

# Development and Validation of a Photonumeric Scale for Evaluation of Facial Skin Texture

LISA DONOFRIO, MD,\* ALASTAIR CARRUTHERS, MA, BM, BCH, FRCPC, FRCP(LON),<sup>†</sup>  
BHUSHAN HARDAS, MD, MBA,<sup>‡</sup> DIANE K. MURPHY, MBA,<sup>‡</sup> JEAN CARRUTHERS, MD,<sup>§</sup>  
DEREK JONES, MD,<sup>||</sup> JONATHAN M. SYKES, MD,<sup>¶</sup> LELA CREUTZ, PhD,<sup>#</sup> ANN MARX, MD,<sup>‡</sup>  
AND SARA DILL, MD<sup>‡</sup>

---

**BACKGROUND** A validated scale is needed for objective and reproducible comparisons of facial skin roughness before and after aesthetic treatment in practice and in clinical studies.

**OBJECTIVE** To describe the development and validation of the 5-point photonumeric Allergan Skin Roughness 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 ( $N = 290$ ) completed during 2 sessions occurring 3 weeks apart.

**RESULTS** A score difference of  $\geq 1$  point was shown to reflect a clinically meaningful difference (mean [95% confidence interval] absolute score difference 1.09 [0.96–1.23] for clinically different image pairs and 0.53 [0.38–0.67] for not clinically different pairs). Intrarater agreement between the 2 validation sessions was almost perfect (weighted kappa = 0.83). Interrater agreement was almost perfect during the second rating session (0.81, primary end point).

**CONCLUSION** The Allergan Skin Roughness Scale is a validated and reliable scale for physician rating of midface skin roughness.

*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.*

---

Facial skin that is smooth and appears nontextured is an important component of facial attractiveness.<sup>1</sup> Intrinsic structural changes in the skin due to aging and extrinsic environmental factors (e.g., sun exposure) can lead to skin unevenness (e.g., roughness).<sup>2–6</sup> Rejuvenation procedures to improve skin texture have become increasingly popular with

the growing availability of advanced skin care products and minimally invasive therapies (e.g., topical retinoids,<sup>7,8</sup> intradermal filler injections,<sup>9,10</sup> mesotherapy,<sup>11,12</sup> photodynamic therapy,<sup>13</sup> chemical peels,<sup>14,15</sup> and laser skin resurfacing<sup>16</sup>). The number of such treatments in the United States has steadily increased over the past decade, with more than 2.3

\*Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut; <sup>†</sup>Department of Dermatology and Skin Science, University of British Columbia, Vancouver, British Columbia, Canada; <sup>‡</sup>Allergan plc, Irvine, California; <sup>§</sup>Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, British Columbia, Canada; <sup>||</sup>Division of Dermatology, University of California at Los Angeles, Los Angeles, California; <sup>¶</sup>UC Davis Medical Group, Sacramento, California; <sup>#</sup>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:S219–S226 • DOI: 10.1097/DSS.0000000000000852

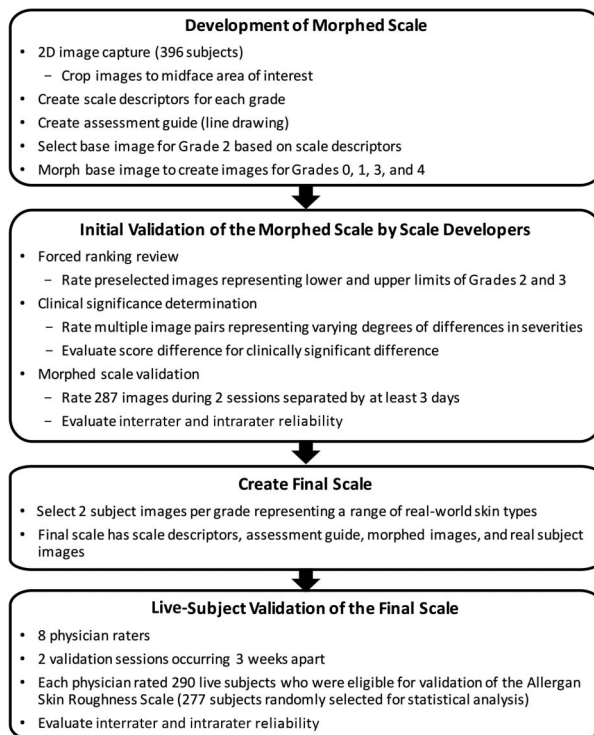
million soft-tissue filler procedures, 1.25 million chemical peels, and 500,000 laser skin resurfacing procedures reported to the American Society of Plastic Surgeons in 2014.<sup>17</sup>

