Original Investigation

Development and Psychometric Validation of the FACE-Q Skin, Lips, and Facial Rhytids Appearance Scales and Adverse Effects Checklists for Cosmetic Procedures

Anne F. Klassen, DPhil; Stefan J. Cano, PhD; Jonathan A. Schwitzer, MD; Stephen B. Baker, MD, DDS; Alastair Carruthers, MD; Jean Carruthers, MD; Anne Chapas, MD; Andrea L. Pusic, MD, MSc

IMPORTANCE Patient-reported outcomes data are needed to determine the efficacy of cosmetic procedures.

OBJECTIVE To describe the development and psychometric evaluation of 8 appearance scales and 2 adverse effect checklists for use in minimally invasive cosmetic procedures.

DESIGN, SETTING, AND PARTICIPANTS We performed a psychometric study to select the most clinically sensitive items for inclusion in item-reduced scales and to examine reliability and validity with patients. Recruitment of the sample for this study took place from June 6, 2010, through July 28, 2014. Data analysis was performed from December 11, 2014, to December 22, 2015. Pretreatment and posttreatment patients 18 years and older who were consulting for any type of facial aesthetic treatment were studied. Patients were from plastic surgery and dermatology outpatient clinics in the United States and Canada (field-test sample) and a clinical trial of a minimally invasive lip treatment in the United Kingdom and France (clinical trial sample).

MAIN OUTCOMES AND MEASURES The FACE-Q scales that measure appearance of the skin, lips, and facial rhytids (ie, overall, forehead, glabella, lateral periorbital area, lips, and marionette lines), with scores ranging from 0 (lowest) to 100 (highest), and the FACE-Q adverse effects checklists for problems after skin and lip treatment.

RESULTS Of 783 patients recruited, 503 field-test patients (response rate, 90%) and 280 clinical trial participants were studied. The mean (SD) age of the patients was 47.4 (14.0) years in the field-test sample and 47.7 (12.3) years in the clinical trial sample. Most of the patients were female (429 [85.3%] in the field-test sample and 274 [97.9%] in the clinical trial sample). Rasch Measurement Theory analyses led to the refinement of 8 appearance scales with 66 total items. All FACE-Q scale items had ordered thresholds and acceptable item fit. Reliability, measured with the Personal Separation Index (range, 0.88-0.95) and Cronbach α (range, 0.93-0.98), was high. Lower scores for appearance scales that measured the skin ($r = -0.48$, $P < .001$), lips ($r = -0.21$, $P = .001$), and lip rhytids ($r = -0.32$, $P < .001$) correlated with the reporting of more skin- and lip-related adverse effects. Higher scores for the 8 appearance scales correlated (range, 0.70-0.28; $P < .001$) with higher scores on the core 10-item FACE-Q satisfaction with facial appearance scale. In the pretreatment group, older age was significantly correlated with lower scores on 5 of the 6 rhytids scales (exception was forehead rhytids) (range, −0.28 to −0.65; $P = .03$ to $<.001$). Pretreatment patients reported significantly lower scores on 7 of the 8 appearance scales compared with posttreatment patients (exception was skin) ($P < .001$ to .005 on independent sample t tests).

CONCLUSIONS AND RELEVANCE The FACE-Q appearance scales and adverse effects checklists can be used in clinical practice, research, and quality improvement to incorporate cosmetic patients' perspective in outcome assessments.

JAMA Dermatol. 2016;152(4):443-451. doi:10.1001/jamadermatol.2016.0018
Published online March 2, 2016.

Supplemental content at jamadermatology.com

Author Affiliations: Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada (Klassen); Modus Outcomes, Boston, Massachusetts (Cano); Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York (Schwitzer, Pusic); Department of Plastic Surgery, Georgetown University Hospital, Washington, DC (Schwitzer, Baker); Department of Dermatology, University of British Columbia, Vancouver, British Columbia, Canada (A. Carruthers); Department of Ophthalmology, University of British Columbia, Vancouver, British Columbia, Canada (J. Carruthers); Department of Dermatology, New York University School of Medicine, New York, New York (Chapas).

Corresponding Author: Anne F. Klassen, DPhil, McMaster University, 1280 Main St W, Hamilton, ON L8S 4K1, Canada (aklass@mcmaster.ca).

In 2014, a total of 13.9 million minimally invasive cosmetic procedures were performed in the United States, representing an increase of 3% from the year before.[1] To include the patient voice in the assessment of treatment outcomes in the cosmetics industry, patient-reported outcome (PRO) instruments are needed.[2] A review[3] of PRO instruments in 96 736 registered clinical trials between 2007 and 2013 found that 27% used 1 or more, with 17% as a primary or secondary end point. The choice of which PRO instrument to use in a study is a crucial decision. If the wrong instrument is used, it may appear that a new aesthetic product or intervention has little to no benefit.

