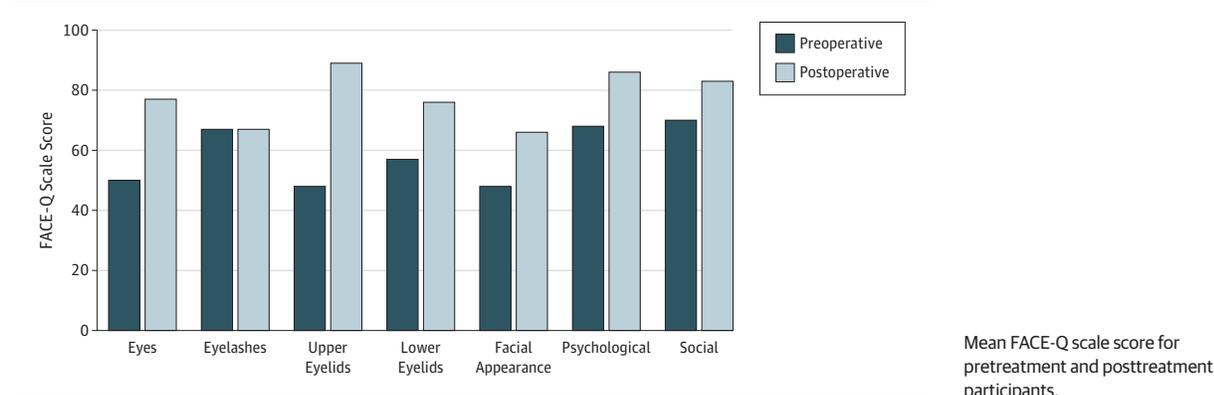
^a Correlation significant at the 0.05 level (2-tailed).

^b Correlation significant at the 0.01 level (2-tailed).

Table 4 shows the correlation findings. A higher number of adverse effects following a blepharoplasty correlated with lower scores on the 4 eye scales as well as the Satisfaction with Facial Appearance, Psychological Function, and Social Function scales. A higher score on the 4 eye scales were significantly correlated with higher scores on the Satisfaction with Facial Appearance scale. Higher scores on the 3 eye scales (no data for eyelash scale) correlated with higher scores on the Psychological Function and Social Function scales.

In the pretreatment group, older age correlated with lower scores on the eye overall ($R = -0.42$; $P = .001$), upper eyelids ($R = -0.51$; $P < .001$), and lower eyelids ($R = -0.42$; $P < .001$) scales as well as Satisfaction with Facial Appearance scale ($R = -0.35$; $P = .001$). In the posttreatment group, the correlations between older age and appearance was significant for the lower eyelid scale ($R = 0.21$; $P = .01$) and on the Satisfaction with Facial Appearance scale ($R = 0.21$; $P = .01$). Here, older age was associated with reporting more satisfaction with appearance.

Figure. Mean FACE-Q Scale Scores



The **Figure** shows the mean scores for pretreatment and posttreatment participants for the 4 eye scales, as well as the Satisfaction with Facial Appearance, Psychological Function, and Social Function scales. For 6 of the 7 scales (exception, eyelash scale), the *P* value for independent samples *t* tests were significant ($P \leq .002$).

Discussion

The FACE-Q is currently the only PROM developed following international recommendations that measures appearance and other concepts of interest important to patients who have undergone facial aesthetic procedures. The psychometric analyses described in this article provide evidence of reliability and validity of the 4 scales that compose the FACE-Q Eye Module. For blepharoplasty patients, appearance of the eyes was found to correlate with the number of postoperative adverse effects experienced. Pretreatment patients reported lower scores for appearance of the eyes and upper and lower eyelids and satisfaction with facial appearance overall and psychological and social function compared with patients who underwent a cosmetic treatment. Prospective studies of clinical change are now needed to determine the magnitude of change for cosmetic eye treatments instead of blepharoplasty.

As cosmetic surgery providers in the United Kingdom proceed to collect PROM data on a national level for the first time ever, normative data on important patient outcomes for a range of cosmetic surgery procedures will be compiled. In the United Kingdom quality initiative, the recommended PROM to measure outcomes following blepharoplasty is the FACE-Q 7-item satisfaction with eyes overall, which should be administered before and after treatment. In addition, the FACE-Q adverse effects checklist for eyes was recommended for use postoperatively. Such data can be used to empower patients, inform decision making, identify patients most likely to respond to treatment, and support quality improvement. For example,

findings about the incidence of bothersome adverse effects for different cosmetic surgical procedures could be used by plastic surgeons to assure that patients are properly educated prior to surgery.

