

Development and Validation of a Photonumeric Scale for Assessment of Chin Retrusion

JONATHAN M. SYKES, MD,* ALASTAIR CARRUTHERS, MA, BM, BCH, FRCPC, FRCP(LON),[†]
BHUSHAN HARDAS, MD, MBA,[‡] DIANE K. MURPHY, MBA,[‡] DEREK JONES, MD,[§]
JEAN CARRUTHERS, MD,^{||} LISA DONOFRIO, MD,[¶] LELA CREUTZ, PHD,[#] ANN MARX, MD,[‡]
AND SARA DILL, MD[‡]

BACKGROUND A validated scale is needed for objective and reproducible comparisons of chin appearance before and after chin augmentation in practice and clinical studies.

OBJECTIVE To describe the development and validation of the 5-point photonumeric Allergan Chin Retrusion Scale.

METHODS The Allergan Chin Retrusion Scale was developed to include an assessment guide, verbal descriptors, morphed images, and real subject images for each scale grade. The clinical significance of a 1-point score difference was evaluated in a review of multiple image pairs representing varying differences in severity. Interrater and intrarater reliability was evaluated in a live-subject validation study ($N = 298$) completed during 2 sessions occurring 3 weeks apart.

RESULTS A difference of ≥ 1 point on the scale was shown to reflect a clinically meaningful difference (mean [95% confidence interval] absolute score difference, 1.07 [0.94–1.20] for clinically different image pairs and 0.51 [0.39–0.63] for not clinically different pairs). Intrarater agreement between the 2 live-subject validation sessions was substantial (mean weighted kappa = 0.79). Interrater agreement was substantial during the second rating session (0.68, primary end point).

CONCLUSION The Allergan Chin Retrusion Scale is a validated and reliable scale for physician rating of severity of chin retrusion.

Supported by Allergan plc, Dublin, Ireland. Editorial support for this article was provided by Peloton Advantage, Parsippany, New Jersey, and was funded by Allergan plc. The authors received an honorarium for participating in scale development and validation. B. Hardas, A. Marx, and D.K. Murphy are employees of Allergan plc. L. Creutz provided medical writing services at the request of the authors, which was funded by Allergan plc. The remaining authors have indicated no significant interest with commercial supporters. The opinions expressed in this article are those of the authors. The authors received no honorarium or other form of financial support related to the development of this article.

Chin projection is an important component of facial appearance and affects the overall balance and harmony of the face.^{1,2} Lack of adequate chin projection (i.e., chin retrusion) may be considered unattractive or a sign of aging.^{3–5} In addition to projection, the straightness, breadth, and length of the

chin are important components of chin attractiveness.^{2,6} Although the appearance of the chin is often unrelated to aging (e.g., microgenia),⁴ age-related changes in dentition and resorption of the mandible and maxilla may further alter chin appearance.^{7,8}

**Department of Otolaryngology/Facial Plastic Surgery, UC Davis Medical Group, Sacramento, California; [†]Department of Dermatology and Skin Science, University of British Columbia, Vancouver, British Columbia, Canada; [‡]Allergan plc, Irvine, California; [§]Division of Dermatology, University of California at Los Angeles, Los Angeles, California; ^{||}Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, British Columbia, Canada; [¶]Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut; [#]Peloton Advantage, LLC, Parsippany, New Jersey*

© 2016 by the American Society for Dermatologic Surgery, Inc. Published by Wolters Kluwer Health, Inc. All rights reserved.
ISSN: 1076-0512 • *Dermatol Surg* 2016;42:S211–S218 • DOI: 10.1097/DSS.0000000000000849

Chin augmentation procedures include a range of surgical techniques, such as osteotomy of the bony mentum or insertion of soft-tissue allografts or alloplastic implants, and more recently, less-invasive injectable soft-tissue filler treatment for chin correction.¹ Treatment of chin retrusion with filler injections has been reported as a component of full facial rejuvenation.^{9,10} There are currently no published studies of chin treatment that used a validated scale to objectively rate chin appearance before and after treatment. There is a need for a validated scale for rating severity of chin retrusion to help physicians objectively and reproducibly compare chin appearance before and after treatment in practice, as well as in clinical studies.