Given the increasing popularity of skin texture treatments, there is a great need for a validated scale for objective and reproducible comparison of skin quality pretreatment and post-treatment. Although a number of biophysical techniques have been used for objectively measuring skin quality, including skin replicas<sup>18,19</sup> and optical image analyses,<sup>2,4,10,19,20</sup> use of these methods requires access to costly technical equipment that limits their utility in clinical practice. This report describes the development and validation of a photometric scale for the evaluation of facial skin roughness (Allergan Skin Roughness Scale), created to meet FDA requirements for outcome assessments in clinical trials of new treatments<sup>21</sup> and to provide a tool that physicians can use for practical and long-term assessments of patients. 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 skin roughness in live subjects.

## Methods

### Scale Development

Figure 1 summarizes key steps in the creation and validation of the Allergan Skin Roughness 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. Full-facial, 2-dimensional (2D) images were obtained using a 2D custom studio suite



**Figure 1.** Development and validation processes for the Allergan Skin Roughness Scale.

for facial imaging (Nikon D7100 Hi Res SLR). 2D images taken at 45° of the left side of the face were cropped horizontally from the mid glabella to the tragus and vertically from the lower eyelids to just below the chin to produce images of the midface area of interest for the Allergan Skin Roughness Scale.

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.

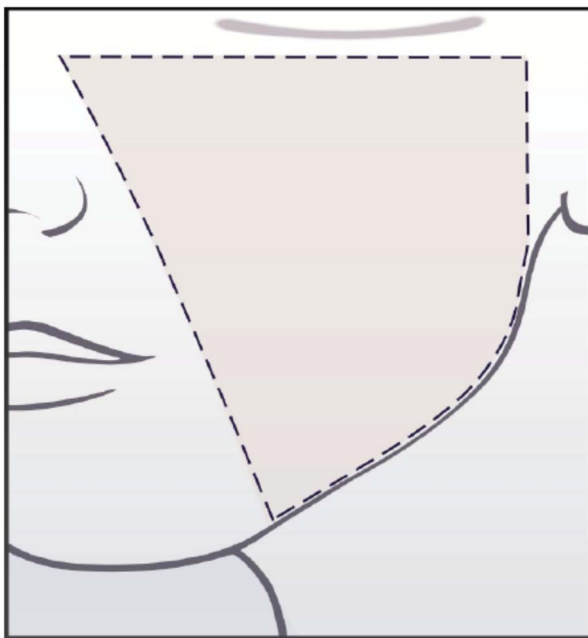
An assessment guide with a line drawing of anatomic markers demarcating the area of interest was created by Canfield based on detailed instructions from the Allergan team regarding anatomic markers (Figure 2). The drawing was then revised by Canfield multiple times based on careful review by the Allergan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

**TABLE 1. Descriptors for the Allergan Skin Roughness Scale**

<i>Grade</i>	<i>Term</i>	<i>Descriptor</i>
0	None	Smooth visual skin texture
1	Minimal	Slightly coarse and uneven visual skin texture
2	Moderate	Moderately coarse and uneven visual skin texture; may have early elastosis
3	Severe	Severely coarse visual skin texture, crosshatched fine lines; may have some elastosis
4	Extreme	Extremely coarse visual skin texture, crosshatched deep creases; extreme elastosis

Roughness Scale was defined as the area between the nasolabial fold to the preauricular cheek and from the inferior orbital rim to the mandible.