The FACE-Q study has certain limitations that have been previously reported.^{14-16,19-21} Specifically, the study sample described here varied by age, sex, timing of assessment, and type of treatment. These factors limit our ability to report findings beyond instrument development and validation. Cosmetic surgery is sought mainly by women and white patients.¹ As such, our sample of patients in this FACE-Q study sample had more women than men and primarily white participants. Additionally, it is possible that the office staff recruiting patients for the field test sample may have been biased in their selective recruitment of patients. Finally, as mentioned above, although we included pretreatment and posttreatment patients in the study, the number of patients who provided data before and after treatment was too small to compute change scores. Elsewhere, in a sample of close to 1000 cosmetic patients, we showed that the mean scores on the FACE-Q Satisfaction with Appearance scale for 5 treatments (botulinum toxin type A, filler, rhinoplasty, facelift, or blepharoplasty) were significantly higher among those who underwent treatment compared with those who did not.³⁵ Responsiveness research is now needed to document the benefits of cosmetic eye treatments.

Conclusions

As the cosmetic surgery industry continues to expand worldwide, collection of evidence-based information regarding patient outcomes is essential. The FACE-Q Eye Module, developed and validated using state-of-the-art qualitative and quantitative psychometric research methods, is a valuable new tool that can help researchers, clinicians, and regulatory bodies accomplish this goal using eye-specific PROMs.

ARTICLE INFORMATION

Accepted for Publication: June 27, 2016.

Published Online: September 15, 2016.
doi:10.1001/jamafacial.2016.1018

Author Affiliations: McMaster University, Hamilton, Ontario L8S4K1, Canada (Klassen); Can to Modus Outcomes, Letchworth Garden City, England (Cano); Grotting & Cohn Plastic Surgery,

Birmingham, Alabama (Grotting); MedStar Georgetown University Hospital, Washington, DC (Baker); Department of Ophthalmology, University of British Columbia, Vancouver, Canada (J. Carruthers); Department of Dermatology and Skin Science, University of British Columbia, Vancouver, Canada (A. Carruthers); Division of Plastic Surgery, University of British Columbia, Vancouver, BC, Canada (Van Laeken); UC Davis Medical Center, Sacramento, California (Sykes); MedStar Georgetown University Hospital, Washington, DC (Schwitzer); Memorial Sloan Kettering Cancer Center, New York, New York (Pusic).

Author Contributions: Drs Klassen and Pusic had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Klassen, Cano, Van Laeken, Sykes, Schwitzer, Pusic.

Acquisition, analysis, or interpretation of data: Klassen, Cano, Grotting, Baker, Carruthers, Carruthers, Schwitzer, Pusic.

Drafting of the manuscript: Klassen, Cano, Schwitzer, Pusic.

Critical revision of the manuscript for important intellectual content: All Authors.

Statistical analysis: Klassen, Cano, Schwitzer, Pusic.

Obtaining funding: Pusic.

Administrative, technical, or material support: Baker, Van Laeken, Sykes, Schwitzer, Pusic.

Study supervision: Carruthers, Pusic.

Provide patients for the study, aid in study design: Van Laeken.

Editorial critique of manuscript: Grotting.

Conflict of Interest Disclosures: The FACE-Q is owned by Memorial Sloan Kettering Cancer Center. Drs Cano, Klassen and Pusic are codevelopers of the FACE-Q and, as such, receive a share of any license revenues as royalties based on Memorial Sloan Kettering Cancer Center's inventor sharing policy. Financial disclosure (all other relationships): Dr Cano is cofounder of Modus Outcomes, an outcomes research and consulting firm that provides services to pharmaceutical, medical device, and biotechnology companies. Drs Alastair Carruthers and Jean Carruthers are consultants and investigators for Allergan, Merz, Kythera and Alphasen.

Funding/Support: This study was supported by a grant from the Plastic Surgery Foundation. Dr Pusic received support from the NIH/NCI Cancer Center (support grant No. P30 CA008748).

Role of the Funder/Sponsor: The funders/sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. The American Society of Plastic Surgeons. 2015 Plastic Surgery Statistics Report. <http://www.plasticsurgery.org/Documents/news-resources/statistics/2015-statistics/plastic-surgery-statistics-full-report.pdf>. Accessed July 20, 2016.
2. The British Association of Aesthetic Plastic Surgeons. Tweak not tuck. <http://baaps.org.uk/about-us/press-releases/2039-auto-generate-from-title>. Accessed March 8, 2016.

