

Development and Validation of a Photonumeric Scale for Evaluation of Infraorbital Hollows

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

BACKGROUND A validated scale is needed for objective and reproducible comparisons of infraorbital hollows (i.e., tear troughs) before and after treatment in practice and clinical studies.

OBJECTIVE To describe the development and validation of the 5-point photonumeric Allergan Infraorbital Hollows Scale.

METHODS The scale was developed to include an assessment guide, verbal descriptors, morphed images, and real subject images for each grade. The clinical significance of a 1-point score difference was evaluated in a review of image pairs representing varying differences in severity. Interrater and intrarater reliability was evaluated in a live-subject validation study ($N = 297$) completed during 2 sessions occurring 3 weeks apart.

RESULTS A score difference of ≥ 1 point was shown to reflect a clinically significant difference (mean [95% confidence interval] absolute score difference, 0.90 [0.79–1.02] for clinically different image pairs and 0.33 [0.19–0.46] for not clinically different pairs). Intrarater agreement between the 2 validation sessions was substantial (mean weighted kappa = 0.79). Interrater agreement was substantial during the second rating session (0.70, primary end point).

CONCLUSION The Allergan Infraorbital Hollows Scale is a validated and reliable scale for physician rating severity of hollowing in the infraorbital area.

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.

Tear troughs, which appear due to hollowing of the infraorbital area, are one of the most common areas for which patients seek aesthetic treatment of the face.¹ Several methods have been used to treat infraorbital hollowing, including subdermal filler injections,^{2–4} autologous fat transfer,^{5,6} radiosurgery,⁷ topical agents,⁸ lower eyelid blepharoplasty,⁹ and

laser resurfacing.¹⁰ The high demand for treatment of the infraorbital area calls for the use of validated reliable instruments to objectively assess severity of hollowing before and after treatment.

This report describes the development and validation of a new Allergan photonumeric scale for rating the

*Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut; [†]Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, British Columbia, Canada; [‡]Allergan plc, Irvine, California; [§]Division of Dermatology, University of California, Los Angeles, Los Angeles, California; ^{||}UC Davis Medical Group, Sacramento, California; [¶]Department of Dermatology and Skin Science, University of British Columbia, Vancouver, British Columbia, Canada; [#]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:S251–S258 • DOI: 10.1097/DSS.0000000000000856

severity of infraorbital hollows. The scale was developed to meet FDA requirements for outcome assessments to be used in clinical trials.^{11,12} The objectives of this study were to determine the clinically significant difference in scale scores and to establish the interrater and intrarater reliability of the scale for rating the severity of infraorbital hollows in live subjects.

Methods

Scale Development

Figure 1 summarizes key steps in the creation and validation of the Allergan Infraorbital Hollows 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 untreated 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 photograph consent before image collection. Subjects were excluded if they had anything that would interfere with visual assessment of the area of interest. Full three-dimensional (3D) images of the face (0° frontal image with raked lighting) were obtained using a VECTRA M3 Camera with 3D Capture Software. The 3D images were cropped horizontally from the midline to the temporal line/hairline and vertically starting 1 cm above the superior orbital rim down to the subnasale to produce 2D images of the eye area.

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 anatomical markers demarcating the area of interest between the

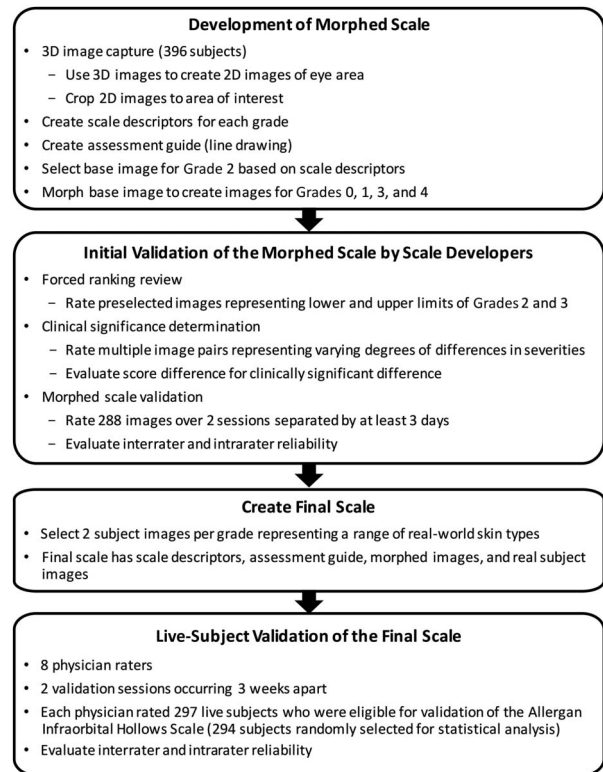


Figure 1. Development and validation processes for the Allergan Infraorbital Hollows Scale.

medial canthus line and the lid–cheek junction extending to the lateral bony orbital rim was created by Canfield based on instructions from the Allergan team (Figure 2). The drawing was then revised by Canfield multiple times after careful review by the Allergan team.

