Development and validation of a novel patient-reported treatment satisfaction measure for hyperfunctional facial lines: facial line satisfaction questionnaire

Farrah Pompilus, MA,1 Somali Burgess, PhD,2,† Stacie Hudgens, MA,3,‡ Benjamin Banderas, BS,1 & Selena Daniels, PharmD, MS4

1Adelphi Values, Boston, MA, USA
2Strategic Outcomes Services, Mission Viejo, CA, USA
3Clinical Outcomes Solutions, Tucson, AZ, USA
4Allergan Inc., Irvine, CA, USA

Summary

Background Facial lines or wrinkles are among the most visible signs of aging, and minimally invasive cosmetic procedures are becoming increasingly popular.

Aims The aim of this study was to develop and validate the Facial Line Satisfaction Questionnaire (FLSQ) for use in adults with upper facial lines (UFL).

Methods A literature review, concept elicitation interviews (n = 33), and cognitive debriefing interviews (n = 23) of adults with UFL were conducted to develop the FLSQ. The FLSQ comprises Baseline and Follow-up versions and was field-tested with 150 subjects in a US observational study designed to assess its psychometric performance. Analyses included acceptability (item and scale distribution [i.e. missingness, floor, and ceiling effects]), reliability, and validity (including concurrent validity).

Results In total, 69 concepts were elicited during patient interviews. Following cognitive debriefing interviews, the FLSQ-Baseline version included 11 items and the Follow-up version included 13 items. Response rates for the FLSQ were 100% and 73% at baseline and follow-up, respectively: no items had excessive missing data. Questionnaire scale scores were normally distributed. Most domain scores demonstrated good internal consistency reliability (Cronbach’s $\alpha \geq 0.70$). Most items within their respective domains exhibited good convergent (item-scale correlations $> 0.40$) and discriminant (items had higher correlation with their hypothesized scales than other scales) validity. Concurrent validity correlation coefficients of the FLSQ domain scores with the associated concurrent measures were acceptable (range: $r = 0.40–0.70$). Six FLSQ items demonstrated reliability and validity as stand-alone items outside their domains.

Conclusions The FLSQ is a valid questionnaire for assessing treatment expectations, satisfaction, impact, and preference in adults with UFL.

Keywords: patient satisfaction, facial lines, cosmetic

Introduction

The development of upper facial lines (UFL) or wrinkles is one of the most visible signs of aging. These lines form slowly over time: repeated contraction of the underlying facial musculature, coupled with the loss of collagen and elastin fibers over time, leads to skin inelasticity and the formation of facial lines or wrinkles.1 With age, facial lines tend to become static and are visible even without facial muscle contraction (i.e. at rest).1–3 There
are three discrete UFL most noticed by patients and clinicians: lateral canthal lines in the periorbital region (hereafter referred to as crow’s feet lines [CFL]), transverse forehead lines (FHL), and glabellar lines (GL).2

Minimally invasive esthetic procedures have become increasingly popular as aging and its impact on physical appearance continue to be of concern to the general populace. Traditionally, physicians have placed an emphasis on optimizing the physical outcome of these esthetic procedures and minimizing associated side effects; however, a range of patient-oriented factors influence the perception of treatment success, including patient satisfaction.4

Patient satisfaction has been demonstrated to be an important health outcome in a variety of therapeutic areas,5 and it is likely equally if not more relevant in the field of facial esthetics, due to the subjective nature of esthetic perception. Satisfaction is a multidimensional concept, comprising a number of elements that vary from patient to patient, such as treatment expectations, perceived convenience, and aesthetic preference.6 High patient satisfaction with esthetic procedures has also been associated with improvements in self-esteem and body image.7,8,9 Although many treatment satisfaction questionnaires exist, few have been developed in accordance with the United States (US) Food and Drug Administration’s (FDA’s) Guidance for Industry – Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims, published in 2009 (hereafter referred to as FDA PRO Guidance).10

In order to address this gap, a new facial line treatment satisfaction questionnaire was developed to assess treatment expectations and satisfaction, impact, and preference in adults with UFL (hereafter referred to as facial lines), including CFL, FHL, and GL. In this article, the development and psychometric evaluation of the Facial Line Satisfaction Questionnaire (FLSQ) are described.

Methods

The FLSQ was developed in two phases: (1) qualitative and (2) quantitative. The qualitative phase comprised a review of the literature and elicitation of interviews to identify relevant concepts regarding satisfaction, convenience, and preference. This was followed by face-to-face cognitive interviews to test the validity of the questionnaire. The quantitative phase comprised defining the study population, data collection, and psychometric analyses of the newly developed questionnaires. Figure 1 outlines the phases of the development and validation processes.

Qualitative phase

Literature review

Prior to the conduct of qualitative research activities, in accordance with FDA recommendations,10 a preliminary conceptual framework was developed. The conceptual framework (which contains concepts and domains of a disease/state) was initially informed by a brief review of the literature (a search of MEDLINE® for relevant articles) and further refined in light of qualitative feedback from patients and clinicians.11 The framework served as the basis for the development of the concept elicitation (CE) qualitative interview guide, as well as for the analysis of qualitative interviews.