A base image to demonstrate Grade 2 skin roughness 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. Alignment of the morphed images with the scale descriptors was achieved through an interactive process with the Allergan team.



**Figure 2.** Assessment guide for the Allergan Skin Roughness Scale.

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 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 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. The scale developers rated a total of 287 images (120 images in Session 1 and 167 images in Session 2). 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, 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 contained 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 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 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 disregard acne scars, prominent pores, and epidermal lesions such as seborrheic keratoses, nevi, sebaceous hyperplasia, etc., and to focus on the texture or roughness of the facial epidermis itself.

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 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 area of interest; any treatment with toxin/fillers, dental procedures, or surgery that would alter the areas of interest within 2 weeks of the first evaluation session or plans to have one of these procedures between the 2 sessions; or diagnosis of pregnancy. 2D images of each subject were collected at the first live validation session using a 2D custom studio suite for facial imaging with Nikon



**Figure 3.** The Allergan Skin Roughness Scale assigns a grade from none (0) to extreme (4) that describes the severity of skin coarseness, crosshatching, and elastosis in the midface area delineated in the diagram shown in the upper right corner.

D7100 Hi Res SLR. 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<sup>22–28</sup>). 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 significant differences in skin roughness, 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-subject scale validation study, intrarater reliability was compared between Round 1 and Round 2 scores by calculating weighted kappa scores using Fleiss-Cohen weights.<sup>29</sup> 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.<sup>30</sup> 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.<sup>31</sup> 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.<sup>32</sup> With up to 10 raters and an ICC of 0.5,

a total of 66 subjects were needed 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 290 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, 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**

The mean (95% CI) absolute difference in scores was 1.09 (0.96–1.23) for image pairs deemed clinically different and 0.53 (0.38–0.67) for image pairs not deemed clinically different (Table 2). The 95% CIs for the pairs deemed 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 290 subjects were eligible for the Allergan Skin Roughness Scale validation and were rated using the scale; 277 of these 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 (80%), and had Fitzpatrick skin Type III (28%) or IV

**TABLE 2. Difference in Scores on the Allergan Skin Roughness Scale for Image Pairs Deemed Clinically Different or Not Clinically Different**

	<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	181	1.09 (0.899)	0–3	0.96–1.23
Not clinically different pairs	59	0.53 (0.568)	0–2	0.38–0.67

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

(31%). Median age was 48 years, and a broad span of ages was represented (range, 18–83 years).

Intrarater agreement between the 2 live-subject rating sessions was almost perfect (mean weighted kappa = 0.83) (Table 4). Interrater agreement was substantial during the first rating session (ICC = 0.77) and almost perfect during the second session (ICC = 0.81; primary end point) (Table 4).

## Discussion

The almost-perfect level of interrater and intrarater agreement observed in the validation study indicate

that the Allergan Skin Roughness Scale is reliable and reproducible for the classification of skin roughness in the cheek and midface areas of live subjects. 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 skin roughness. The Allergan Skin Roughness Scale is the first validated scale for the assessment of facial skin texture or roughness. The increasing popularity of procedures intended to improve skin texture,<sup>17</sup> along with an aging population, signal the need for a practical and reliable visual assessment tool accompanied by detailed verbal descriptions for pretreatment and post-treatment use by clinicians. Such a scale can have multiple uses including teaching, training, and importantly, informing and building trust with patients.<sup>33,34</sup>

The Allergan Skin Roughness Scale describes visual changes in skin texture in words and requires that raters visually assess skin roughness, a characteristic that is tactile in nature. Rough skin may have a pebbly stucco-like texture and skin unevenness and may be