This report describes the development and validation of a new photometric scale designed to rate the severity of chin retrusion (Allergan Chin Retrusion Scale) using a combination of real and morphed subject images over a range of Fitzpatrick skin types. The objectives of this study were to determine the differences in scale scores that reflect clinically significant differences and to establish the validity and interrater and intrarater reliability of this scale for rating chin retrusion in live subjects.

Methods

Scale Development

Figure 1 summarizes key steps in the creation and validation of the Allergan Chin Retrusion Scale. A 9-member team comprising 5 external members (3 board-certified dermatologists, 1 board-certified facial plastic surgeon, and 1 board-certified oculoplastic surgeon) and 4 Allergan employees (2 dermatologists, 1 plastic surgeon, and 1 clinical scientist) developed the scale from a pool of subject images captured by Canfield Scientific, Inc. (Canfield, Fairfield, NJ). A total of 396 men and women aged 18 years or older with Fitzpatrick skin Types I through VI and in good general health volunteered for image capture. All subjects provided informed photo consent before image collection. Subjects were excluded if they had anything that would interfere with visual assessment of the area of interest (e.g., piercings, tattoos). Full 3-dimensional (3D) images of the face were obtained

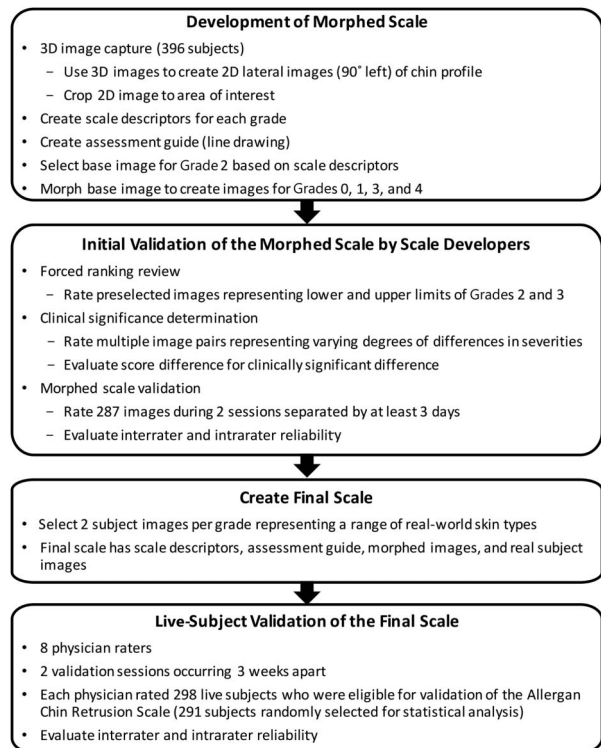


Figure 1. Development and validation processes for the Allergan Chin Retrusion Scale.

using a VECTRA M3 Camera with 3D Capture Software. The 3D images were used to produce lateral 2D images of the chin profile (90° left) and cropped from the tragus out to 0.5 cm from the tip of the nose and from the top ear to 2 cm below the chin.

Scale descriptors were created for each of the 5 grades of the scale (Table 1). Two members of the Allergan team met with each member of the scale development team for preliminary input on each scale grade. After preliminary scale grades were established, all 9 individuals involved in scale creation had a collaborative discussion about the scale grades and descriptors. The wording for each grade was then finalized by the Allergan team.

An assessment guide with a line drawing of anatomic markers demarcating the chin was created by Canfield based on detailed instructions from the Allergan team regarding anatomic markers (Figure 2). The drawing was then revised by Canfield multiple times after careful review by the Allergan team. The chin area of assessment was defined as the area between the lower

TABLE 1. Scale Descriptors for the Allergan Chin Retrusion Scale

Grade	Term	Descriptor
0	None	No chin retrusion; chin midpoint at or in front of the lower vermilion border vertical line
1	Minimal	Minimal chin retrusion; chin midpoint is between the labiomental sulcus vertical line and lower vermilion border vertical line
2	Moderate	Moderate chin retrusion; chin midpoint is at labiomental sulcus vertical line
3	Severe	Severe chin retrusion; chin midpoint slightly behind labiomental sulcus vertical line
4	Extreme	Extreme chin retrusion; chin midpoint significantly behind labiomental sulcus vertical line

Chin midpoint, the midpoint between the labiomental sulcus and the inferior point of the chin.