Concept elicitation interviews

Following the literature review, a series of CE interviews were conducted to identify relevant concepts related to facial line treatment satisfaction with respect to three UFL: CFL, FHL, and GL. Ethics approval from an independent review board (IRB) was obtained for all study documents and the study protocol prior to initiating these interviews. Before any study activities commenced, subjects were required to provide written consent to be a participant in the study. A total of 33 subjects were recruited through clinicians at three separate sites in the United States. Interviews were conducted in a private room at the recruiting clinician’s office or a nearby interview facility. Inclusion and exclusion criteria are provided in Table 1. Recruitment for the study began in June 2010, and all data were acquired by the end of July 2010. All 33 recruited subjects were interviewed in person, with each interview lasting approximately 60 min. All interviews were audio- and video-recorded, as well as transcribed and anonymized with subjects’ prior written consent. Transcripts were analyzed on an ongoing basis using a modified grounded theory approach.12,13 ATLAS.ti, a computerized qualitative data analysis package (Berlin, Germany), was used to facilitate this process. Subjects were compensated $100 each for their time. Demographic and Health Information Forms were also completed by each patient.

Sample size sufficiency was confirmed based on saturation (the point at which no new or distinctive categories, high level concepts, or substantive codes are likely to be generated from further qualitative data collection). Concept saturation was assessed by documenting concept emergence across sets of successive interviews, according to established methods.14–16
Item generation
Findings from CE interviews were used to generate items for a draft questionnaire. Based on the FDA PRO Guidance, items were generated and included according to the following criteria: saturation; translatability; applicability; lack of jargon, double-enquiry, and value-laden terms; response option consistency; and frequency of spontaneous mention.

Cognitive interviews
Following item generation, cognitive debriefing interviews (CIs) were conducted to assess subjects’ interpretation and understanding of the newly developed facial line treatment satisfaction PRO questionnaire (i.e. FLSQ) and its instructions. Twenty-three new subjects were recruited (from the same three sites in the United States as the CE interviews) and interviewed. Inclusion and exclusion criteria were identical to those for the CE interviews. Recruitment for the study began in September 2010, and all data were acquired by the end of November 2010. The length of each interview was approximately 90 min, and all interviews were conducted and analyzed using methods similar to those for the CE interviews. Subjects were compensated $100 each for their time. Demographic and Health Information Forms were also completed by each patient. In the CIs, subjects were asked to complete the FLSQ using a “think aloud” method to reduce interviewer
Participating subjects must have/be:
- Inclusion Criteria
- Signed Informed Consent Form
- 18-65 years old
- Moderate or severe CFL at maximum smile, FHL at maximum eyebrow elevation and/or GL at maximum frown
- Fluent in US English (both reading and writing)
- Capable and willing to participate in a face-to-face interview

Facial line types.

Range from 0-3: none, mild, moderate, and severe on each of the *A proprietary photo-numeric rating scale. Ratings on the scale

Exclusion Criteria
- Participating subjects must not have:
  - Severe resting CFL, FHL, and/or GL as measured by the investigator-rated FWS*
  - One or more of the following:
    - Marked facial asymmetry
    - Dermatochalasis
    - Deep dermal scarring
    - Excessively thick sebaceous skin
    - Brow or eyelid ptosis
    - Infection or skin disorder at injection sites
    - History of facial nerve palsy
    - History of alcohol abuse
    - Cognitive impairment that would interfere with participation in an interview
    - Severe psychiatric disorder
    - Other condition or situation that may put the subject at significant risk, confound study results, or interfere significantly with the subject’s participation in the study (e.g., inability to complete study visits or read and understand PRO instruments
  - Previously undergone any of the following aesthetic procedures:
    - In the three months prior to enrollment:
      - Non-ablative resurfacing laser or light treatment
      - Microdermabrasion
      - Superficial peels
    - In the six months prior to enrollment:
      - Cosmetic procedures with medium depth to deep facial chemical peels (e.g., TCA and phenol)
      - Mid-facial or periorbital skin resurfacing or permanent make-up
    - In the 12 months prior to enrollment:
      - Mid-facial or periorbital treatment with non-permanent soft tissue fillers
    - Prior to enrollment:
      - Periorbital surgery
      - Facial lift (full face or mid-face)
      - Brow lift or related procedure (e.g., eyelid and/or eyebrow surgery)
      - Mid-facial or periorbital treatment with permanent soft tissue fillers, gortex implantation, and/or autologous fat transplantation

| CFL = Crow’s Feet Lines, FHL = Forehead Lines, GL = Glabellar Lines.
* A proprietary photo-numeric rating scale. Ratings on the scale range from 0-3: none, mild, moderate, and severe on each of the facial line types.

Criteria from the FDA PRO Guidance were considered, including clarity and ease of comprehension on first reading, absence of ambiguity, absence of clinical terminology, and the absence of double-entendre. The draft PRO questionnaire was revised based on the results of the CIs, and all modifications were documented.

