OPEN FORUM

Cosmetic Medicine

Validating a Series of Photonumeric Rating Scales for Use in Facial Aesthetics Using Statistical Analysis of Intra- and Inter-rater Reliability

Z. Paul Lorenc, MD, FACS; Derek Jones, MD; Jeongyun Kim, MS; Hee Min Gwak, MS; Samixa Batham, MBBS; and Ginny Vachon, PhD

Aesthetic Surgery Journal Open Forum 2021, 1–8
© 2021 The Aesthetic Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
https://doi.org/10.1093/asjof/ojab039
www.asjopenforum.com

OXFORD UNIVERSITY PRESS

Abstract

Background: Growing demand for minimally invasive aesthetic procedures to correct age-related facial changes and optimize facial proportions has been met with innovation but has created an unmet need for objective assessment tools to evaluate results empirically.

Objectives: The purpose of this study is to establish the intra- and inter-rater reliability of ordinal, photonumeric, 4-, or 5-point rating scales for clinical use to assess facial aesthetics.

Methods: Board-certified plastic surgeons and dermatologists (3 raters) performed live validation of jawline contour, temple volume, chin retrusion, nasolabial folds, vertical perioral lip lines, midface volume loss, lip fullness, and crow's feet dynamic- and at rest-rating scales over 2 rounds, 2 weeks apart. Subjects selected for live validation represented the range of scores and included 54 to 83 subjects for each scale. Test-retest reliability was quantitated through intra- and inter-rater reliability, determined from the mean weighted kappa and round 2 intraclass correlation coefficients, respectively. The clinical significance of a 1-grade difference was assessed through rater comparison of 31 pairs of side-by-side photographs of subjects with the same grade or a different grade on the developed scales.

Results: The study demonstrated substantial to near-perfect intra- and inter-rater reliability of all scales when utilized by trained raters to assess a diverse group of live subjects. Furthermore, the clinical significance of a 1-point difference on all the developed scales was established.

Conclusions: The high test-retest reliability and intuitive layout of these scales provide an objective approach with standardized ratings for clinical assessment of various facial features.

Editorial Decision date: October 14, 2021, online publish-ahead-of-print October 21, 2021.

Demand for minimally invasive aesthetic procedures that aim to correct age-related changes and optimize facial proportions, such as dermal fillers and botulinum toxin, has increased dramatically over the last 2 decades.[1,2] To meet this demand, companies within the aesthetic space have sought to develop innovative technologies to assist in facial rejuvenation, and the number of devices and treatments continues to grow. With this growth comes a need for objective assessment of treatment results and, therefore, development of tools for evidence-based evaluation

Dr Lorenc is a plastic surgeon in private practice in New York, NY, USA. Dr Jones is a dermatologist in private practice in Beverly Hills, CA, USA. Ms Kim and Ms Gwak are data management/STAT leads and Dr Batham is a senior manager in the Medical Services Department, LG Chem, Ltd., Life Sciences, Seoul, Korea. Dr Vachon is a medical writer at Principal Medvantage Writing, LLC in Atlanta, GA, USA.

Corresponding Author:
Dr Z. Paul Lorenc, 983 Park Avenue, New York, NY 10028, USA.
E-mail:lorenc@lorenc.com

of new techniques for a range of anatomical areas.[3] Without empirical measurements of success, judgments of effectiveness rely on subjective perceptions of the physician and patient.[4] Rating scales must be designed to capture increasingly pronounced stages of aging pathology, which can be difficult given the wide array of facial structures and manifestations of aging. While aging causes predictable changes in soft tissue, fat, and bone, these changes depend on variations in both intrinsic and extrinsic variables, such as genetics, ethnicity, and sun exposure.[2,5] To date, few rating scales have been developed and validated to measure each area of the face objectively and are somewhat lacking for jawline contour and temple hollowing.[5-11]

When developing scales, it is important to take into consideration the attributes of specific anatomical areas that drive perceptions of aging and/or youthfulness and attractiveness. For example, in the midface, early signs of aging, such as changes in the upper cheek contour due to bone resorption and/or decreasing fat pad volume, can lead to a deflated or tired appearance.[12] Lower face aging can become evident in the highly animated perioral zone due to an imbalance between tissue firmness and activity of the mimetic muscles.[12] Additionally, nasolabial folds, pre-jowl sulci, and jowls become more pronounced, and skin and soft tissue laxity can compound the descent of facial features arising from atrophy and the age-related recession of the jaw, chin, and cheeks.[2,12]