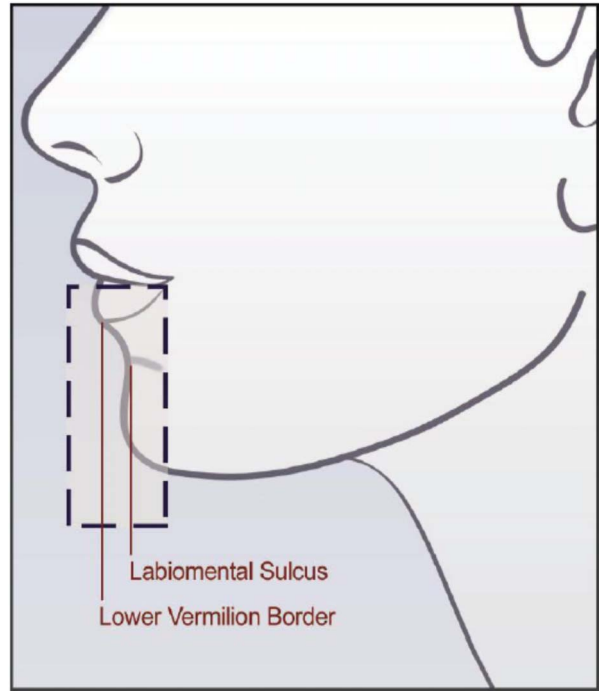


Figure 2. Assessment guide for the Allergan Chin Retrusion Scale.

lip vermilion border, the most projected part of the chin (bony pogonion), and the most inferior point of the chin (bony menton) from a lateral view.

A base image to demonstrate Grade 2 chin retrusion was selected, and this image was morphed to represent all 5 Grades of the scale. A Canfield graphics technician morphed the anatomic area of interest in the base image to match the descriptors provided for Grades 0, 1, 3, and 4. Anatomic correctness and photo alignment with the scale descriptors of the morphed images were achieved through an interactive process with the Allergan team.

A forced ranking review was performed to delineate the range of severity between Grades 2 and 3 and to confirm the selection of the best representative image to be used as Grade 2 on the scale. The 5 external scale developers performed the web-based forced ranking exercise on preselected images that represented the upper and lower boundaries of Grades 2 and 3.

To determine whether there was a clinically significant difference between grades of the scale, the 5 external

scale developers were asked to perform an on-line clinical significance review. Multiple image pairs were selected to represent varying degrees of differences in severity (ranging from no difference to a 4-point difference). During the session, the scale developers determined whether there was a clinically significant difference (Yes/No) between images for each pair. After the session, the individual images from all image pairs were randomly mixed in with other images to be used in the morphed image scale validation (described in the following paragraph) and assigned a score by scale developers so that score differences between each image in each pair could be calculated.

The morphed image scale was validated by having the scale developers use the scale to rate randomized images representing all scale grades during 2 web-based sessions occurring at least 3 days apart. A total of 287 images (120 in session 1 and 167 in session 2) were rated. The scale had acceptable interrater and intrarater agreement (>0.5), so scale development proceeded using the morphed images.

For both the clinical significance review and the morphed image scale validation review, scale

developers were provided uniform hardware by Canfield to complete the reviews. Before the reviews, the scale developers completed web-based PowerPoint training to familiarize themselves with the hardware, the review platform, and the purpose of the clinical significance and morphed image validation reviews. The scale developers were not allowed to discuss the review with one another, and each completed the image review independently.

After the morphed scale was created, 2 subject photos representing each grade of the scale were selected to represent diversity in sex and Fitzpatrick skin type per grade. The final scale contains the scale descriptors for each grade, an assessment guide, the morphed images, and the real subject images (Figure 3).