**Quantitative phase**

**Study population and data collection**

Approval from an IRB was obtained for all study documents and the study protocol prior to recruitment. Similar to the qualitative phase and prior to any research activities, subjects were to complete a written informed consent form in order to participate in the study. Recruitment for the study began in February 2011, and all data were acquired by the end of June 2011.

The first assessment was conducted in the clinic for the purpose of screening subjects into the study. The second assessment occurred 1–2 weeks later and consisted of a packet of questionnaires completed at home and returned by mail. A number of instruments were administered in parallel with the FLSQ for descriptive and/or concurrent validity analysis: DHIF, Facial Line Outcomes (FLO-11) Questionnaire, Treatment Satisfaction Questionnaire for Medication vII (TSQM), and the investigator-rated Facial Wrinkle Scale with Photonumeric Guide (FWS). The subject and clinician schedule of assessments is provided in Table 2.

The DHIF, FLO-11, TSQM, FWS, and FLSQ (Baseline version) were administered during each subject’s initial visit. A follow-up assessment consisted of the FLSQ (Follow-up version), FLO-11, and TSQM. Descriptive and traditional psychometric analyses were performed using SAS version 9.2 (SAS Institute, Cary NC).

**Traditional psychometric analyses**

Traditional psychometric analyses were outlined in a statistical analysis plan and performed on the FLSQ domain and select single-item scores to determine the reliability and validity of the FLSQ. Specifically, the following psychometric measurement properties were examined: acceptability (item and scale distribution [i.e., missingness no > 10%, floor and ceiling effects no > 20%]), reliability (internal consistency reliability [Cronbach’s ≥ 0.70]), and validity (including concurrent validity r = 0.40–0.70). Several items were identified for individual reliability and validity assessment, as they were considered potential “stand-alone” items. These items were included on the FLSQ-Follow-up version: satisfaction with treatment onset; satisfaction with treatment duration; satisfaction with receiving a natural look; overall treatment satisfaction;
likelihood to continue treatment; and extent to which treatment expectations were met. Tests and criteria are summarized in Table 3.

Results

Qualitative phase

Literature review

The preliminary conceptual framework developed on the basis of a brief review of the literature initially contained 31 concepts divided across eight domains (Treatment Procedure, Outcomes [Treatment Efficacy], Emotional Outcomes, Social Outcomes, Willingness to Continue Treatment, Treatment Preference, Convenience and Cost) encompassing evaluations of both treatment outcomes and satisfaction. Additional data on the conceptual framework can be provided by the authors on request.

Concept elicitation interviews

Study population. Thirty-three subjects (female = 90.9%; mean age = 45.1 ± 16.9 years) were recruited and participated in CE interviews. These subjects were included in the analysis. Subjects’ demographic and health information is summarized in Table 4.

Table 2 Validation schedule of assessments

| Protocol activities and forms completed | Study recruitment (visit 1) | Follow-up (visit 2; 7–10 days postvisit 1) |
|----------------------------------------|-----------------------------|------------------------------------------|
| Information letter                      | C, S                        | –                                        |
| Clinician profile                      | C                           | –                                        |
| Subject information and consent form   | C, S                        | –                                        |
| including an authorization to use and disclose personal health information for research form | | |
| Demographic and health information form | S                           | –                                        |
| Case report form                       | C                           | –                                        |
| Investigator-rated FWS                 | C                           | –                                        |
| FLSQ (baseline version)                | S                           | –                                        |
| FLSQ (follow-up version)               | –                           | S                                        |
| FLO-11                                 | S                           | S                                        |
| TSQM                                   | S                           | S                                        |

C, clinician; S, subject; FLO-11, facial line outcomes questionnaire; FLSQ, facial line satisfaction questionnaire; FWS, facial wrinkle scale with photonumeric guide; TSQM, treatment satisfaction questionnaire measure.

Interview results. Subjects’ description of their facial lines and associated impacts and their reactions and expectations regarding treatment (Fig. 2) were recorded for each facial line (CFL, FHL, and GL). Sixty-nine concept codes and their associated verbatim quotes were reviewed to identify whether certain concepts were more strongly associated with particular line types, and redundant codes were merged. The most frequently mentioned concepts common to all facial lines are presented in Figure 2, which contains the complete listing of consolidated concepts and their frequencies.

Overall, subjects were acutely aware of the appearance of their facial lines, and in most cases, a plurality indicated that their facial lines made them look and feel older, with both impacts reported most commonly for forehead lines (n = 15, 45.5% and n = 11, 33.3%, respectively). Subjects also reported using a variety of treatments for their UFL, including moisturizers, serums, chemical peels, lasers, facials, fillers, and other topical therapies.