The ideal jawline is free of sulci or excess/sagging skin, well-defined, and smooth along the angle of the mandible until the chin, which balances the lower face—accordingly, adequate chin projection, straightness, and size influence facial beauty.[2,6,12] Lip shape and fullness also contribute to an aesthetically youthful appearance, and age-related volume loss results in flattening, shrinking, and fading of the lips and can worsen the appearance of perioral lines.[12,13] On the upper face, the ideal forehead has a gentle convex curve with flat temples free of any depressions and hollowing.[12] Age-related soft tissue volume loss and bone remodeling lead to protrusion of the upper forehead, flattening of the lower forehead and eyebrows, and temporal hollowing.[12] A deficit in any one area of the face can have an impact on global aesthetic, whether through indirect effects (eg, loss of midface volume contributing to laxity in the lower face) or through disturbance of symmetry or balance. The interconnected nature of facial volume, boney features, muscle activity, and skin quality in the aging face makes the careful diagnosis and treatment of aesthetic deficits central to the art of helping patients maintain a refreshed and vibrant appearance at any age.

To be useful for clinical assessment, rating scales should have high test-retest reliability, which can be quantitated through intra- and inter-rater reliability. Intra-rater reliability can be evaluated by Cohen's weighted kappa coefficient, which is a number between 0 and 1, where 1 corresponds to exact reproducibility and 0 corresponds to a random distribution of repeat score.[14] Inter-rater reliability can be assessed by the intraclass correlation coefficient (ICC), which is also a number between 0 and 1 with higher values reflecting increased reliability of the method between users.[15] The purpose of this study is to validate several ordinal, photonumeric rating scales for clinical use to assess jawline contour, temple volume, and chin retrusion, in addition to nasolabial folds, vertical perioral lip lines, midface volume, lip fullness, and crow's feet, at rest and dynamic.

METHODS

Scale Development

Several photonumeric rating scales were developed by the authors with LG Chem Ltd. (Seoul, South Korea) to validate for clinical use between January 2019 and February 2021. Volume rating scales were developed as 4- or 5-point ordinal scales for jawline contour, temple volume, and chin retrusion (Table 1 and Appendix A), in addition to nasolabial folds, vertical perioral lip lines, midface volume, and lip fullness (Supplemental Table 1, Appendix B). Crow's feet rating scales (CFRS) were developed as at rest and dynamic 5-point ordinal scales (Supplemental Table 1, Appendix B). Initial image collection was conducted by photographers from Canfield Scientific, Inc. (Parsippany-Troy Hills, NJ). Full facial 2-dimensional and 3-dimensional images were collected for participants without make-up to generate a repository for scale creation—181 participants for Jawline Contour Rating Scale (JCRS), Temple Volume Rating Scale (TVRS), Chin Retrusion Rating Scale (CRRS), and Vertical Perioral Lip Lines Rating Scale (VPLRS); 142 participants for the Lip Fullness Rating Scale (LFRS), Midface Volume Loss Rating Scale (MFVLRS), and Nasolabial Folds Rating Scale (NLFRS); 200 participants for Crow's Feet Rating Scale-at Rest and Dynamic. Participants were at least 18 years old and represented both sexes and Fitzpatrick skin types 1 through VI. The purpose, procedure, risks, benefits, and alternatives to participation were communicated to each participant before obtaining written consent for photograph use. Captured images were reviewed for technical adequacy and cropped to ensure primary focus on the fully visible area of interest. The scale developers independently evaluated the images for each scale indication to establish differences between grades and create scale descriptors. Two subjects per grade were chosen as actual patient image examples representing diversity in gender and Fitzpatrick skin type. For each scale, base images demonstrating grade 2 moderate were selected by scale developers and morphed by a Canfield graphics technician to match the description provided for each numeric grade. The morphed images were reviewed and confirmed by the scale developers. Final scales contained the scale

Table 1. Volume Rating Scales Developed as 4- or 5-Point Ordinal Scales for Jawline Contour, Temple Volume, and Chin Retrusion