Scale Validation

The interrater and intrarater reliability of the final scale was evaluated in a live-subject rating validation study. Eight external physician raters experienced in using aesthetic photonumeric scales who were not involved in scale development participated in two 2-day live validation sessions occurring 3 weeks apart. Before the first live validation session, all physician raters were trained on the use of the scale in an interactive group training session using 4 example subjects. Raters were instructed that the chin midpoint was defined as the vertical point halfway between the labiomental sulcus and the most inferior point of the chin. For someone with a severely retruded chin, the most inferior part of the chin may be quite far back from the vertical line. For grade 0 and grade 2, which describe the chin midpoint “at” either the lower vermilion border vertical line or the labiomental sulcus vertical line, the term “at” was defined as ± 1 mm.

All subjects who qualified for the initial image capture events were invited to attend the live validation sessions. Subjects were instructed to arrive at the study center clean-shaven, to remove make-up and jewelry, to wear dark pants or jeans and a provided black T-shirt, to not drink alcohol excessively before the sessions, to try not to alter their usual routine (e.g., their facial care routine and normal sleep or hydration patterns) between sessions, and to not have tanning

Allergan Chin Retrusion Scale

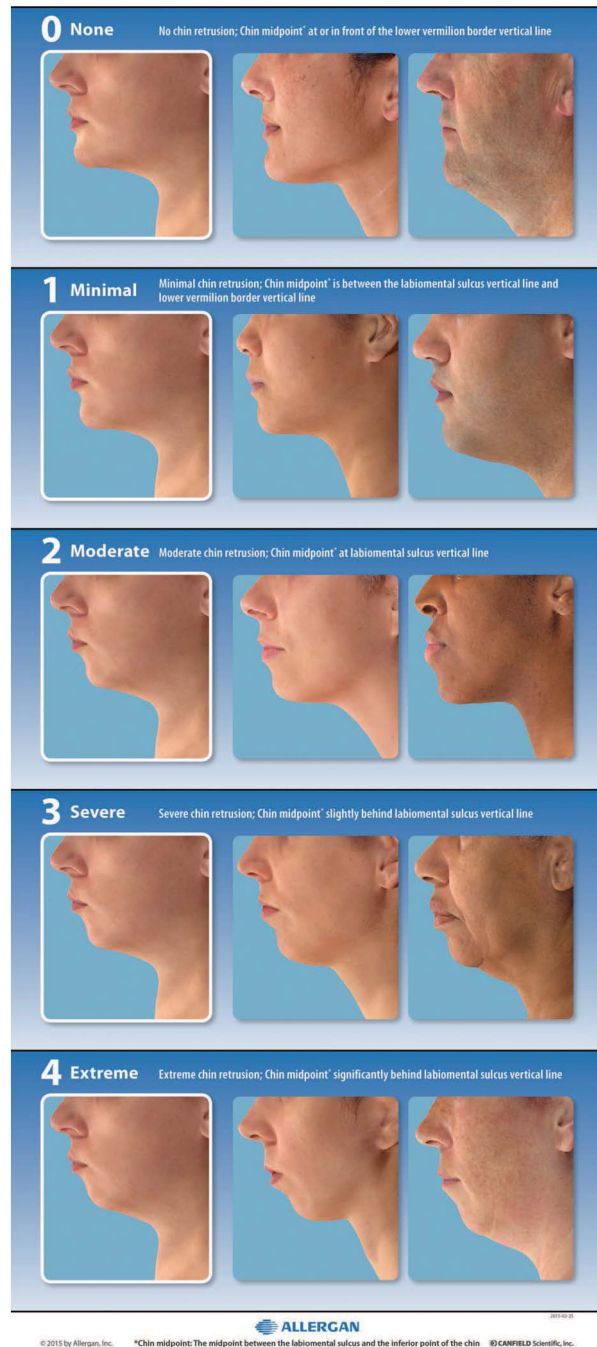
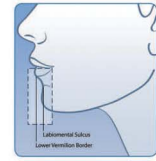


Figure 3. The Allergan Chin Retrusion Scale assigns a grade from none (0) to extreme (4) that describes the degree of the chin's placement posterior to the normal position, as defined in the diagram in the upper right corner.

sessions or extensive sun exposure between sessions. On arrival at the study center for the first live validation session, subjects signed informed consent and were assessed for eligibility, age, sex, race (as reported by the subject), and Fitzpatrick skin type (determined by the investigator). Subjects were excluded if they had their photographs included in the scale; anything that would interfere with visual assessment of the chin; any treatment with toxin/fillers, dental procedures, or surgery that would alter chin appearance within 2 weeks of the first validation session or plans to have one of these procedures between the 2 validation sessions; or diagnosis of pregnancy. 3D images of each subject were collected at the first live validation session using a VECTRA M3 Camera with 3D Capture Software. The first 5 subjects rated during the first validation session were considered run-in training subjects and were excluded from the analysis.