Engaging patients in the identification of issues that matter to them and using their stories to develop PRO instruments can help to ensure content validity.[4-6] Unfortunately, few such instruments are available for cosmetic treatments. A literature review[7] to identify PRO instruments for cosmetic procedures found 9 of which 3 met international recommendations for how such tools should be developed and validated (ie, BREAST-Q,[8,9] FACE-Q,[10] and Skindex[11]). The review concluded that research dedicated to the evaluation of PRO instruments in cosmetic surgery is urgently required.

The FACE-Q[10,12-16] is a PRO instrument that includes more than 40 scales and checklists designed to measure appearance, adverse effects, health-related quality of life, and experience of health care. These domains form the basis of the FACE-Q conceptual framework. Each domain contains multiple scales and checklists. Because of the large number of scales, validation results are being published as a series of articles, each of which describes clinically relevant groupings. The aim of this article is to describe the set of the FACE-Q scales and checklists that can be used to evaluate minimally invasive cosmetic procedures. Specifically, we describe our psychometric findings for 8 appearance scales designed to evaluate skin, lips, and facial rhytids (overall, forehead, glabella, lateral periorbital area, lips, and marionette lines). We also describe 2 checklists designed to measure adverse effects for skin and lip treatment.

Key Points

Question: Do the FACE-Q scales provide a means to measure appearance of the skin, lips, and facial rhytids (ie, overall, forehead, glabella, lateral periorbital area, lips, and marionette lines)?

Findings: In this study of 783 participants, psychometric analysis supported the reliability and validity of the FACE-Q scales. Adverse effects after specific cosmetic treatments were also identified.

Meaning: The FACE-Q can be used to involve patients in the assessment of treatment outcomes in the cosmetics industry.

Methods

Before study commencement, research ethics approval was obtained at The New School in New York City, New York, and University of British Columbia in Vancouver, British Columbia, Canada. Completion of the FACE-Q questionnaire implied consent.

The FACE-Q was developed by following the US Food and Drug Administration guidance to industry[2,17] and other guidance documents.[18-20] We describe our methods elsewhere.[10,13-16] Briefly, a systematic review,[21] qualitative interviews with 50 patients with facial aesthetics, and input from 26 experts were used to develop the FACE-Q conceptual framework and scales and checklists. The content of each scale was then refined through cognitive interviews with 35 patients. We developed 4 response options in keeping with best practice.[22] Instructions ask respondents to answer in relation to the past week.

The scales for skin and lips measure satisfaction with appearance. The 6 scales that measure appearance of rhytids (overall, forehead, glabella, lateral periorbital area, lips, and marionette lines) and the adverse effects checklists (skin and lips) evaluate how bothered someone is by these concepts. eTable 1 in the Supplement lists the content and response options for the scales and checklists.

For validation purposes, we included 3 additional FACE-Q scales: 10-item satisfaction with facial appearance scale, 10-item psychological function scale, and 8-item social function scale. These scales previously demonstrated reliability, validity, and the ability to detect change.[8,15] Participants were also asked questions so the sample could be characterized by age, sex, and ethnicity.

Study 1: Data Collection

To be included in the study, patients had to be 18 years or older with a pretreatment or posttreatment status for 1 or more of any type of surgical or nonsurgical facial aesthetic treatment. For minimally invasive treatments, returning patients asked to participate, those who had received botulinum toxin treatment more than 4 months ago, and those who had received soft-tissue fillers more than 9 months ago were considered pretreatment participants in our study sample. Participants were recruited from 4 dermatology and 11 plastic surgery offices in the United States and Canada from June 6, 2010, through July 28, 2014. Data analysis was performed from December 11, 2014, to December 22, 2015. For 11 clinics, staff provided a questionnaire booklet to complete in the waiting room at check-in. The remaining clinics invited patients to participate via a postal survey that included a personalized letter from the relevant health care professional alongside a questionnaire booklet with up to 3 mailed reminders. Potential participants were provided a $5 coffee card in appreciation of their participation. Completion of the FACE-Q questionnaire implied consent.

Study 2: Data Collection

An international, randomized, 2-arm, active-controlled study[23] recruited patients 18 years and older for a volume enhancement lip treatment (clinical trial sample). Participants were recruited from 12 sites in the United Kingdom and France. The treatment injection volume was based on clinical experience and lip treatment goals. Vermilion body

and border were the primary treatment sites; additional perioral sites could also be treated. This study was approved by Ethics Committee Address and Chairperson National Research Ethics Service. All participants provided written informed consent. The data were deidentified. More details about the study sample and methods are published elsewhere.[23]

The scales that measured lips and satisfaction with facial appearance were administered on days 0, 30, and 90. The scales that measured lip rhytids and psychological and social function were administered on days 0, 14, 30, and 90. The adverse effects checklist for lips was administered on days 14 and 30. These scales were translated into French by MAPI Research Trust, following their linguistic validation method, which includes 2 separate forward translations by 2 qualified translators, a reconciliation process, and 1 backward translation by a qualified translator.[24]