Item generation

In order to gain an accurate and meaningful appreciation of patient treatment expectations and satisfaction, the initial draft instrument was divided into two versions, the FLSQ-Baseline and FLSQ-Follow-up, to be administered before and after treatment, respectively. The draft FLSQ-Baseline and FLSQ-Follow-up contained 16 items and 21 items, respectively; response options were three-, five-, or six-point Likert-type/adjectival scales. Recall period was defined as the present time (i.e., at the time of questionnaire administration). Wording was derived from subject language recorded in CE interviews. Once the items were drafted, two domains were identified in the FLSQ-Baseline (Treatment Expectations and Impact) and five domains in the FLSQ-Follow-up (Treatment Satisfaction, Impact, Met Treatment Expectations, Continue Treatment, and Recommend to Others).

FLSQ-Baseline. Item concepts related to Treatment Expectations were as follows: looking the age one wants to look, improved facial looks, improved facial appearance, treatment onset, treatment duration, receiving a natural look, receiving a fresh look, treatment effect on facial lines, improved self-esteem, overall expectations for treatment, and improved self-esteem. Item concepts related to Impact of Facial Lines were as follows: feeling older, negatively affected self-esteem, looking tired, being bothered by facial lines, feeling unhappy about facial lines, and looking angry.
Follow-up assessment was conducted 7–10 days following baseline assessment. Item concepts related to Treatment Satisfaction were as follows: looking the age one wants to look, improved facial looks, improved facial appearance, treatment onset, treatment duration, receiving a natural look, receiving a fresh look, treatment effect on facial lines, improved self-esteem, overall treatment results, and overall treatment satisfaction. Item concepts related to Impact of Facial Lines were as follows: feeling older, negatively affected self-esteem, looking tired, being bothered by facial lines, feeling unhappy about facial lines, and looking angry.

The initial FLSQ-Follow-up also included the following item concepts (comprising the remaining domains of Met Treatment Expectation, Continue Treatment, and Recommend to Others): treatment convenience, likelihood of continuing treatment, likelihood of recommending treatment to a friend, extent to which expectations for treatment results were met, and extent to which treatment expectations were met.

Cognitive interviews

Study population. Twenty-three subjects (female = 95.6%; mean age = 45.9 ± 14.7 years) participated in CIs and were included in the analysis. Subjects’ demographic and health information is summarized in Table 4.

Interview results. Based on the feedback from the subjects, five items on the FLSQ-Baseline and seven items on the FLSQ-Follow-up were deleted due to redundancy with other items; one item on the FLSQ-Follow-up was also eliminated due to multiple interpretations.
### Table 4 Demographic characteristics