A base image to demonstrate Grade 2 infraorbital hollows was selected, and this image was morphed to represent all 5 grades of the scale. A Canfield graphics technician morphed the infraorbital area of interest in the base image to match the descriptors provided for Grades 0, 1, 3, and 4. Alignment of the morphed images with the scale descriptors was 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 a web-based forced ranking exercise on preselected images that represented the upper and lower boundaries of Grades 2 and 3.

TABLE 1. Allergan Infraorbital Hollows Scale Descriptors

<i>Grade</i>	<i>Term</i>	<i>Descriptor</i>
0	None	No visible hollowing or volume loss medially or laterally
1	Minimal	Presence of hollowing with some volume loss medial to the mid-pupillary line; smooth lateral lid-cheek transition
2	Moderate	Defined hollowing extending laterally beyond the mid-pupillary line with moderate volume loss; smooth lateral lid-cheek transition with mild volume loss
3	Severe	Defined hollowing extending laterally beyond the mid-pupillary line with moderate volume loss creating a defined groove along the lid-cheek junction
4	Extreme	Defined hollowing extends from medial to lateral canthus; severe volume loss creates a marked step along the lid-cheek junction

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 the external scale developers so that score differences between each image in each pair could be calculated.

The morphed image scale was validated by having the 5 external scale developers use the scale to rate randomized images representing all grades of the scale during 2 web-based sessions occurring at least 3 days apart. A total of 288 images were rated (120 images in Session 1 and 168 images in Session 2). The scale had acceptable interrater and intrarater agreement (>0.5), so scale development proceeded using the morphed images.

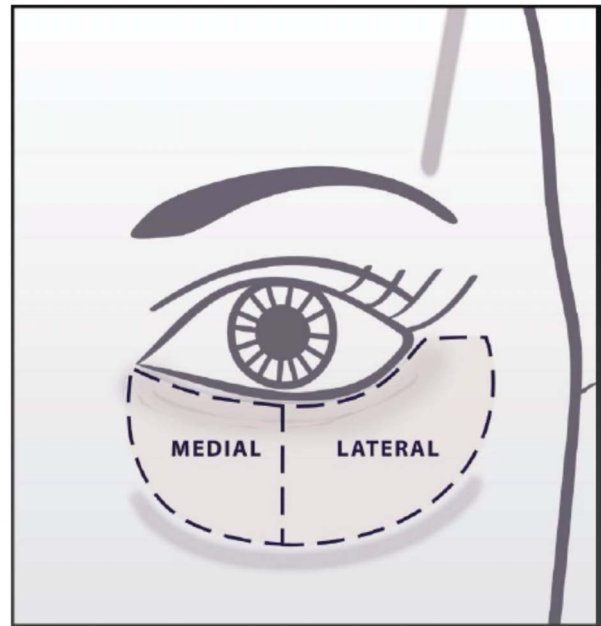


Figure 2. Assessment guide for the Allergan Infraorbital Hollows Scale.

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 external 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 external 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 photographs representing each grade of the scale were selected to represent diversity in sex and Fitzpatrick skin type per grade. The final scale includes 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 physician raters experienced in using aesthetic photonumeric scales who were not involved in scale development participated in two



Figure 3. The Allergan Infraorbital Hollows Scale assigns a grade from none (0) to extreme (4) describing the degree of volume deficit within the infraorbital area shown in the diagram in the upper right corner.

2-day live validation sessions occurring 3 weeks apart. Before the first live evaluation 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 to focus their attention on the tear trough and lateral lid–cheek junction and to disregard the prominence of the globe, hypertrophy of the orbicularis oculi,

prominent fat pads, and/or skin redundancy, as these may mask or detract from the presence or absence of true infraorbital hollowing.