Statistical Analysis

For the adverse effects checklists, the proportion of responses for each response option was computed. For the appearance scales, Rasch Measurement Theory (RMT)[25,26] was conducted within RUMM2030 statistical software.[27] Rasch Measurement Theory examines the difference between observed and predicted item responses to determine whether data from a sample fit the Rasch model.[28] The results from a range of statistical and graphical tests were examined, with the evidence considered together to make a decision about each scale's overall quality.[28-30] We performed the following:

1. *Threshold for item response options*: We examined the ordering of thresholds, which are the points of crossover between adjacent response categories (eg, between somewhat satisfied and very satisfied) to determine whether successive integer scores increased for the construct measured.
2. *Item fit statistics*: For each scale, we examined 3 indicators of fit to determine whether the scale's items worked together to map out a clinically important construct: (1) log residuals (item-person interaction), (2) χ^2 values (item-trait interaction), and (3) item characteristic curves. The criteria for fit residuals should fall between −2.5 and +2.5. The χ^2 value for each item should be nonsignificant after Bonferroni adjustment.
3. *Dependency*: Residual correlations among items in a scale can artificially inflate reliability. We examined residual correlations among items, which should be below 0.30.[26]
4. *Stability:* Differential item functioning (DIF) measures the degree to which item performance remains stable across subgroups. A χ^2 value significant after Bonferroni adjustment can indicate an item with potential DIF. We examined DIF by age, sex, and country.
5. *Targeting:* Targeting can be examined by inspecting the spread of person (range of the construct reported by the sample) and item (range of the construct measured by the items) locations. Items in a scale should be evenly spread across a reasonable range that matches the range of the construct experienced by the sample.
6. *Person separation index (PSI)*: We examined reliability using the PSI, a statistic that is comparable to the Cronbach α.[31] The PSI measures error associated with the measurement of people in a sample. Higher values indicate greater reliability.

We also computed a Cronbach α for each scale, which provides a measure of how closely related a set of items are as a group.[31] Rasch logit scores for each participant were transformed into scores from 0 (worst) to 100 (best). The scoring algorithm is available from the authors. Pearson correlations to examine associations among scores and 2-tailed independent sample *t* tests used to test for differences among means were used to test the following hypotheses:

1. Higher scores on the appearance scales would correlate with higher scores for satisfaction with facial appearance, psychological function, and social function.
2. Lower scores on the skin scale would correlate with more adverse effects for skin. Similarly, lower scores on the lips and lip rhytids scales would correlate with more adverse effects for lips on the day 14 assessment.
3. Before treatment, older participants would report lower scores on the 6 rhytids scales compared with younger participants.
4. Pretreatment participants would report lower scores on all 8 scales compared with posttreatment patients.

$P < .05$ was considered statistically significant.

Results

Response Rate

A total of 503 of 558 patients invited to participate completed a FACE-Q booklet that contained 1 more of the scales described in this study (response rate, 90%). In addition, 280 individuals participated in the lip enhancement clinical trial, for a total of 783 participants. **Table 1** gives the sample characteristics. When we compared the field-test sample with the clinical trial sample, mean age did not differ ($P = .77$ on 2-tailed independent sample *t* test), but sex did ($P < .001$ on the χ^2 test). Specifically, the clinical trial sample had fewer than expected men (9.1% vs 2.1%).

Adverse Effects

The checklist that measured adverse effects of the skin was completed by 74 participants a mean (SD) of 2.4 (3.6) months after skin treatment (range, immediate to 12 months). The top 3 items endorsed included redness, uneven skin tone, and skin sensitivity (**Table 2**). On day 14 in the lip sample, the most common adverse effects were lips that did not feel smooth, look symetric, or look normal.

RMT Analysis

The RMT analysis supported the reliability and validity of the appearance scales. All 66 items had ordered thresholds, providing evidence that each scale's response options worked as a continuum that increased for the construct

Table 1. Patient Characteristics[a]