| Characteristic                      | Concept elicitation total (N = 33) | Cognitive interviews total (N = 23) | Psychometric validation total (N = 155) |
|-------------------------------------|------------------------------------|------------------------------------|----------------------------------------|
|                                     | Concept elicitation total (N = 33) | Cognitive interviews total (N = 23) | Psychometric validation total (N = 155) |
| Gender n (%)                        | Male 3 (9.1)                       | 1 (4.3)                            | 22 (14.2)                              |
|                                     | Female 30 (90.9)                   | 22 (95.7)                          | 133 (85.8)                             |
| Age                                 | Range 24–64                        | 36–61                              | 27.1–62.7                              |
|                                     | Mean (SD) 45.1 (16.9)              | 45.9 (14.7)                        | 47.0 (6.4)                              |
| Race n (%)                          | Asian 1 (3.0)                      | 1 (4.3)                            | 4 (2.6)                                |
|                                     | Black or African American 1 (3.0)  | 1 (4.3)                            | 1 (0.6)                                |
|                                     | Hispanic/Latino (of any race) 3 (9.1) | 2 (8.7)                            | 6 (3.9)                                |
|                                     | Multiracial N/A                    | N/A                                | 7 (4.5)                                |
|                                     | Native Hawaiian or Pacific Islander N/A | N/A                              | 1 (0.6)                                |
|                                     | North African or Middle Eastern N/A | 1 (4.3)                            | 1 (0.6)                                |
|                                     | White/Caucasian 28 (84.8)          | 16 (69.6)                          | 134 (86.5)                             |
|                                     | Other N/A                          | N/A                                | 1 (0.6)                                |
|                                     | Missing/No response N/A            | N/A                                | N/A                                    |
| Education n (%)                     | High school diploma (or GED) or less 7 (21.2) | 1 (4.3)                            | 6 (3.9)                                |
|                                     | Some college or certificate program N/A | 5 (21.7)                            | 30 (19.4)                              |
|                                     | College or university degree (2- or 4-year) 19 (57.6) | 13 (56.5)                           | 71 (45.8)                              |
|                                     | Graduate degree 7 (21.2)           | 3 (13.0)                           | 44 (28.4)                              |
|                                     | Missing/No response N/A            | 1 (4.3)                            | N/A                                    |
| Work Status* n (%)                  | Employed (full-time) 14 (42.4)     | 13 (56.5)                          | 6 (3.9)                                |
|                                     | Employed (part-time) 6 (18.2)      | 4 (17.4)                           | 96 (61.9)                              |
|                                     | Homemaker 5 (15.2)                 | 4 (17.4)                           | 19 (12.3)                              |
|                                     | Student 1 (3.0)                    | N/A                                | 17 (11.0)                              |
|                                     | Retired 1 (3.0)                    | N/A                                | 6 (3.9)                                |
|                                     | Unemployed 3 (9.1)                 | N/A                                | 3 (1.9)                                |
|                                     | Other 4 (12.1)                     | 1 (4.3)                            | 4 (2.6)                                |
|                                     | Missing/No response N/A            | 1 (4.3)                            | N/A                                    |
| Annual household income n (%)       | Under $25 000 3 (9.1)              | N/A                                | 1 (0.6)                                |
|                                     | $25 000 to $49 999 3 (9.1)         | 3 (13.0)                           | 15 (9.7)                               |
|                                     | $50 000 to $74 999 8 (24.2)        | 6 (26.1)                           | 24 (15.5)                              |
|                                     | $75 000 to $99 999 4 (12.1)        | 5 (21.7)                           | 13 (8.4)                               |
|                                     | $100 000 and over 14 (42.4)        | 8 (34.8)                           | 98 (63.2)                              |
|                                     | Missing/No response N/A            | 1 (4.3)                            | N/A                                    |
| Treatment(s)* n (%)                 | None 4 (12.1)                      | N/A                                | N/A                                    |
|                                     | Cosmetics 8 (24.2)                 | N/A                                | N/A                                    |
|                                     | Creams/moisturizers 23 (69.7)      | 14 (60.9)                          | 87 (56.1)                              |
|                                     | Cleansers/exfoliating washes 19 (57.6) | 11 (47.8)                         | 65 (41.9)                              |
|                                     | Chemical Peels 9 (27.3)            | 7 (30.4)                           | 24 (15.5)                              |
|                                     | BOTOX® 19 (57.6)                   | 20 (87.0)                          | 148 (95.5)                             |
|                                     | Dysport® 6 (18.2)                  | 6 (26.1)                           | 46 (29.7)                              |
|                                     | Other 2 (6.1)                      | 2 (8.7)                            | 6 (3.9)                                |
| Severity of CRL n (%)               | Not applicable/no 1 (3.0)          | 2 (8.7)                            | 6 (3.9)                                |
|                                     | Mild 13 (39.4)                     | 5 (21.7)                           | 58 (37.4)                              |
|                                     | Moderate 15 (45.5)                 | 14 (60.9)                          | 85 (54.8)                              |
|                                     | Severe 4 (12.1)                    | 2 (8.7)                            | 6 (3.9)                                |
| Severity of FHL n (%)               | Not applicable/no N/A              | 1 (4.3)                            | 7 (4.5)                                |
|                                     | Mild 8 (24.2)                      | 9 (39.1)                           | 49 (31.6)                              |
|                                     | Moderate 20 (60.6)                 | 8 (34.8)                           | 84 (54.2)                              |
|                                     | Severe 5 (15.2)                    | 5 (21.7)                           | 13 (8.4)                               |
|                                     | Missing/No response N/A            | N/A                                | 2 (1.3)                                |
addition, three items of the FLSQ-Baseline and three of the FLSQ-Follow-up were modified to improve clarity or prevent misinterpretations.

Revised questionnaires and updated conceptual framework. The final questionnaires consisted of 11 items (FLSQ-Baseline), and 13 items (FLSQ-Follow-up); the number of domains in each remained unchanged. The 11 items of the Baseline version include the following item concepts: improved facial appearance, treatment onset, treatment duration, receiving a natural look, overall treatment satisfaction, improved self-esteem, feeling older, negatively affected self-esteem, looking tired, feeling unhappy about facial lines, and looking angry.

The 13 items of the Follow-up version include the following concepts: improved facial appearance, treatment onset, treatment duration, receiving a natural look, treatment effect on facial lines, feeling older, negatively affected self-esteem, looking tired, feeling unhappy about facial lines, looking angry, likelihood of continuing treatment, likelihood of recommending treatment to a friend, and extent to which treatment expectations were met. The final list of concepts and corresponding domains included in the questionnaires are shown in Table 5.

Quantitative phase

Study population

A total of 155 subjects (female = 85.8%; mean age = 47.0 ± 6.4 years) participated in the first and second assessments. Subjects' demographic and health information is summarized in Table 4.

---

**Table 4 (continued)**

| Characteristic                              | Concept elicitation total ($N = 33$) | Cognitive interviews total ($N = 23$) | Psychometric validation total ($N = 155$) |
|---------------------------------------------|--------------------------------------|--------------------------------------|-----------------------------------------|
| Severity of GL n (%)                        |                                      |                                      |                                         |
| Not applicable/none                         | 1 (3.0)                              | 2 (8.7)                              | 8 (5.2)                                 |
| Mild                                        | 6 (18.2)                             | 4 (17.4)                             | 48 (31.0)                               |
| Moderate                                    | 19 (57.6)                            | 13 (56.5)                            | 75 (48.4)                               |
| Severe                                      | 7 (21.2)                             | 4 (17.4)                             | 23 (14.8)                               |
| Missing/No response                         | N/A                                  | N/A                                  | 1 (0.6)                                 |

CFL, Crow’s feet lines; FHL, forehead lines; GL, glabellar lines.