During the first and second live validation sessions, each physician rater evaluated all subjects on all scales (7 additional scales for other anatomic features were evaluated at the same sessions and are reported separately¹¹⁻¹⁷). Raters had separate evaluation stations with an examination lamp, table, a stool for subject seating, supplies, and the photometric scale mounted and displayed for use in subject evaluation. Subjects presented themselves to each rater individually and proceeded from one rating station to the next in the same order until evaluated by all 8 raters. Raters were instructed to not discuss ratings with subjects or other raters. The raters took at least a 10-minute break every hour and at least a 30-minute lunch break to avoid rater fatigue.

Statistics

To determine the utility of the scale grades for detecting clinically meaningful differences in chin appearance, absolute score differences for the image pairs deemed “clinically different” or “not clinically different” during scale development were summarized (mean, SD, range, 95% confidence interval [CI]). For the live scale validation study, intrarater reliability was compared between round 1 and round 2 scores by calculating weighted kappa scores using Fleiss-Cohen weights.¹⁸ Kappa scores within the range of 0.0 to 0.20 indicate slight agreement, 0.21 to 0.40 indicate fair

agreement, 0.41 to 0.60 indicate moderate agreement, 0.61 to 0.80 indicate substantial agreement, and 0.81 to 1.00 indicate almost perfect agreement.¹⁹ Interrater agreement was measured by determining the intraclass correlation coefficient (ICC [2,1]) and 95% CIs calculated using the formula described by Shrout and Fleiss.²⁰ The a priori primary end point for the interrater agreement analysis was ICC (2,1) for the second rating session. SAS version 9.3 (Cary, NC) was used for all statistical analyses.

Sample Size Considerations

The sample size for the live-subject validation sessions was calculated using the method described by Bonett.²¹ With up to 10 raters and an ICC of 0.5, a total of 66 subjects were needed for the scale to have a 95% CI with a width of 0.2 for interrater reliability. Considering potential loss of subjects between the 2 rounds, at least 80 subjects were to be enrolled for the scale. Because 298 subjects were eligible for the chin scale validation analysis, the number of subjects evaluated using the scale was substantially larger than the preplanned sample size of 80, and the overall number of assessments for some grades of the scale was larger than those for the other grades. To minimize imbalance in the number of subjects across scale grades and to meet the sample size requirement, the mean score across the 8 raters for each subject was used to assign an overall grade for each subject, and a subset of 81 subjects with minimum imbalance across the grades (~16 subjects per each of the 5 scale grades) was randomly selected from the eligible subjects using a prespecified procedure. This random selection of the subset was performed 20 times. Interrater and intrarater agreements calculated for each of the 20 subsets were combined using SAS procedure PROC MIANALYZE to obtain the overall interrater and intrarater agreements.

Results

Clinical Significance Determination by Scale Developers

Mean (95% CI) absolute difference in scores was 1.07 (0.94–1.20) for image pairs identified as clinically different and 0.51 (0.39–0.63) for image pairs identified as not clinically different (Table 2). The 95% CIs

TABLE 2. Difference in Scores for Image Pairs Deemed Clinically Different or Not Clinically Different Using the Allergan Chin Retrusion Scale

	<i>n</i> *	<i>Absolute Difference in Scores</i>		
		<i>Mean (SD)</i>	<i>Range</i>	<i>95% CI for Mean</i>
Clinically different pairs	154	1.07 (0.83)	0–3	0.94–1.20
Not clinically different pairs	86	0.51 (0.57)	0–2	0.39–0.63

**N* = 240 = 48 pairs × 5 raters; *n* = no. of pairs in each category.
CI, confidence interval; SD, standard deviation.

for the pairs deemed to be clinically different did not overlap with the CIs for the pairs deemed not clinically different, confirming that a 1-point difference in scores is clinically significant.