Characteristic	Field-test Sample (n = 503)	Clinical Trial Sample (n = 280)
Age, mean (SD) [range], y	47.4 (14.0) [18-80]	47.7 (12.3) [18-76]
Sex[b]		
Female	429 (85.3)	274 (97.9)
Male	46 (9.1)	6 (2.1)
Missing	28 (5.6)	...[c]
Ethnicity		
White	357 (71.0)	...
Other	105 (20.9)	...
Missing	41 (8.2)	...
Country		
United States	347 (69.0)	...
Canada	156 (31.0)	...
United Kingdom	...	96 (34.3)
France	...	184 (65.7)
No. of assessments	536	1082
1	496 (94.4)	...
2	23 (4.6)	...
3	5 (1.0)	...
4	...	280 (100)
Timing of assessment		
Pretreatment	215 (40.1)	280 (25.9)
Posttreatment	321 (59.9)	802 (74.1)
Procedure		
Minimally invasive	397 (74.1)	280 (100.0)
Surgical	127 (23.7)	...
Surgical and minimally invasive	12 (2.2)	...
Type of minimally invasive procedure		
Botulinum toxin	181 (33.8)	...
Filler	123 (23.0)	280 (100.0)
Skin treatment	124 (23.1)	...
Other	4 (0.1)	...
Type of surgical procedure		
Antiaging: face-lift, blepharoplasty, brow-lift, neck lift	96 (17.9)	...
Other surgery: rhinoplasty, jaw, chin	43 (8.0)	...
Scales or checklists completed		
Skin	130 (24.3)	...
Lips	...	809 (74.8)
Facial rhytids overall	163 (30.4)	...
Forehead rhytids	188 (35.1)	...
Glabella rhytids	173 (32.3)	...
Lateral periorbital area rhytids	210 (39.2)	...
Lip rhytids	...	1076 (99.5)
Marionette lines	280 (52.2)	...
Skin adverse effects	74 (13.8)	...
Lip adverse effects	...	534 (49.4)
Facial appearance	523 (97.6)	800 (73.9)
Psychological function	177 (33.0)	1076 (99.5)
Social function	178 (33.2)	1075 (99.4)

[a] Data are presented as number (percentage) of study participants unless otherwise indicated.

[b] Clinical trial sample had fewer men than the field-test sample ($P < .001$ on χ^2 test).

[c] Ellipses indicate data not available.

Table 2. Adverse Effect Reports by the 74 Participants Who Completed the Skin Adverse Effects Checklist and the 280 Participants Who Completed the Lips Adverse Effects Checklist 14 Days After a Minimally Invasive Treatment

	No. (%) of Patients			
Adverse Effect	Not at All	A Little	Moderately	Extremely
Skin adverse effects				
Redness	20 (27.0)	34 (45.9)	13 (17.6)	7 (9.5)
Uneven skin tone (darker and lighter areas)	22 (29.7)	25 (33.8)	19 (25.7)	8 (10.8)
Skin sensitivity (eg, to sunlight, skin products)	26 (35.6)	22 (30.1)	17 (23.3)	8 (11.0)
Parts of face looking blotchy	27 (37.0)	23 (31.5)	19 (26.0)	4 (5.5)
Parts of face not looking smooth	28 (37.8)	30 (40.5)	14 (18.9)	2 (2.7)
Parts of face not feeling smooth to the touch	35 (47.3)	24 (32.4)	14 (18.9)	1 (1.4)
Tightness	39 (53.4)	24 (32.9)	8 (11.0)	2 (2.7)
Itching	47 (65.3)	17 (23.6)	6 (8.3)	2 (2.8)
Parts of face looking scarred	47 (63.5)	15 (20.3)	9 (12.2)	3 (4.1)
Burning	54 (74.0)	12 (16.4)	6 (8.2)	1 (1.4)
Lip adverse effects[a]				
Lips not feeling smooth (lumps, bumps)	180 (66.7)	67 (24.8)	16 (5.9)	7 (2.6)
Lips looking asymmetric (uneven)	201 (74.4)	59 (21.9)	9 (3.3)	1 (0.4)
Lips not looking smooth (lumps, bumps)	205 (76.2)	47 (17.5)	13 (4.8)	4 (1.5)
Swelling	223 (82.6)	31 (11.5)	15 (5.6)	1 (0.4)
Lips feeling unnatural	225 (83.3)	34 (12.6)	9 (3.3)	2 (0.7)
Skin adverse effects				
Numbness	245 (90.7)	19 (7.0)	5 (1.9)	1 (0.4)
Difficulty moving lips (eg, laughing, smiling)	249 (92.2)	15 (5.6)	5 (1.9)	1 (0.4)
Lips feeling too big	251 (93.0)	17 (6.3)	1 (0.4)	1 (0.4)

[a] Lip adverse effects checklist data from the clinical trial sample are from the day 14 assessment.

measured. Fit residuals were within the −2.5 to +2.5 recommended range for 50 of the 66 items (eTable 2 in the Supplement), and 66 of the 66 items were not significant in terms of the adjusted χ^2 *P* values, providing evidence that the items fit the expectations of the Rasch model for each scale. The 16 items with fit outside the recommended range were retained because of their clinical importance. The item residuals were above 0.30 (range, 0.35-0.59) for 6 pairs of items within 5 scales. Subtests performed on the pairs of items revealed marginal effect on scale reliability (0 to 0.01 difference in PSI value). For the scale that measured satisfaction with lips, DIF was detected for age and/or country on 5 items. When these items were split on the variable with DIF and the new person locations for the scale were correlated with the original person locations, the DIF had a negligible effect (Pearson correlates were 0.99).