*Counts not mutually exclusive.

---

**Figure 2** Most common impacts of facial lines.
Traditional psychometric analyses

Acceptability. Each FLSQ domain fulfilled psychometric criteria for missingness and floor and ceiling effects. A good response rate was demonstrated with a low frequency of missing responses (<10%) across items and subjects. Floor or ceiling effects above the desired criteria were seen in some individual FLSQ items targeting satisfaction and impact. Items in the FLSQ-Baseline version that had floor or ceiling effects were as follows: treatment onset (ceiling effect), treatment duration (ceiling effect), receiving a natural look (ceiling effect), improved self-esteem (ceiling effect), and looking angry (floor effect). Items in the FLSQ-Follow-up version that had floor or ceiling effects were as follows: treatment onset (ceiling effect), treatment duration (ceiling effect), receiving a natural look (ceiling effect), improved self-esteem (ceiling effect), feeling older (ceiling effect), negatively affected self-esteem (ceiling effect), looking tired (ceiling effect), feeling unhappy about facial lines (ceiling effect), and looking angry (floor effect).

Reliability. Cronbach’s α coefficients (>0.70) and interitem correlations (r < 0.80) supported domain reliability. (Note: the Treatment Expectations domain was marginally below the recommended criteria for internal consistency reliability [α = 0.63]) (Table 6).

Validity. The multitrait Pearson’s correlation coefficients (relating items to their respective domains) (0.42–0.75), consistent concurrent correlations (as predicted), and hypothesized group differences (as predicted) provide evidence toward the construct validity of the FLSQ (Table 7). Testing of the single items produced similar validity results.

Further information on data quality, scaling assumptions, and validity are available upon request from the authors.

Discussion

Wrinkling of the skin occurs slowly over time, as repeated contraction of the underlying facial musculature (i.e. repeated facial expressions) can lead to the formation of facial lines or wrinkles. In the upper face, there are three main types of hyperfunctional facial lines: CFL, FHL, and GL. With age, facial lines tend to become static and are visible even when muscles are at rest.1–3 To address concerns with facial wrinkling, individuals may seek cosmetic surgery and minimally invasive nonsurgical procedures to delay the aging process. To better understand the patient’s voice with regard to the impact of these facial lines as well as satisfaction of minimally invasive nonsurgical procedures, a de novo PRO instrument was created.

This study describes the content development and psychometric evaluation of a new PRO questionnaire, the FLSQ, using rigorous qualitative and quantitative methods that comply with the current FDA PRO Guidance. The FLSQ was developed to address the absence of facial line-specific satisfaction PRO questionnaires. Patient input was incorporated at each step in the FLSQ development. The conceptual framework of the FLSQ was developed to comprehensively encompass relevant concepts and domains for facial line treatment satisfaction. Based on patient feedback and item generation, the final conceptual framework contained seven domains across the FLSQ-Baseline and Follow-up: Treatment Expectations and Impact of Facial Lines (Baseline version), and Treatment Satisfaction, Impact of Facial Lines, Continue Treatment, Recommend to Others, and Met Expectations (Follow-up version). The resulting 11-item and 13-item questionnaires (Baseline and Follow-up, respectively) are suitable for use among adults with UFL (CFL, GL, and FHL). The FLSQ has been demonstrated to be content valid after a literature
Facial line satisfaction questionnaire • F Pompilus et al.

Table 6 Psychometric test results

| FLSQ domain       | Internal consistency reliability (Cronbach’s α) | Concurrent validity FLO-11 total score (Spearman’s r) | Concurrent validity TSQM domain (Spearman’s r) |
|-------------------|-----------------------------------------------|------------------------------------------------|-----------------------------------------------|
|                   |                                               | Crow’s lines | Forehead lines | Glabellar lines | Effectiveness | Side effects | Convenience | Global satisfaction |
| Baseline (n = 155) |                                               |              |               |                |              |              |              |                    |
| Treatment Expectations | 0.63                                    | r = –0.207 | r = –0.069 | r = –0.211 | r = 0.228 | r = 0.116 | r = 0.284 | r = 0.352 |
| Impact (Baseline) | 0.82                                       | r = –0.483 | r = 0.573 | r = –0.549 | r = 0.171 | r = –0.005 | r = 0.066 | r = 0.231 |
| Follow-up (n = 112) |                                              |              |               |                |              |              |              |                    |
| Treatment satisfaction | 0.83                                | r = 0.495  | r = 0.283 | r = 0.206 | r = 0.557 | r = 0.312 | r = 0.350 | r = 0.481 |
| Impact (Follow-up) | 0.86                                       | r = –0.294 | r = –0.696 | r = –0.654 | r = –0.145 | r = –0.082 | r = –0.237 | r = –0.020 |
| Met Treatment Expectations | N/A                                         | r = 0.295  | r = –0.100 | r = 0.422  | r = 0.506 | r = 0.058 | r = 0.194 | r = 0.470 |
| Continue Treatment | N/A                                        | r = 0.225  | r = 0.040 | r = 0.120 | r = 0.490  | r = 0.152 | r = 0.192 | r = 0.482 |
| Recommend To others | N/A                                        | r = 0.397  | r = 0.008 | r = 0.160 | r = 0.472  | r = 0.144 | r = 0.345 | r = 0.555 |