Scale	Grade	Term	Descriptor
JCRS	0	None	Smooth uninterrupted jawline contour
	1	Mild	Mildly visible anterior and/or posterior jowl sulcus
	2	Moderate	Moderate volume loss of anterior and/or posterior jowl sulcus
	3	Severe	Severe volume loss of anterior and/or posterior jowl sulcus
	4	Extreme	Extreme volume loss of anterior and/or posterior jowl sulcus with pronounced jowl laxity
TVRS	0	Convex	Temple with convex contour
	1	Flat	Temple with flat contour
	2	Moderate Concavity	Temple with moderate concavity
	3	Severe Concavity	Temple with severe concavity and visible bony landmarks
	4	Extreme Concavity	Temple with extreme concavity and pronounced bony landmarks
CRRS	0	None	No chin retrusion
	1	Mild	Minimal chin retrusion
	2	Moderate	Moderate chin retrusion
	3	Severe	Severe chin retrusion

CRRS, Chin Retrusion Rating Scale; JCRS, Jawline Contour Rating Scale; TVRS, Temple Volume Rating Scale.

descriptors, morphed images, and actual subject images. This study was not IRB approved because there is no intervention. Photographs were taken and each person photographed filled out a release form.

Scale Validation

Board-certified plastic surgeons and dermatologists were trained as raters by the scale developer through a web training program. Following training completion, raters completed scale validation for jawline contour, temple hollowing, chin retrusion, and other scales by performing live validation over 2 rounds, 2 weeks apart. Subjects were selected to represent the range of volume scores for each scale, including 68 subjects to constitute all jawline contour scores, 65 subjects with varied temple volume scores, and 83 subjects corresponding to the different severities of chin retrusion. Each scale validator had a separate rating station with a displayed photonumeric scale to reference during participant evaluation. Raters separately assigned an integer rating of 0 to 4 to the left- and right-hand sides of each patient, assessing jawlines for the JCRS and temples for the TVRS. To validate the CRRS, raters assigned an integer rating of 0 to 3 to each patient for chin retrusion. Raters assessed the same subjects during the second round but in a different random sequence. Scales for nasolabial folds, vertical perioral lip lines, midface volume, and lip fullness were validated in the same way as jawline and temple hollow (Supplemental Table 2). Scale validation for crow's feet was performed similarly, except that its severity was assessed at dynamic expression and at rest, with separate ratings for the left and right periorbital areas (Supplemental Table 2). Following the second round, the data were compiled and statistically analyzed.

To demonstrate the clinical significance of a 1-grade difference, the raters compared side-by-side photographs of subjects with the same grade or a different grade on the developed scales (1-3 grade difference for the CRRS and 1-4 grade difference for other scales). During the session, raters were presented with 31 photograph pairs and judged whether the individual photographs had a clinically significant difference (Different/Not Different). After the session, all reviewed photographs were randomly sequenced using a standardized randomization program (Canfield Scientific), and the raters assigned each one a score based on the corresponding facial scale to assess observed clinically significant differences in the 2 photographs. At the conclusion of the review, the data were compiled and statistically analyzed.

Data Analysis and Statistical Methods

Intra-rater reliability between rounds 1 and 2 was evaluated for each rater, the median of all raters, and all raters combined by determining the percentage of agreement (exact and ≥1-grade difference), weighted kappa statistic

with 95% CI, and ICC with 95% CI. Inter-rater reliability for each pair of raters and each rater against the median score of all 3 raters was determined using the same 3 calculations. Weighted kappa statistics were calculated using the weights proposed by Fleiss and Cohen, whereas ICCs were calculated using the ICC method of Shrout and Fleiss. All statistical analysis was performed using SAS Version 9.4.[14,16]

Overall intra-rater reliability was calculated from the mean and 95% CI of mean of weighted kappa between rounds 1 and 2 for all raters. Conversely, overall inter-rater reliability was determined for rounds 1 and 2 separately, using rating scores from each round as the dependent variable and the rater and subject identification numbers as independent variables. Criteria for satisfactory intra- and inter-rater reliability were determined from the mean weighted kappa for intra-rater and the ICC values for inter-rater. Intra-rater mean weighted kappa and inter-rater ICC values > 0 and ≤ 0.2 indicate slight agreement, >0.2 and ≤0.4 indicate fair agreement, >0.4 and ≤0.6 indicate moderate agreement, >0.6 and ≤0.8 indicate substantial agreement, and >0.8 and ≤1.0 indicate almost perfect agreement.[17]

To determine whether the difference in the severity of each assessed feature is clinically significant using the corresponding scale, the absolute difference in scores between each paired photograph was calculated based on the actual ratings from the 3 independent raters. The differences were summarized using descriptive statistics for the photograph pairs deemed as clinically different and not clinically different according to the rater's assessment. The mean, standard deviation (SD), and 95% CI of the mean were reported.