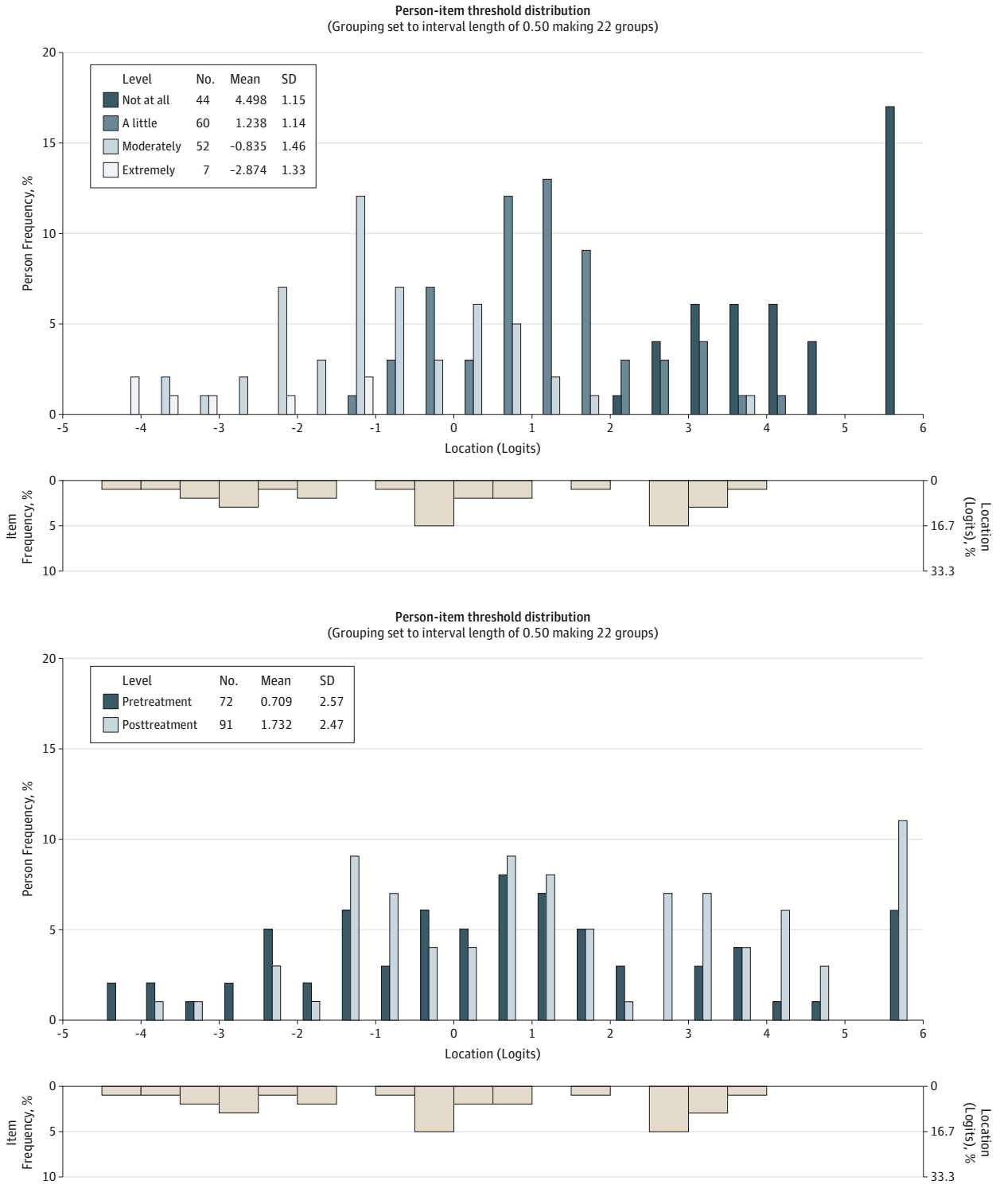
Figure 1 shows the person-item threshold distribution for the scale that measured facial rhytids overall as an example of targeting. The x-axis represents the construct (facial rhytids appearance), with higher scores (less bothered) increasing to the right. The y-axis represents the frequency of person measure locations (top histogram) and item locations (bottom histogram). The sample was divided into 4 groups based on their answer (not at all, a little, moderately, or extremely) to a stand-alone item that asked how much participants were bothered by, "How the lines on your face look overall?" and into pretreatment and posttreatment groups. These examples provide evidence that most of the sample lay inside the range in which the scale provided measurement.

The *P* values for fit to the Rasch model were not significant for 7 of the 8 scales, which indicates that the data satisfied the requirements of the Rasch model. The *P* value for the scale that measured lip rhytids was significant (*P* = .02). The 8 scales evidenced high reliability. The PSI and Cronbach α values were as follows: skin, 0.93 and 0.93; lips, 0.95 and 0.97; rhytids overall, 0.93 and 0.95); forehead rhytids, 0.88 and 0.95; glabella rhytids, 0.91 and 0.96; lateral periorbital area rhytids, 0.92 and 0.96; lip rhytids, 0.93 and 0.97; and marionette lines, 0.92 and 0.98, respectively.