Table 7 Clinical validity: known-groups methods at baseline and follow-up

| Severity* | n | Baseline |       | Follow-up |       |
|-----------|---|----------|-------|-----------|-------|
|           |   | Treatment expectation mean (SD) | Impact domain mean (SD) | Treatmet satisfaction mean (SD) | Impact mean (SD)** | Met expectations MEAN (SD)** | Continue treatment mean (SD)†† | Recommend to others mean (SD)‡‡ |
| Crow’s lines at maximum³ |  | n |   | n |       |       |       |       |
| None/Mild | 15 | 71.7 (17.2) | 57.3 (18.7) | 9 | 69.4 (34.2) | 52.8 (24.0) | 66.7 (25.0) | 97.2 (8.3) | 94.4 (11.0) |
| Moderate | 100 | 64.7 (13.9) | 53.9 (21.5) | 72 | 72.3 (17.6) | 46.3 (24.0) | 61.1 (25.5) | 89.6 (17.2) | 82.3 (23.9) |
| Severe    | 17 | 60.6 (11.7) | 49.4 (20.5) | 13 | 73.5 (15.2) | 44.2 (21.6) | 50.0 (20.4) | 86.5 (13.0) | 78.8 (24.7) |
| Forehead lines at maximum brow |  | n |   | n |       |       |       |       |
| None/Mild | 26 | 71.8 (11.8) | 58.7 (23.0) | 18 | 78.6 (20.5) | 49.7 (25.6) | 63.9 (28.7) | 90.3 (19.4) | 93.1 (14.4) |
| Moderate/Severe | 101 | 64.9 (13.6) | 52.6 (20.4) | 74 | 70.1 (19.9) | 44.4 (23.8) | 60.8 (22.3) | 88.5 (16.6) | 82.4 (23.3) |
| Glabellar lines at maximum brow |  | n |   | n |       |       |       |       |
| None/Mild | 52 | 67.2 (13.7) | 56.9 (22.2) | 39 | 73.8 (20.9) | 50.9 (25.7) | 65.4 (26.0) | 91.7 (16.6) | 87.2 (23.6) |
| Moderate | 62 | 66.2 (13.1) | 53.0 (20.6) | 41 | 71.7 (16.6) | 44.2 (22.2) | 57.3 (23.9) | 87.8 (21.7) | 81.1 (24.9) |
| Severe    | 12 | 65.9 (14.9) | 56.3 (22.2) | 10 | 66.0 (22.0) | 41.5 (25.1) | 60.0 (21.1) | 77.5 (24.9) | 77.5 (29.9) |

FLSQ, facial line satisfaction questionnaire.
*As rated on the investigator-rated FWS at baseline.
†Higher scores on the FLSQ domain represent high expectations. FLSQ domain scores vary from 0 to 100.
‡Higher scores on the FLSQ domain represent high negative impact. FLSQ domain scores vary from 0 to 100.
§Using the most severe score from right or left.
¶Higher scores on the FLSQ domain represent high satisfaction with treatment. FLSQ domain scores vary from 0 to 100.
**Higher scores on the FLSQ domain represent treatment expectations being met better than expected. FLSQ domain scores vary from 0 to 100.
††Higher scores on the FLSQ domain represent treatment being continued. FLSQ domain scores vary from 0 to 100.
‡‡Higher scores on the FLSQ domain represent treatment being recommended. FLSQ domain scores vary from 0 to 100.

review, concept elicitation interviews, item generation, and cognitive interviews.

After completion of the qualitative research, the psychometric performance of the FLSQ was tested. Traditional psychometric methods were utilized to provide evidence to support the new PRO questionnaire in accordance with FDA recommendations, as well as to generate data to compare its performance to existing treatment satisfaction questionnaires. In general, the psychometric evaluation demonstrated that the FLSQ mostly satisfies recommended psychometric criteria according to traditional methodologies. The FLSQ domain and single-item scores of interest fulfilled traditional psychometric criteria for acceptability, reliability, and validity. For items in the FLSQ-Follow-up version that showed apparent ceiling effects, the responses at
the highest score range (i.e. the ceiling effect) likely reflect the true distribution rather than the presence of a ceiling effect. For example, subjects were satisfied with the effect of treatment on their facial lines with regard to treatments they had received in the past, identified in the inclusion/exclusion criteria, or were currently receiving. No domain that had multiple items had excessive floor and ceiling responses; all responses were <20% for all scales.

Although one of the domains (Treatment Expectations) had internal consistency reliability estimates below the recommended criteria, it was marginal, thus mostly fulfilling the minimum criteria and indicating adequate reliability.