RESULTS

Subject demographics for the live validation population for each scale are summarized in Table 2, Supplemental Table 3. Subjects encompassed both sexes and a wide range of age, race, ethnicities, and Fitzpatrick skin types, with the majority being white, non-Hispanic females. Self-reported Fitzpatrick skin types II through IV were most prevalent in the evaluated population. Subjects' ages ranged from 18 to 88 to represent most of the patient population, with the mean age of 47 to 50 representing patients frequently seen in practice.

The intra-rater reliability for live validation of the scales is characterized by high weighted kappa scores, denoting substantial to almost perfect agreement within the 3 raters for each scale (Table 3, Supplemental Table 4). Moreover, the mean weighted kappa values indicate substantial agreement for live validation of the jawline contour (left jawline: κ, 0.776; 95% CI, 0.561-0.991 and right jawline: κ, 0.775; 95% CI, 0.582-0.968) and chin retrusion scales (κ, 0.756; 95% CI, 0.522-0.991), and almost perfect agreement for validation of the temple volume scale (left temple hollowing: κ, 0.886; 95% CI, 0.813-0.959 and right temple hollowing: κ, 0.884; 95% CI, 0.831-0.937).

Table 2. Summary of Demographics

	JCRS (N = 68)	TVRS (N = 65)	CRRS (N = 83)
Age (years)			
Mean (SD)	49.5 (15.8)	49.0 (15.2)	50.1 (14.8)
Median	49.0	49.0	53.0
Minimum	18	21	22
Maximum	88	88	87
Sex, n (%)			
Male	29 (42.7)	25 (38.5)	26 (31.3)
Female	39 (57.4)	40 (61.5)	57 (68.7)
Ethnicity, n (%)			
Hispanic or Latino	2 (2.9)	6 (9.2)	8 (9.6)
Not Hispanic or Latino	66 (97.1)	59 (90.8)	75 (90.4)
Race, n (%)			
Asian	1 (1.5)	1 (1.5)	4 (4.8)
White	57 (83.8)	54 (83.1)	68 (81.9)
Black or African American	6 (8.8)	5 (7.7)	6 (7.2)
American Indian or Alaska Native	1 (1.5)	—	2 (2.4)
Native Hawaiian or Other Pacific Islander	1 (1.5)	—	2 (2.4)
Other Indian	1 (1.5)	1 (1.5)	1 (1.2)
Not Reported	1 (1.5)	4 (6.2)	—
Fitzpatrick Skin Type, n (%)			
Type I	0	0	2 (2.4)
Type II	8 (11.8)	9 (13.9)	30 (36.1)
Type III	36 (52.9)	32 (49.2)	20 (24.1)
Type IV	17 (25.0)	19 (29.2)	19 (22.9)
Type V	1 (1.5)	0	9 (10.8)
Type VI	6 (8.8)	5 (7.7)	3 (3.6)

CRRS, Chin Retrusion Rating Scale; JCRS, Jawline Contour Rating Scale; TVRS, Temple Volume Rating Scale.

Table 3. Intra-rater Reliability

Scale	Round 1 vs round 2		
	Percentage exact agreement	Percentage within 1 grade	Weighted kappa coefficient (95% CI)
JCRS (N = 67)			
Left jawline			
Rater 1	67.2	92.6	0.711 (0.515, 0.908)
Rater 2	76.1	97.0	0.874 (0.779, 0.969)
Rater 3	59.7	91.0	0.743 (0.604, 0.881)
Mean			0.776 (0.561, 0.991)
Right jawline			
Rater 1	70.1	91.0	0.711 (0.514, 0.907)
Rater 2	76.1	97.0	0.861 (0.751, 0.971)
Rater 3	59.7	92.5	0.753 (0.615, 0.891)
Mean			0.775 (0.582, 0.968)
TVRS (N = 65)			
Left temple hollowing			
Rater 1	69.2	96.9	0.864 (0.795, 0.934)
Rater 2	80.0	100.0	0.920 (0.877, 0.962)
Rater 3	67.7	100.0	0.875 (0.819, 0.931)
Mean			0.886 (0.813, 0.959)
Right temple hollowing			
Rater 1	67.7	98.5	0.887 (0.832, 0.942)
Rater 2	75.4	100.0	0.903 (0.854, 0.953)
Rater 3	64.6	100.0	0.861 (0.800, 0.923)
Mean			0.884 (0.831, 0.937)
CRRS (N = 83)			
Rater 1	66.3	98.8	0.680 (0.555, 0.805)
Rater 2	79.5	100.00	0.862 (0.796, 0.927)
Rater 3	61.4	98.8	0.727 (0.626, 0.829)
Mean			0.756 (0.522, 0.991)

CI, confidence interval; CRRS, Chin Retrusion Rating Scale; JCRS, Jawline Contour Rating Scale; TVRS, Temple Volume Rating Scale.