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 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 then 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 infraorbital area; any treatment with toxin/fillers or surgery that would alter appearance within 2 weeks of the first evaluation session or plans to have one of these procedures between the 2 evaluation sessions; or diagnosis of pregnancy. 3D images of each subject were collected at the first live validation session. 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 scale 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 separately^{13–19}). Raters had separate evaluation stations with an examination lamp, table, a stool for subject seating, supplies, and the photonic 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 significant differences, 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.²⁰ 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.²¹ 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.²² 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.²³ 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. Because 297 subjects were eligible for the 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 this scale was larger than that 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. A subset of 104 subjects with minimal 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

The mean (95% CI) absolute difference in scale scores was 0.90 (0.79–1.02) for clinically different image pairs and 0.33 (0.19–0.46) for pairs deemed not clinically different (Table 2). The 95% CIs 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 297 subjects were eligible for the scale validation analysis, and 294 subjects were selected in at least one of the 20 random subsets for analysis of intrarater and interrater agreement. Demographic

TABLE 2. Differences in Scores for Image Pairs Deemed Clinically Different or Not Clinically Different Using the Allergan Infraorbital Hollows Scale

	Absolute Difference in Scores			
	<i>n</i> *	Mean (SD)	Range	95% CI for Mean
Clinically different pairs	182	0.90 (0.79)	0–3	0.79–1.02
Not clinically different pairs	58	0.33 (0.51)	0–2	0.19–0.46

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

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 (28%) or IV (32%). 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 rating session (ICC = 0.66) and second rating session (ICC = 0.70; primary end point) (Table 4).

Discussion

This study demonstrated substantial interrater and intrarater agreement for the Allergan Infraorbital Hollows Scale, indicating that multiple assessments of the infraorbital area of the same subject and across different raters are reliable. A 1-point difference in scale ratings was shown to reflect clinically significant differences, demonstrating that the scale grades have

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

Characteristic	N = 294
Sex, n (%)	
Female	198 (67.3)
Male	96 (32.7)
Age, yrs	
Median	48
Range (min, max)	18–83
Fitzpatrick skin type, n (%)	
I	24 (8.2)
II	59 (20.1)
III	81 (27.6)
IV	93 (31.6)
V	22 (7.5)
VI	15 (5.1)
Race, n (%)	
Caucasian	232 (78.9)
Hispanic or Latino	30 (10.2)
African American	16 (5.4)
Asian	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 Infraorbital Hollows Scale (Validation Testing With Live Subjects)

Intrarater agreement	
Mean weighted kappa (95% CI)	0.79 (0.667–0.904)
Interrater agreement	
Round 1, ICC (95% CI)	0.66 (0.581–0.736)
Round 2,* ICC (95% CI)	0.70 (0.629–0.772)
*Primary end point.	
CI, confidence interval; ICC, intraclass correlation coefficient.	

the sensitivity to detect clinically meaningful changes in hollowing of the infraorbital area.

Numerous scales have been used to assess infraorbital hollows,^{2,8,24–26} including one validated scale.^{2,8,24} The interrater and intrarater reliability of the Allergan Infraorbital Hollows Scale is similar to that reported for the one other validated infraorbital hollows scale.²⁴ In that study, 12 experts rated photographs of 50 subjects, demonstrating substantial interrater (0.66–0.72) and intrarater (0.77) reliability. However, the scale was not validated in live subjects and did not include real subject images. The Allergan Infraorbital Hollows Scale demonstrated substantial reliability for rating live subjects in this study and is the first validated scale to include real images of subjects representing both sexes or multiple ethnicities and skin types.

The Allergan Infraorbital Hollows Scale assesses one of the most concerning aesthetic areas for patients.¹ Infraorbital hollowing manifests as loss of volume causing a shadow and darkness in the infraorbital area that can result in an overly tired, stressed, and aged appearance. In the authors' experience, treatment to improve infraorbital hollowing using dermal fillers is very common and popular with patients. The authors have found that it is essential not to overfill the area but rather to aim for a mild undercorrection. Use of standardized scales is important from the patient's perspective, as it can minimize variability among clinics in the information patients receive regarding their appearance and treatment options, and also establish realistic expectations for treatment outcomes.^{27,28}

Study Limitations

The clinical significance of scale scores was determined solely by the scale developers; although a 1-point change on each scale was considered meaningful to the scale developers, it may or may not be meaningful to subjects. A less than 1-point change 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 validated FACE-Q appraisal of lower eyelids scale²⁹ may be helpful for capturing the patient's perspective on appearance before and after treatment of infraorbital hollows. The verbal descriptors for each grade on the scale are subjective. However, the descriptors were developed and refined by extensive feedback from 9 experienced clinicians, minimizing inherent subjectivity.