3. Morley D, Jenkinson C, Fitzpatrick R. A structured review of patient-reported outcome measures used in cosmetic surgical procedures: report to Department of Health, 2013. <http://phi.uhce.ox.ac.uk/pdf/Cosmetic%20Surgery%20PROMs%20Review2013.pdf>. Accessed March 22, 2016.
4. US Department of Health and Human Services. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. <http://www.ispor.org/workpaper/FDA%20PRO%20Guidance.pdf>. Accessed March 8, 2016.
5. Alsarraf R, Larrabee WF Jr, Anderson S, Murakami CS, Johnson CM Jr. Measuring cosmetic facial plastic surgery outcomes: a pilot study. *Arch Facial Plast Surg*. 2001;3(3):198-201.
6. Alsarraf R. Outcomes research in facial plastic surgery: a review and new directions. *Aesthetic Plast Surg*. 2000;24(3):192-197.
7. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167.
8. Department of Health. Guidance on the routine collection of patient reported outcome measures. http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_092647. Accessed March 8, 2016.
9. Department of Health. Review of the Regulation of Cosmetic Interventions. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192028/Review_of_the_Regulation_of_Cosmetic_Interventions.pdf. Accessed March 22, 2016.
10. Royal College of Surgeons. Patient Reported Outcome Measures. <http://www.rcseng.ac.uk/surgeons/surgical-standards/working-practices/cosmetic-surgery/datasets-and-qproms/patient-reported-outcome-measures>. Accessed March 22, 2016.
11. Private Healthcare Information Network. <http://www.phin.org.uk/>. Accessed March 22, 2016.
12. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg*. 2009;124(2):345-353.
13. Cano SJ, Klassen AF, Scott AM, Cordeiro PG, Pusic AL. The BREAST-Q: further validation in independent clinical samples. *Plast Reconstr Surg*. 2012;129(2):293-302.
14. Pusic A, Klassen AF, Scott AM, Cano SJ, 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(2):249-260.
15. 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(1):21-30.
16. Klassen AF, Cano SJ, East CA, et al. Development and psychometric evaluation of the FACE-Q scales for patients undergoing rhinoplasty. *JAMA Facial Plast Surg*. 2016;18(1):27-35.

17. Klassen AF, Cano SJ, Soldin M, et al. Klassen AF, Cano SJ, Alderman A, et al. The BODY-Q: a patient-reported outcome instrument for weight loss and body contouring treatments. *Plast Reconstr Surg Glob Open*. 2016;4(4):e679.
18. Klassen AF, Cano SJ, Scott A, Tsangaris E, Pusic AL. Assessing outcomes in body contouring. *Clin Plast Surg*. 2014;41(4):645-654.
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(4):443-451.
20. Klassen AF, Cano SJ, Schwitzer JA, Scott AM, Pusic 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(2):375-386.
21. Panchapakesan V, Klassen AF, Cano SJ, Scott AM, Pusic AL. Development and psychometric evaluation of the FACE-Q Aging Appraisal Scale and Patient-Perceived Age Visual Analog Scale. *Aesthet Surg J*. 2013;33(8):1099-1109.
22. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res*. 2009;18(9):1263-1278.
23. 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(8):967-977.
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 2—assessing respondent understanding. *Value Health*. 2011;14(8):978-988.
25. Cano SJ, Klassen A, Pusic AL. The science behind quality-of-life measurement: a primer for plastic surgeons. *Plast Reconstr Surg*. 2009;123(3):98e-106e.
26. Kosowski TR, McCarthy C, Reavey PL, et al. A systematic review of patient-reported outcome measures after facial cosmetic surgery and/or nonsurgical facial rejuvenation. *Plast Reconstr Surg*. 2009;123(6):1819-1827.
27. Khadka J, Gothwal VK, McAlinden C, Lamoureux EL, Pesudovs K. The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes*. 2012;10:80.
28. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmark: Danish Institute for Education Research; 1960.
29. Andrich D, Sheridan B. RUMM2030. Perth (Australia): RUMM Laboratory; 1997-2011. <http://www.rummlab.com.au/>. Accessed July 13, 2016.
30. Wright B, Masters G. *Rating scale analysis: Rasch measurement*. Chicago, Illinois: MESA; 1982.
31. Andrich D. *Rasch Models for Measurement*. Newbury Park, California: Sage; 1988.