Analysis of live validation of the scales through inter-rater reliability also indicated substantial to almost perfect agreement (Table 4, Supplemental Table 5). Substantial agreement was established between the 3 raters by overall ICC values ranging from 0.743 to 0.767 for live scale validation assessing chin retrusion and left and right temple hollowing. Furthermore, almost perfect inter-rater reliability was determined by overall ICC values of 0.845 and 0.843 for the left and right jawlines, respectively.

Table 4. Inter-rater Reliability

Scale	ICC (95% CI)	
	Round 1	Round 2
JCRS		
Left jawline	0.750 (0.651, 0.829)	0.845 (0.772, 0.898)
Right jawline	0.764 (0.667, 0.839)	0.843 (0.760, 0.900)
TVRS		
Left temple hollowing	0.750 (0.572, 0.852)	0.763 (0.558, 0.868)
Right temple hollowing	0.739 (0.581, 0.840)	0.767 (0.573, 0.868)
CRRS	0.736 (0.647, 0.810)	0.743 (0.651, 0.818)

CI, confidence interval; CRRS, Chin Retrusion Rating Scale; ICC, intraclass correlation coefficient; JCRS, Jawline Contour Rating Scale; TVRS, Temple Volume Rating Scale.

Additionally, almost perfect intra- and inter-rater reliabilities were established for the NLFRS, VPLRS, MFVLRS, LFRS, and CFRS (Supplemental Tables 4, 5).

Absolute differences in scores between pairs deemed clinically different vs not clinically different according to the rater's scaled-based assessment are summarized in Table 5, Supplemental Table 6. For all the assessed scales, the 95% CI for the pairs considered clinically different does not overlap with that of the pairs classified as not clinically different. These observations suggest that most ratings between the 2 photographs had an approximate 1-point difference for photograph pairs perceived as clinically different. In contrast, most ratings were identical for photograph pairs assessed as not clinically different. Therefore, the results confirm the clinical significance of a 1-point difference for the JCRS, TVRS, and CRRS (Table 5), as well as the NLFRS, VPLRS, MFVLRS, LFRS, and CFRS, at rest and dynamic (Supplemental Table 6).

DISCUSSION

Live validation of the presented series of LG Chem rating scales was performed using board-certified plastic surgeons and dermatologists as raters of a diverse pool of live subjects, as opposed to validation of the scales using computer-generated images or small number of representative patient images. The study demonstrated substantial to almost perfect intra- and inter-rater reliability of the JCRS, the TVRS, and the CRRS when utilized by trained raters to assess live subjects. Test-retest reliability was almost perfect for the NLFRS, VPLRS, MFVLRS, LFRS, and CFRS. For scales applied to the left- and right-hand sides or upper

Table 5. Clinical Significance

Scale	Absolute difference in scores			
	n	Mean (SD)	Range	95% CI
JCRS				
Clinically different pairs	63	1.44 (0.96)	0-4	1.20, 1.69
Not clinically different pairs	30	0.47 (0.68)	0-2	0.21, 0.72
TVRS				
Clinically different pairs	74	1.22 (0.82)	0-4	1.03, 1.41
Not clinically different pairs	19	0.16 (0.50)	0-2	0.08, 0.40
CRRS				
Clinically different pairs	72	1.22 (0.95)	0-3	1.00, 1.45
Not clinically different pairs	21	0.38 (0.59)	0-2	0.11, 0.65

CI, confidence interval; CRRS, Chin Retrusion Rating Scale; JCRS, Jawline Contour Rating Scale; TVRS, Temple Volume Rating Scale; SD, standard deviation.

and lower lips, features were rated with comparable reliability by all raters. Furthermore, the clinical significance of a 1-point difference on all the scales was established.