Live-Subject Scale Validation

A total of 298 subjects were eligible for the chin scale validation analysis, and 291 subjects were selected in at least one of the 20 random subsets for analysis of intrarater and interrater agreement. Demographic characteristics of subjects in the final scale validation set are shown in Table 3. Most subjects were female (67%), Caucasian (79%), and had Fitzpatrick skin Type III (29%) or IV (31%). Median age was 48 years, and a broad span of ages was represented (18–83 years).

Intrarater agreement between the 2 live-subject rating sessions was substantial (mean weighted kappa = 0.79) (Table 4). Interrater agreement was substantial during the first (0.71) and second (0.68, primary end point) rating sessions (Table 4).

Discussion

This study demonstrated substantial interrater and intrarater agreement for the Allergan Chin Retrusion Scale, suggesting that multiple assessments for the same subject and across different raters are reliable. A 1-point difference in ratings was shown to reflect clinically significant differences, indicating that the scale has sufficient sensitivity for detecting clinically significant changes in the severity of chin retrusion.

The chin and submental area of the neck are common facial areas for aesthetic treatment,²² and interest in chin profiles has increased with the recent approval of

a nonsurgical treatment (deoxycholic acid injection) for moderate to severe convexity or fullness associated with submental fat.^{23–25} The Allergan Chin Retrusion Scale provides standardized ratings, along with verbal descriptors for each grade describing the degree of the chin's location posterior to the "ideal" position and a facial diagram clearly defining chin landmarks. These factors likely contributed to the high interrater reliability and may translate to ease of use by clinicians. The use of morphed images to represent each grade helps to focus the rater's attention on the change from one grade to the next, as all other features remain

TABLE 3. Demographics of Subjects in the Live Scale Validation Study

<i>Characteristic</i>	<i>Number (%) of Subjects, N = 291</i>
Sex, <i>n</i> (%)	
Female	194 (66.7)
Male	97 (33.3)
Age, yrs	
Median	48
Range (min, max)	18–83
Fitzpatrick skin type, <i>n</i> (%)	
I	23 (7.9)
II	57 (19.6)
III	83 (28.5)
IV	91 (31.3)
V	23 (7.9)
VI	14 (4.8)
Race, <i>n</i> (%)	
Caucasian	230 (79.0)
Hispanic or Latino	30 (10.3)
Asian	15 (5.2)
African American	14 (4.8)
American Indian or Alaska Native	1 (0.3)
Caucasian/Hispanic or Latino	1 (0.3)

TABLE 4. Physician Intrarater and Interrater Agreement on the Allergan Chin Retrusion Scale (Validation Testing with Live Subjects)

Intrarater agreement	
Mean weighted kappa (95% CI)	0.79 (0.693–0.882)
Interrater agreement	
Round 1, ICC (95% CI)	0.71 (0.627–0.786)
Round 2,* ICC (95% CI)	0.68 (0.605–0.765)
*Primary end point.	
CI, confidence interval; ICC, intraclass correlation coefficient.	

constant across scale grades. The inclusion of real-world images representing a diverse range of skin types across sexes and races is also important, as morphed images may not always translate clinically to the broad array of physical appearances or physical changes observed in the aging face.

The treatment-planning process for improvement in chin appearance should include aesthetic assessment of all aspects of the lower face, including chin length and width, as well as the lateral regions (i.e., the pre-jowl sulcus) and the labiomental sulcus, to optimize outcomes of treatment.^{8,26,27} It should be noted that the Allergan Chin Retrusion Scale determines chin projection in relation to the labiomental sulcus, the shape and depth of which can vary with chin length and may change after treatment.^{27,28}

Attractiveness scores rated by clinicians may be significantly different from layperson and patient ratings, as clinicians may or may not be as critical as non-treating observers or patients.²⁹ A lack of standardized methods may lead to significant variability in what patients are told about their appearance and treatments offered from one practice to another.³⁰ Use of a validated scale for formalized and reproducible consultation procedures may empower patients to make informed treatment decisions³¹ and potentially lead to overall improvements in patient satisfaction. The FACE-Q is a validated patient satisfaction scale with a chin subscale that may be helpful for capturing the patient's perspective on appearance before and after chin treatment.³² In the experience of the authors, patient satisfaction with improvements in chin appearance is highly subjective, with many patients

expressing dissatisfaction with unnatural appearance and sudden drastic changes in the chin, especially after surgery. Use of fillers for chin augmentation, which allows for more gradual changes in stages, may help avoid patient dissatisfaction when subtle changes are desired.