32. Hobart J, Cano S. Improving the evaluation of therapeutic intervention in MS: the role of new psychometric methods. *Health Technol Assess*. 2009;13(12):1-177.

33. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297-334.

34. Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11:571-585.

35. Klassen AF, Cano SJ, Pusic AL. FACE-Q Satisfaction with Appearance Scores from Close to

1000 Facial Aesthetic Patients. *Plast Reconstr Surg*. 2016;137(3):651e-652e.

Invited Commentary

Using Patient-Reported Outcome Measures to Optimize Results—From Good to Great

Lisa E. Ishii, MD, MHS

As surgeons we strive for improvement with every procedure, always aiming for better outcomes on the next procedure than the one that immediately preceded it. We typically use multiple methods of performance improvement, including utilizing preoperative planning with photographs, morphing, and sketches, as well as detailed surgical documentation. We evaluate our technical modifications by examining our surgical results, frequently drawing on standardized surgical outcome instruments. Fortunately for us, we have an increasing number of patient-reported outcome measures (PROMs) in our toolbox, a previously underused tool in our approach to performance improvement.¹ By inviting the direct feedback from patients through validated PROMs we gain critical information from the population whose opinion matters the most. In this issue of *JAMA Facial Plastic Surgery*, Klassen et al² share valuable information on the development and psychometric testing of the FACE-Q Cosmetic Eye Module.

The FACE-Q is a PROM for facial cosmetic procedures, consisting of more than 40 independently functioning scales and checklists that measure appearance, health-related quality of life, and adverse outcomes.³ Developed using Rasch measurement theory,⁴ it consists of multiple modules addressing different areas, including rhinoplasty, rhytidectomy, and cosmetic eye procedures. The development and psychometric testing of the rhinoplasty module was described in a previous issue of *JAMA Facial Plastic Surgery*.¹ The FACE-Q module described by Klassen et al² is the PROM following cosmetic eye treatments.

Briefly, Klassen et al² describe the development and psychometric evaluation of the module over a 4-year period from 2010 to 2014 in a number of plastic surgery practices across the United States and Canada. Patients completed the scales and checklists before and after procedures, with 233 patients participating specifically in the cosmetic eye treatment section. Psychometric evaluation showed ordered thresholds and good item fit.

Notably, Klassen et al² identified a correlation between higher scores on the 4 eye scales and higher scores on the Satisfaction with Facial Appearance scale and also between higher scores on 3 eye scales (no data for the eyelash scale) and higher

scores on the Psychological Function and Social Function Scales. These are important correlations to consider because the driving factor for patients to engage in facial rejuvenation procedures, including eye procedures, may ultimately be to improve their overall appearance and their social integration. An awareness of this link may benefit prospective patients considering the risks and benefits for the procedures under consideration.

The FACE-Q Eye Module is a promising tool for clinical practice and outcome assessment related to cosmetic eye procedures. Feasible to administer to patients, it requires only a few minutes for completion. Presumably completion could occur in the office setting or through forms mailed or emailed to patients. If the office is the preferred setting, it can easily be added into the clinic flow through tablet administration or with paper forms as a single module or in association with other modules or forms. Thus, surgeons can hear directly from patients about their own perception of outcome. These data can then be incorporated into the ultimate evaluation of surgical outcome.

It is encouraging that PROMs are increasing in availability and acceptance.^{5,6} As we move toward a new patient-centered and evidence-based care model in the United States we can take advantage of the patient's perspective as we never have before. We are on the cusp of fully appreciating the value of structured patient feedback for improving the treatments we provide. Further, when we standardize our data collection forms and agree that specific structured data elements should be gathered for specific clinical encounters, we will then have the ability to aggregate these data in large clinical data registries.^{7,8} For example, Regent, the new qualified clinical data registry for otolaryngology-head & neck surgery, will have a facial plastic surgery module where such data can be aggregated and disseminated. Then large aggregated patient outcome data can be analyzed for quality improvement and research purposes in unprecedented ways. With large sets of aggregated data from thousands of patients at multiple sites, all maintained in a single central data repository, one may have the ability to compare different techniques head to head as they relate to patient reported outcomes. This presents an opportunity for us to advance our specialty in an evidence-driven manner never before possible. Our colleagues in other surgical specialties, such as the Society for Thoracic Surgeons and

 Related article [page 7](#)