Weighted kappa coefficients and ICC values were calculated as a quantitative measure of intra- and inter-rater reliability. High kappa values indicate that the scales can be used by the same rater multiple times for consistent assessment and observation of changes. Conversely, the high ICC values indicate that the scales can be used for dependable evaluation by individual raters at different times.[18] Accordingly, high reliability scores are a primary consideration of a scale's utility and application.[19] Additional considerations include ease of use.[5] LG Chem's rating scales showcase vivid, well-framed images that effectively showcase the feature to be assessed. The 4- or 5-point scales are enhanced with descriptions and multiple images for each point. Furthermore, the images used over the validated scales represent various ethnicities, races, ages, and both sexes. The inherent diversity of LG Chem's scales may have contributed to the high test-retest reliability scores calculated from live validation of the assessed population, which included both men and women of various ages, races, ethnicities, and Fitzpatrick skin types.

Objective assessment tools for clinical use and evidence-based evaluation of new techniques are needed in the field of facial aesthetics.[3] Without such tools, determinants of success are subjective perceptions of the physician and patient.[4] Furthermore, using a validated scale can improve patient understanding of nonsurgical aesthetic procedures and may, therefore, assist clinicians in providing patients with realistic expectations.[20] Validation of the rating scales presented here warrants their potential application in pre-procedure assessment and evaluation of clinical outcomes. Although most of the scales assessed in this study represent changes that occur through the aging process, the chin retrusion scale may also reflect an individual soft tissue feature or bony structure that is present independent of age.

The high reliability, user-friendly layout, relevance to real-life populations, and extensiveness of the presented clinically significant rating scales will likely benefit clinical assessment of facial attractiveness despite the presence of previously published scales to rate various features.[5-11] Unlike the scales shown here, previously published scales of crow's feet, midface, and lower face that encompass nasolabial folds, lip fullness, lip wrinkles, and jawline show only one patient over the scales with computer-generated changes. Moreover, the patients lack diversity in skin type, race, and sex.[5,9,10] Some previously published jaw and temple rating scales do use real-world images; however, the TVRS and JCRS are unique as they display the area of interest from 2 angles.[5,7,8] The distinctive qualities of the LG Chem scales may translate to more straightforward and useful clinical assessment for real-world patients.

Study limitations include the small number of physicians who were recruited to perform live validation. Although the study used 3 board-certified dermatologists and plastic surgeons to effectively assess each rating scale, more aesthetics physicians could test the rating scales to generate inter- and intra-rater reliability scores that reflect a broader cross-section of the profession. Future studies might include live validation by aesthetic technicians who administer nonsurgical aesthetic treatments. Validation of rating

scales with images comprised of patients who underwent specific nonsurgical aesthetic procedures might also be helpful to measure patient outcomes in the clinical setting. Additional study limitations include the verbal descriptors for each grade on the developed scales as they are somewhat subjective to the raters' interpretation.

CONCLUSIONS

Live validation of the LG Chem comprehensive rating scales to assess facial features resulted in substantial to near-perfect inter- and intra-rater reliability for all scales, including the JCRS, TVRS, CRRS, NFLRS, VPLRS, MFVLRS, LFRS, and CFRS, at rest and dynamic. The user-friendly, clinically significant scales incorporate vibrant real-world images that represent diverse subjects, multiple images, and sometimes 2 angles for each scale point, as well as text descriptions to supplement the images. The high test-retest reliability and intuitive layout of these scales will likely contribute to the aesthetics field by providing an objective approach with standardized ratings for the clinical assessment of various facial features.

Supplemental Material

This article contains supplemental material located online at www.asjopenforum.com.

Acknowledgments

The authors would like to thank the scale validators: For chin and crow's feet as well as jawline contouring, perioral line, and temple hollowing, Dr David Bank of The Center for Dermatology, Cosmetic & Laser, Mount Kisco NY, USA; Dr Jason Bloom of Bloom Facial Plastic Surgery, Bryn Mawr, PA, USA; and Dr Robert Weiss of MD Laser Skin and Vein, Hunt Valley, MD, USA. For nasolabial folds, midface, and lips, Dr Deanne Robinson of Modern Dermatology of Connecticut, Westport, CT, USA; Dr Jason Bloom of Bloom Facial Plastic Surgery, Ardmore, PA, USA; and Dr Lawrence Bass, Bass Plastic Surgery, New York, NY, USA. The authors would also like to thank Brigid Stadinski, PhD, of Principal Medvantage, LLC, for her editorial assistance.