Construct Validity

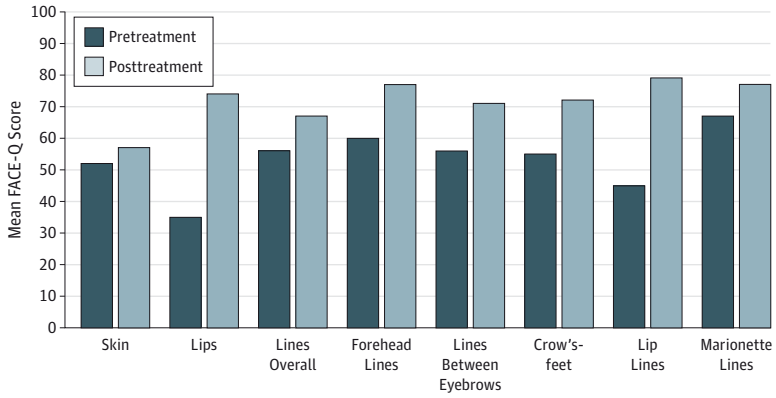
Pearson correlations between the 8 scales and satisfaction with facial appearance scores were significant (*P* < .001)

Figure 1. Person-Item Threshold Distribution for Rhytids Overall by Response to the Question That Asked How Much Participants Were Bothered by, "How the Lines on Your Face Look Overall?" and by Pretreatment and Posttreatment Status



and ranged from 0.70 (skin) to 0.28 (glabella rhytids). Correlations between the 8 scales and psychological function were significant (P = .03 to <.001) for 7 of the 8 scales (exception was glabella rhytids) and ranged from 0.51 (lateral periorbital area rhytids) to 0.32 (rhytids overall). Correlations between the 8 scales and social function

Figure 2. Mean Scores for FACE-Q Scales Comparing Pretreatment With Posttreatment Data



Mean FACE-Q Score	Skin	Lips	Lines Overall	Forehead Lines	Lines Between Eyebrows	Crow's-feet	Lip Lines	Marionette Lines
Pretreatment	52	35	56	60	56	55	45	67
Posttreatment	57	74	67	77	71	72	79	77

Lip and lip line posttreatment data from the clinical trial sample are from the day 14 assessment.

were significant for 3 scales, including lateral periorbital area rhytids (r = 0.40, P < .002), lips (r = 0.35, P < .001), and lip rhytids (r = 0.28, P < .001).

More skin-related adverse effects correlated with lower scores on the skin scale (r = −0.48, P < .001). More lip-related adverse effects correlated with lower scores on the lip (r = −0.21, P = .001) and lip rhytids (r = −0.32, P < .001) scales.

In the pretreatment group, correlations between older age and lower scores for the rhytids scales were significant for 5 of the 6 scales (exception was forehead rhytids): rhytids overall (r = −0.41, P < .001), glabella rhytids (r = −0.28, P = .03), lateral periorbital area rhytids (r = −0.35, P = .001), lip rhytids (r = −0.52, P < .001), and marionette lines (r = −0.65, P < .001). In the posttreatment group, age was not significantly correlated with scores from 5 of the 6 rhytids scales (exception was lip rhytids: r = −0.32, P < .001).

Figure 2 shows the mean scores for the 8 appearance scales for pretreatment and posttreatment data. Pretreatment patients reported significantly lower scores on 7 of the 8 scales (exception was the skin scale) compared with posttreatment patients (P <.001-.005 on 2-tailed independent sample t tests).

Discussion

Increasing acceptance of facial cosmetic treatments has led to an industry that continues to expand. Research is urgently needed to ensure that new treatments are safe and effective. The FACE-Q is a rigorously developed PRO instrument that can be used by academics and other health care professionals to collect evidence-based outcome data from patients with facial aesthetics.

To date, the FACE-Q is currently the only PRO instrument that includes scales that measure facial appearance. Some FACE-Q appearance scales ask about satisfaction with appearance, and other scales, for negative concepts such as facial rhytids, ask about being bothered by appearance. Other PRO instruments used in facial aesthetics research measure appearance-related psychosocial distress rather than appearance per se. For example, the rigorously developed 61-item Skindex[14] measures negative affect, self-esteem, anxiety, physical discomfort, physical limitations, self-consciousness, and intimacy. A PRO instrument that measures psychosocial issues would not be the best choice for measuring change in appearance.