Study Limitations

This scale measures chin retrusion as a variation of horizontal chin projection from the ideal position. As such, the scale and its validation used lateral photographs. The chin is a 3D structure that can be measured horizontally (as in this study), vertically, and transversely. Additionally, chin contour and shape, the position and depth of the labiomental sulcus, and lateral volume can be analyzed. This study provides validated measurement of the horizontal projection of the chin only.

The clinical significance of chin scale scores was determined solely by the scale developers. Although a 1-point change on the scale was considered meaningful to the scale developers, it may or may not be meaningful to subjects. A change of less than 1 point may be meaningful for patients desiring a subtle change, whereas other subjects may perceive only dramatic changes as meaningful; hence, this scale is not recommended for patient self-assessment of meaningful improvement. The verbal descriptors for each grade on the scale are subjective; however, the descriptors were developed and refined by extensive feedback between 9 experts to minimize inherent subjectivity.

Conclusions

The Allergan Chin Retrusion Scale demonstrated substantial interrater and intrarater agreement among physicians, and 1-point score differences were shown to reflect clinically meaningful differences in chin retrusion. This scale includes a user-friendly diagram, detailed verbal descriptions, and morphed and real subject images representative across sexes and skin types to provide standardized ratings that can be uniformly applied in clinical trials and by dermatologists and plastic surgeons who treat men and women seeking enhancement of the chin.

Acknowledgments The authors thank the following physicians for completing the scale validation study: David E. Bank, MD, FAAD; Sue Ellen Cox, MD; Timothy M. Greco, MD, FACS; Z. Paul Lorenc, MD, FACS; David J. Narins, MD, PC, FACS; William B. Nolan, MD; Robert A. Weiss, MD; and Margaret Weiss, MD. Statistical support was provided by Yijun Sun, PhD, and Shraddha Mehta, PhD of Allergan plc, Irvine, California.