Disclosures

Dr Lorenc is a consultant for Allergan (Irvine, CA, USA), Galderma (Fort Worth, TX, USA), Merz (Raleigh, NC, USA), Suneva Medical, Inc (San Diego, CA, USA), and Thermi (Irving, TX, USA) and received honorarium from Canfield Scientific (Parsippany, NJ, USA) for scale development. Dr Jones received honorarium from Canfield Scientific for scale development. Ms Kim, Ms Gwak, and Dr Batham are full-time employees of LG Chem Ltd. Dr Vachon is a professional medical writer and was paid for her assistance in writing this manuscript.

Funding

Funding for this research was provided by LG Chem, Ltd., Seoul, Korea.

REFERENCES

1. Farolch-Prats L, Nome-Chamorro C. Facial contouring by using dermal fillers and botulinum toxin A: a practical approach. *Aesthetic Plast Surg.* 2019;43(3):793-802.
2. Akinbiyi T, Othman S, Familusi O, Calvert C, Card EB, Percec I. Better results in facial rejuvenation with fillers. *Plast Reconstr Surg Glob Open.* 2020;8(10):e2763.
3. Gupta S, Biskup N, Mattison G, Leis A. Development and validation of a clinical assessment tool for platysmal banding in cervicomental aesthetics of the female neck. *Aesthet Surg J.* 2015;35(6):NP141-NP146.
4. Lemperle G, Holmes RE, Cohen SR, Lemperle SM. A classification of facial wrinkles. *Plast Reconstr Surg.* 2001;108(6):1735-1750; discussion 1751.
5. Narins RS, Carruthers J, Flynn TC, et al. Validated assessment scales for the lower face. *Dermatol Surg.* 2012;38(2 Spec No.):333-342.
6. Sykes JM, Carruthers A, Hardas B, et al. Development and validation of a photonumeric scale for assessment of chin retrusion. *Dermatol Surg.* 2016;42 (Suppl 1):S211-S218.
7. Carruthers J, Jones D, Hardas B, et al. Development and validation of a photonumeric scale for evaluation of volume deficit of the temple. *Dermatol Surg.* 2016;42 (Suppl 1):S203-S210.
8. Moradi A, Lin X, Allen S, Fagien S, Norberg M, Smith S. Validation of photonumeric assessment scales for temple volume deficit, infraorbital hollows, and chin retrusion. *Dermatol Surg.* 2020;46(9):1148-1154.
9. Carruthers A, Carruthers J. A validated facial grading scale: the future of facial ageing measurement tools? *J Cosmet Laser Ther.* 2010;12(5):235-241.
10. Carruthers J, Flynn TC, Geister TL, et al. Validated assessment scales for the mid face. *Dermatol Surg.* 2012;38(2 Spec No.):320-332.
11. Lorenc ZP, Bank D, Kane M, Lin X, Smith S. Validation of a four-point photographic scale for the assessment of midface volume loss and/or contour deficiency. *Plast Reconstr Surg.* 2012;130(6):1330-1336.
12. Lipko-Godlewska S, Bolanča Ž, Kalinová L, et al. Whole-face approach with hyaluronic acid fillers. *Clin Cosmet Investig Dermatol.* 2021;14:169-178.
13. Harrar H, Myers S, Ghanem AM. Art or science? An evidence-based approach to human facial beauty a quantitative analysis towards an informed clinical aesthetic practice. *Aesthetic Plast Surg.* 2018;42(1):137-146.
14. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas*. 1973;33(3):613-619. doi:10.1177/001316447303300309
15. Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation—a discussion and demonstration of basic features. *PLoS One.* 2019;14(7):e0219854.

16. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86(2): 420-428.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
18. Lee JH, Choi YS, Park ES, et al. A novel photonumeric hand grading scale for hand rejuvenation. *Arch Plast Surg.* 2019;46(4):359-364.
19. Kaminer MS, Casabona G, Peeters W, et al. Validated assessment scales for skin laxity on the posterior thighs, buttocks, anterior thighs, and knees in female patients. *Dermatol Surg.* 2019;45 (Suppl 1):S12-S21.
20. Jandhyala R. Improving consent procedures and evaluation of treatment success in cosmetic use of incobotulinumtoxinA: an assessment of the treat-to-goal approach. *J Drugs Dermatol.* 2013;12(1):72-78.