The psychometric analyses in this article provided evidence of the reliability and validity of the FACE-Q scales. In addition, and fundamentally, our use of RMT methods to develop the FACE-Q has certain advantages. The RMT methods differ from traditional psychometric methods (based on classic test theory) because their focus is on the association between a person's measurement and the probability of responding to an item, rather than the association between a person's measurement and the observed scale total score.[28] Advantages of using RMT to develop PRO instruments include the following: (1) RMT provides measurements of people that are independent of the sampling distribution of the items used and locates items in a scale independent of the sampling distribution of the people in whom they are developed, (2) RMT improves the potential to diagnose item-level psychometric issues, and (3) RMT allows for a more accurate picture of individual person measurements.[28] These assets, together with the extensive qualitative work performed to create the FACE-Q, are what set the FACE-Q apart from other PRO instruments in the same clinical area.

This study has previously described limitations.[10,13-16] First, the sample was heterogeneous (eg, varied by age, sex, and timing of assessment), which limits the outcome findings we can report. Second, our sample and that of the clinical trial had many more women than men, which

reflects the makeup of patients with cosmetic issues. Third, there could have been bias introduced at the clinic level by office staff who recruited their patients for us. Fourth, few field-test participants completed the FACE-Q before and after treatment. Responsiveness research is needed to document the benefits of treatment for specific facial treatments.

Conclusions

Evidence-based information about patient outcomes for facial aesthetic treatments is needed. The FACE-Q provides the research community and physicians with a PRO instrument they can use to include patients in the assessment of outcomes.

ARTICLE INFORMATION

Accepted for Publication: December 23, 2015.

Published Online: March 2, 2016. doi:10.1001/jamadermatol.2016.0018.

Author Contributions: Drs Pusic and Klassen 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.
Study concept and design: Klassen, Cano, Pusic.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Klassen, Cano, Pusic.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Klassen, Cano, Schwitzer.
Obtained funding: Klassen, Cano, Pusic.
Administrative, technical, or material support: Klassen, Schwitzer, Baker, A. Carruthers, J. Carruthers, Chapas.
Study supervision: Schwitzer, Baker, Pusic.

Conflict of Interest Disclosures: The FACE-Q is owned by Memorial Sloan-Kettering Cancer Center. Drs Cano, Klassen, and Pusic reported being 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. Dr Cano reported being a cofounder of Modus Outcomes, an outcomes research and consulting firm that provides services to pharmaceutical, medical device, and biotechnology companies. Drs A. Carruthers and J. Carruthers reported being consultants and investigators for Allergan, Merz, Kythera, and Alphaeon. No other disclosures were reported.

Funding/Support: This study was supported by a grant from the Plastic Surgery Foundation (Dr Pusic).

Role of the Funder/Sponsor: The funding source 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 the decision to submit the manuscript for publication.

Previous Presentation: Preliminary results of this study were reported the American Society of Plastic Surgeons Annual Meeting; October 12, 2014; Chicago, Illinois.

Additional Contributions: The following physicians recruited their patients into the FACE-Q field-test sample: D. Berson, MD, James C. Grotting, MD, J. M. Kenkel, MD, F. Nahai, MD, Rod J. Rohrich, MD, A. Rossi, MD, Jonathan M. Sykes, MD, Nancy Van Laeken, MD, L. Young, MD, and J. Rivers, MD. Diane Murphy, MPH, at Allergan Medical provided the FACE-Q data from the clinical trial.

REFERENCES

1. American Society of Plastic Surgeons. 2014 Plastic Surgery Statistics Report. http://www.plasticsurgery.org/Documents/news-resources/statistics/2014-statistics/cosmetic-procedure-trends-2014.pdf. Accessed August 4, 2015.

2. US Food and Drug Administration. The FDA's drug review process: ensuring drugs are safe and effective. http://www.fda.gov/Drugs/ResourcesForYou/Consumers/ucm143534.htm. Accessed August 4, 2015.

3. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: evidence from ClinicalTrials.gov (2007-2013). *Contemp Clin Trials*. 2015;43:1-9.

4. US Food and Drug Administration. Clinical Outcome Assessment Qualification Program. http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm. Accessed August 4, 2015.

5. 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.

6. 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.

7. 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 April 1, 2014.

8. 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.

9. 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.

10. Pusic A, Klassen AF, Scott AM, Cano SJ. Development and psychometric evaluation of the FACE-Q Satisfaction with Appearance Scale: a new PRO instrument for facial aesthetics patients. *Clin Plast Surg*. 2013;40(2):249-260.

11. Chren MM, Lasek RJ, Quinn LM, Mostow EN, Zyzanski SJ. Skindex, a quality-of-life measure for patients with skin disease: reliability, validity, and responsiveness. *J Invest Dermatol*. 1996;107(5):707-713.

12. Klassen AF, Cano SJ, Scott A, Snell L, Pusic AL. Measuring patient-reported outcomes in facial aesthetic patients: development of the FACE-Q. *Facial Plast Surg*. 2010;26(4):303-309.

13. 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.

14. 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.

15. 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.

16. Klassen AF, Cano SJ, East C, et al. Development and psychometric evaluation of FACE-Q scales for rhinoplasty patients. *JAMA Facial Plast Surg*. 2016;18(1):27-35.