This study is not without limitations. The FLSQ was developed in a US English-only population. Patient perceptions of facial line treatment satisfaction are not independent of their cultural environment. There are, however, linguistic validation studies that have recently been completed to establish Canadian French/English versions of the FLSQ. Test–retest results were inconclusive in this observational study; however, additional test–retest reliability analyses were performed in an independent clinical study and were proven to meet the standard threshold of an intraclass correlation coefficient of > 0.70. In this same independent study, minimal important difference (MID) development and subsequent testing was conducted for the multidomain that was present in both FLSQ-Baseline and Follow-up (Impact domain). The results from the MID analysis empirically suggest that a 20-point improvement in the FLSQ Impact domain score (on a 0–100 scale) may be a clinically meaningful change.

Continued evaluation of the FLSQ’s psychometric performance in a long-term perspective is recommended.

Conclusion

The FLSQ was created after establishing the need for a PRO questionnaire that measures facial line treatment satisfaction and is in line with currently accepted standards for PRO questionnaire development. The purpose of the present study was to develop and test the psychometric properties of the FLSQ utilizing rigorous qualitative and quantitative methodologies in a prospective, noninterventional study, in accordance with FDA recommendations for the assessment of newly developed PROs. The findings from this study provide empirical evidence that the FLSQ is a reliable and valid PRO questionnaire that can accurately assess the facial line treatment satisfaction in adults with UFL.

Conflict of interest

The research activities presented herein, including the preparation of this manuscript, were funded by Allergan, Inc., and carried out by Adelphi Values.

References

1 Binder WJ. Long-term effects of botulinum toxin type A (Botox) on facial lines: a comparison in identical twins. Arch Facial Plast Surg 2006; 8: 426–31.
2 Coleman KR, Carruthers J. Combination therapy with BOTOX and fillers: the new rejuvenation paradigm. Dermatol Ther 2006; 19: 177–88.
3 Finn CJ, Cox SE, Earl ML. Social implications of hyperfunctional facial lines. Dermatol Surg 2003; 29: 450–5.
4 Sadick NS. The impact of cosmetic interventions on quality of life. Dermatol Online J 2008; 14: 2.
5 Jackson JL, Chamberlin J, Kroenke K. Predictors of patient satisfaction. Soc Sci Med 2001; 52: 609–20.
6 Flynn TC. Botulinum toxin: examining duration of effect in facial aesthetic applications. Am J Clin Dermatol 2010; 11: 183–99.
7 Ching S, Thoma A, McCabe RE et al. Measuring outcomes in aesthetic surgery: a comprehensive review of the literature. Plast Reconstr Surg 2003; 111: 469–80.
8 Cox SE, Finn JC, Stetler L et al. Development of the facial lines treatment satisfaction questionnaire and initial results for botulinum toxin type A-treated patients. Dermatol Surg 2003; 29: 444–9.
9 Imadojemu S, Sarwer DB, Percec I et al. Influence of surgical and minimally invasive facial cosmetic procedures on psychosocial outcomes: a systematic review. JAMA Dermatol 2013; 149: 1325–33.
10 U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville, MD, 2009. Available at: www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed February 4, 2015.
11 Rothman M, Burke L, Erickson P et al. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. Value Health 2009; 12: 1075–83.
12 Glaser BG, Strauss A. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago, IL: Aldine Transaction; 1967.
13 Hammersly M. The Dilemma of Qualitative Method: Herbert Blumer and the Chicago Tradition. London: Routledge; 1990.
14 Lasch KE, Hassan M, Endicott J et al. Development and content validity of a patient reported outcomes measure
to assess symptoms of major depressive disorder. BMC Psychiatry 2012; 12: 34.

15 Charmaz K. Grounded theory. In: JA Smith, R Harre, I. Van Langenhove, eds. Rethinking Methods in Psychology. London: Sage; 1995: pp. 27–49.

16 Glaser B, Strauss AL. The constant comparative method of qualitative analysis. Discovery of Grounded Theory: Strategies for Qualitative Research. New York: Aldine de Gruyter; 1967: pp. 101–16.

17 Kowalski J, Kozma C, Reese PR et al. Initial development of a patient-completed questionnaire to assess outcomes of aesthetic treatment for hyperfunctional facial lines of the upper face [poster]. Presented at: the American Academy of Dermatology Annual Meeting, July 20–24, 2005; Chicago, IL.

18 Carruthers A, Carruthers J. A single-center dose-comparison study of botulinum neurotoxin type a in females with upper facial rhytids: assessing patients’ perception of treatment outcomes. J Drugs Dermatol 2009; 8: 924–9.

19 Atkinson MJ, Kumar R, Cappelleri JC et al. Hierarchical construct validity of the treatment satisfaction questionnaire for medication (TSQM version II) among outpatient pharmacy consumers. Value Health 2005; 8(Suppl. 1): S9–24.

20 Hays RD, Hayashi T. Beyond internal consistency reliability: rationale and user’s guide for Multitrait analysis program on the microcomputer. Behav Res Methods Instrum Comput 1990; 22: 167–75.