References

- Bertossi D, Galzignato PF, Albanese M, Botti C, et al. Chin microgenia: a clinical comparative study. *Aesthet Plast Surg* 2015;39:651–8.
- Mittelman H, Spencer JR, Chrzanowski DS. Chin region: management of grooves and mandibular hypoplasia with alloplastic implants. *Facial Plast Surg Clin North Am* 2007;15:445–60; vi.
- Khosravanifard B, Rakhshan V, Raeesi E. Factors influencing attractiveness of soft tissue profile. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2013;115:29–37.
- Naini FB, Donaldson AN, McDonald F, Cobourne MT. Assessing the influence of chin prominence on perceived attractiveness in the orthognathic patient, clinician and layperson. *Int J Oral Maxillofac Surg* 2012;41:839–46.
- Modarai F, Donaldson JC, Naini FB. The influence of lower lip position on the perceived attractiveness of chin prominence. *Angle Orthod* 2013; 83:795–800.
- Zhang Z, Tang R, Tang X, Yu B, et al. The oblique mandibular chin-body osteotomy for the correction of broad chin. *Ann Plast Surg* 2010; 65:541–5.
- Carruthers JD, Glogau RG, Blitzer A. Advances in facial rejuvenation: botulinum toxin type a, hyaluronic acid dermal fillers, and combination therapies—consensus recommendations. *Plast Reconstr Surg* 2008;121 (Suppl 5):5S–30S.
- Romo T, Yalamanchili H, Sclafani AP. Chin and prejowl augmentation in the management of the aging jawline. *Facial Plast Surg* 2005;21: 38–46.
- Raspaldo H. Volumizing effect of a new hyaluronic acid sub-dermal facial filler: a retrospective analysis based on 102 cases. *J Cosmet Laser Ther* 2008;10:134–42.
- Bae JM, Lee DW. Three-dimensional remodeling of young Asian women's faces using 20-mg/ml smooth, highly cohesive, viscous hyaluronic acid fillers: a retrospective study of 320 patients. *Dermatol Surg* 2013;39:1370–5.
- Jones D, Donofrio L, Hardas B, Murphy DK, et al. Development and validation of a photonic scale for evaluation of volume deficit of the hand. *Dermatol Surg* 2016;42(Suppl 10):S195–202.
- Carruthers J, Jones D, Hardas B, Murphy DK, et al. Development and validation of a photonic scale for evaluation of volume deficit of the temple. *Dermatol Surg* 2016;42(Suppl 10):S203–10.
- Donofrio L, Carruthers A, Hardas B, Murphy DK, et al. Development and validation of a photonic scale for evaluation of facial skin texture. *Dermatol Surg* 2016;42(Suppl 10):S219–26.
- Carruthers J, Donofrio L, Hardas B, Murphy DK, et al. Development and validation of a photonic scale for evaluation of facial fine lines. *Dermatol Surg* 2016;42(Suppl 10):S227–34.
- Jones D, Carruthers A, Hardas B, Murphy DK, et al. Development and validation of a photonic scale for evaluation of transverse neck lines. *Dermatol Surg* 2016;42(Suppl 10):S235–42.
- Carruthers A, Donofrio L, Hardas B, Murphy DK, et al. Development and validation of a photonic scale for evaluation of static horizontal forehead lines. *Dermatol Surg* 2016;42(Suppl 10):S243–50.
- Donofrio L, Carruthers J, Hardas B, Murphy DK, et al. Development and validation of a photonic scale for evaluation of infraorbital hollows. *Dermatol Surg* 2016;42(Suppl 10):S251–58.
- Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measure of reliability. *Educ Psychol Meas* 1973;33:613–9.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
- Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21:1331–5.
- Cosmetic Surgery National Data Bank Statistics. *American Society for Aesthetic Plastic Surgery*, 2014. 2014. Available at: <http://www.surgery.org/sites/default/files/2014-Stats.pdf>. Accessed July 21, 2016.
- Ascher B, Hoffmann K, Walker P, Lippert S, et al. Efficacy, patient-reported outcomes and safety profile of ATX-101 (deoxycholic acid), an injectable drug for the reduction of unwanted submental fat: results from a phase III, randomized, placebo-controlled study. *J Eur Acad Dermatol Venereol* 2014;28:1707–15.
- Rzany B, Griffiths T, Walker P, Lippert S, et al. Reduction of unwanted submental fat with ATX-101 (deoxycholic acid), an adipocytolytic injectable treatment: results from a phase III, randomized, placebo-controlled study. *Br J Dermatol* 2014;170:445–53.
- Kybella [package insert]. Westlake Village, CA: Kythera Biopharmaceuticals, Inc.; 2015.
- Fattahi T. The prejowl sulcus: an important consideration in lower face rejuvenation. *J Oral Maxillofac Surg* 2008;66:355–8.
- Rosen HM. Aesthetic refinements in genioplasty: the role of the labiomental fold. *Plast Reconstr Surg* 1991;88:760–7.
- Frodel JL, Sykes JM, Jones JL. Evaluation and treatment of vertical microgenia. *Arch Facial Plast Surg* 2004;6:111–9.
- Torsello F, Graci M, Grande NM, Deli R. Relationships between facial features in the perception of profile attractiveness. *Prog Orthod* 2010; 11:92–7.
- Williams LM, Alderman JE, Cussell G, Goldston J, et al. Patient's self-evaluation of two education programs for age-related skin changes in the face: a prospective, randomized, controlled study. *Clin Cosmet Investig Dermatol* 2011;4:149–59.
- 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;12:72–80.
- Schwitzer JA, Klassen AF, Cano SJ, Baker SB, et al. Measuring satisfaction with appearance: validation of the FACE-Q scales for the nose, forehead, cheekbones, and chin. *Plast Reconstr Surg* 2015;136:140–1.

Address correspondence and reprint requests to: Jonathan M. Sykes, MD, Department of Otolaryngology/Facial Plastic Surgery, UC Davis Medical Center, 2521 Stockton Boulevard, 6th Floor, Suite 6206, Sacramento, CA 95817, or e-mail: jmsykes@ucdavis.edu