17. International Society for Pharmacoeconomics and Outcomes Research. 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 June 10, 2015.

18. 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.

19. Lasch KE, Marquis P, Vigneux M, et al. PRO development: rigorous qualitative research as the crucial foundation. *Qual Life Res*. 2010;19(8):1087-1096.

20. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002;11(3):193-205.

21. 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.

22. 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.

23. Raspaldo H, Chantrey J, Belhaouari L, Saleh R, Murphy DK. Juvéderm volbella with lidocaine for lip and perioral enhancement: a prospective, randomized, controlled trial. *Plast Reconstr Surg Glob Open*. 2015;3(3):e321.

24. Acquadro C, Conway K, Girourdet C, Mear I. *Linguistic Validation Manual for Patient-Reported Outcomes (PRO) Instruments*. Lyon, France: MAPI Research Trust; 2004.

25. Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? *Med Care*. 2004;42(1)(suppl):I-7-I-16.

26. Wright B, Masters G. *Rating Scale Analysis: Rasch Measurement*. Chicago, IL: Mesa Press; 1982.

27. Andrich D, Sheridan B. *RUMM2030* [software program]. Perth, Australia: RUMM Laboratory; 1997-2011.

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

29. Andrich D. *Rasch Models for Measurement*. Newbury Park, CA: Sage; 1988.

30. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmark: Danish Institute for Education Research; 1960.

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

NOTABLE NOTES

The Emperor's Itch

Mindy X. Wang, BS; Samantha Hsieh, BS; Eric L. Maranda, BS; Victoria Lim, BS; Joaquin Jimenez, MD

Standing with his right hand tucked inside his waistcoat, Napoleon I of France looks off in deep thought, in Jacques-Louis David's 1812 painting *Napoleon in His Study*. Though this stance appears regal, there are speculations that this iconic pose was merely a facade for his scratching convenience. At a time when scabies ran rampant, it is no surprise that some believe that Napoleon probably had "the itch."[1]

We now know "the itch" is attributed to scabies, which came from the Latin term *scabere*—"to scratch." This condition manifests as pruritic rashes in the human host following colonization of the mite *Sarcoptes scabiei* in the skin as a result of skin-to-skin contact with an infected host.[2] This infestation is classically seen as papules or burrows, which are thread-like tunnels under the skin filled with egg cases and mite fecal pellets. When these burrows are not concealed by papular or vesicular lesions, the parasite itself can be seen at the end of the tunnel as a black dot.[2]

Giovan Cosimo Bonomo first described the mite in 1687. However, the causal relationship between the mite and the itching symptoms was not elucidated until 1834 when Simon François Renucci, a Corsican medical student, used a needle probe to remove a mite from a young female patient with "the itch." The key component of his procedure was based on the needle technique of Corsican market women: excavating near the vesicle's center.[2] Our understanding of the disease was furthered by entomologist Kenneth Mellanby, who noted the small mite burden in the scabietic soldiers he studied and established the person-to-person body contact mode of disease transmission.[2] Today, scabies continues to be a problem in impoverished communities and in sexually active, immunosuppressed, or elderly individuals.[2]

But was Napoleon's pose truly attributed to scabies? Reuben Friedman, an American physician, debunked this theory in 1940.[1] He presented some evidence: Napoleon took many sulfur baths, and Empress Josephine was not known to have scabies. What he actually suffered from, Friedman argued, was dermatitis herpetiformis,[1] which is now clearly linked with celiac disease. Blistering rashes accompanied by intense itching and burning are its hallmarks.[3]

Was Napoleon's stance a consequence of his urge to scratch? No one knows for sure. Whether it was scabies or dermatitis herpetiformis, the Emperor's affliction is a reality for many patients today. Both conditions can be masked by common features of papules, vesicles, and erosions, prompting misdiagnoses.[2,3] Fortunately, thanks to histological examination, the Emperor's itch is now an unraveled mystery and treatable affliction.

Author Affiliations: Department of Dermatology and Cutaneous Surgery, University of Miami Miller School of Medicine, Miami, Florida.

Corresponding Author: Eric L. Maranda, BS, Department of Dermatology and Cutaneous Surgery, University of Miami Miller School of Medicine, 1475 NW 12th Ave, Miami, FL 33136 (emaranda@med.miami.edu).

1. Friedman R. *The Emperor's Itch: The Legend Concerning Napoleon's Affliction With Scabies*. New York, NY: Froben Press; 1940.

2. Currier RW, Walton SF, Currie BJ. Scabies in animals and humans: history, evolutionary perspectives, and modern clinical management. *Ann N Y Acad Sci*. 2011;1230:E50-E60.

3. Antiga E, Caproni M. The diagnosis and treatment of dermatitis herpetiformis. *Clin Cosmet Investig Dermatol*. 2015;8:257-265.