Conclusions

The Allergan Infraorbital Hollows Scale demonstrated substantial intrarater and interrater agreement among physicians and a 1-point score difference was shown to reflect clinically significant differences in infraorbital hollowing. This unique infraorbital hollowing scale includes user-friendly diagrams, detailed verbal descriptions, and morphed and real subject images representative of both sexes and across races and Fitzpatrick skin types to provide standardized ratings that may be uniformly applied by clinicians who treat patients seeking treatment for hollowing in the infraorbital area.

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, CA.

References

- Narurkar V, Shamban A, Sissins P, Stonehouse A, et al. Facial treatment preferences in aesthetically aware women. *Dermatol Surg* 2015;41(Suppl 1):S153–S60.
- Hevia O, Cohen BH, Howell DJ. Safety and efficacy of a cohesive polydensified matrix hyaluronic acid for the correction of infraorbital hollow: an observational study with results at 40 weeks. *J Drugs Dermatol* 2014;13:1030–6.
- Bernardini FP, Cetinkaya A, Devoto MH, Zambelli A. Calcium hydroxyl-apatite (Radiesse) for the correction of periorbital hollows, dark circles, and lower eyelid bags. *Ophthal Plast Reconstr Surg* 2014;30:34–9.
- Viana GA, Osaki MH, Cariello AJ, Damasceno RW, et al. Treatment of the tear trough deformity with hyaluronic acid. *Aesthet Surg J* 2011;31:225–31.
- Ciuci PM, Obagi S. Rejuvenation of the periorbital complex with autologous fat transfer: current therapy. *J Oral Maxillofac Surg* 2008;66:1686–93.
- Kranendonk S, Obagi S. Autologous fat transfer for periorbital rejuvenation: indications, technique, and complications. *Dermatol Surg* 2007;33:572–8.
- Pak CS, Lee YK, Jeong JH, Kim JH, et al. Safety and efficacy of ulthera in the rejuvenation of aging lower eyelids: a pivotal clinical trial. *Aesthet Plast Surg* 2014;38:861–8.
- Seidel R, Moy RL. Reduced appearance of under-eye bags with twice-daily application of epidermal growth factor (EGF) serum: a pilot study. *J Drugs Dermatol* 2015;14:405–10.
- Grant JR, Laferriere KA. Periocular rejuvenation: lower eyelid blepharoplasty with fat repositioning and the suborbicularis oculi fat. *Facial Plast Surg Clin North Am* 2010;18:399–409.
- Ma G, Lin XX, Hu XJ, Jin YB, et al. Treatment of venous infraorbital dark circles using a long-pulsed 1,064-nm neodymium-doped yttrium aluminum garnet laser. *Dermatol Surg* 2012;38:1277–82.
- Kane MA, Blitzer A, Brandt FS, Glogau RG, et al. Development and validation of a new clinically-meaningful rating scale for measuring lateral canthal line severity. *Aesthet Surg J* 2012;32:275–85.
- U.S. Department of Health and Human Services, Food and Drug Administration. *Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. 2009. Available from: <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf>. Accessed July 21, 2016.
- 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.
- Sykes JM, Carruthers A, Hardas B, Murphy DK, et al. Development and validation of a photonic scale for assessment of chin retrusion. *Dermatol Surg* 2016;42(Suppl 10):S211–18.
- 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.

20. 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.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
22. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
23. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21:1331–5.
24. Carruthers J, Flynn TC, Geister TL, Gortelmeyer R, et al. Validated assessment scales for the mid face. *Dermatol Surg* 2012;38:320–32.
25. Mally P, Czyz CN, Wulc AE. The role gravity periorbital midfacial aging. *Aesthet Surg J* 2014;34:809–22.
26. Sadick NS, Bosniak SL, Cantisano-Zilkha M, Glavas IP, et al. Definition of the tear trough and the tear trough rating scale. *J Cosmet Dermatol* 2007;6:218–22.
27. 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.
28. 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.
29. Pusic AL, Klassen AF, Scott AM, Cano SJ. Development and psychometric evaluation of the FACE-Q satisfaction with appearance scale: a new patient-reported outcome instrument for facial aesthetics patients. *Clin Plast Surg* 2013;40:249–60.

Address correspondence and reprint requests to: Lisa Donofrio, MD, Department of Dermatology, Yale University School of Medicine, The Savin Dermatology Center, 134 Park Street, New Haven, CT 06511, or e-mail: dr_donofrio@att.net