**TABLE 3. Demographics of Subjects in the Live Scale Validation Study for the Allergan Skin Roughness Scale**

<i>Characteristic</i>	<i>N = 277</i>
<b>Sex, <i>n</i> (%)</b>	
Female	186 (67.1)
Male	91 (32.9)
<b>Age, yrs</b>	
Median	48
Range (min–max)	(18–83)
<b>Fitzpatrick skin type, <i>n</i> (%)</b>	
I	22 (7.9)
II	57 (20.6)
III	77 (27.8)
IV	86 (31.0)
V	22 (7.9)
VI	13 (4.7)
<b>Race, <i>n</i> (%)</b>	
Caucasian	221 (79.8)
Hispanic or Latino	28 (10.1)
African American	14 (5.1)
Asian	12 (4.3)
American Indian or Alaska Native	1 (0.4)
Caucasian/Hispanic or Latino	1 (0.4)

**TABLE 4. Physician Intrarater and Interrater Agreement on the Allergan Skin Roughness Scale (Validation Testing With Live Subjects)**

<b>Intrarater agreement</b>	
Mean weighted kappa score (95% CI)	0.83 (0.596–1.072)
<b>Interrater agreement</b>	
Round 1, ICC (95% CI)	0.77 (0.706–0.834)
Round 2,* ICC (95% CI)	0.81 (0.759–0.865)

\*Primary end point.  
CI, confidence interval; ICC, intraclass correlation coefficient.

associated with a loss of elasticity. Although users of the scale may find it helpful to imagine what the skin would feel like if they ran their fingers across the subject's skin, the scale requires that texture be assessed only visually, without touching the skin. This may be a challenge, but the high interrater and intrarater reliability demonstrated in the scale validation study show that this approach is effective for the evaluation of skin roughness. It is important to note that acne scars, prominent pores, epidermal lesions such as seborrheic keratoses, nevi, sebaceous hyperplasia, and fine lines and wrinkles (other than cross-hatching) were ignored when selecting a grade on the Allergan Skin Roughness Scale. A separate validated scale, the Allergan Fine Lines Scale, was developed for the assessment of facial skin appearance based on the density of superficial fine lines (published separately<sup>25</sup>).

### Study Limitations

Clinical difference image comparisons by the scale developers showed that a 1-point change on the scale is indicative of a clinically significant difference in skin texture by experienced clinicians. However, meaningful differences may vary from the subject's perspective, with a <1-point change considered meaningful for patients desiring a subtle change and other subjects perceiving only dramatic changes as meaningful. This scale is not intended for patient self-assessment of clinically meaningful improvements. The FACE-Q satisfaction with skin scale may help to evaluate patient satisfaction with skin texture treatment.<sup>35,36</sup> Although the verbal descriptors for each grade on the Allergan Skin Roughness Scale are subjective, the descriptions were developed and refined by extensive discussion among a team of 9 experts to minimize inherent subjectivity.

### Conclusions

Because of the increasing popularity of skin texture treatments, there is a need for a validated scale for assessment of skin texture. The Allergan Skin Roughness Scale demonstrated almost perfect intrarater and interrater agreement among physicians, and 1-point score differences on the scale were shown to reflect clinically significant differences in skin texture.

Although the scale was developed to assess treatment effectiveness in clinical trials as per FDA guidelines, the scale also provides standardized ratings that may be of practical use for dermatologists and plastic surgeons who treat men and women seeking to improve facial skin texture.

**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

1. Fink B, Grammer K, Thornhill R. Human (*Homo sapiens*) facial attractiveness in relation to skin texture and color. *J Comp Psychol* 2001;115:92–9.
2. Trojahn C, Dobos G, Lichterfeld A, Blume-Peytavi U, et al. Characterizing facial skin ageing in humans: disentangling extrinsic from intrinsic biological phenomena. *Biomed Res Int* 2015;2015:318586.
3. Fisher GJ, Wang ZQ, Datta SC, Varani J, et al. Pathophysiology of premature skin aging induced by ultraviolet light. *N Engl J Med* 1997;337:1419–28.
4. Callaghan TM, Wilhelm KP. A review of ageing and an examination of clinical methods in the assessment of ageing skin. Part 2: clinical perspectives and clinical methods in the evaluation of ageing skin. *Int J Cosmet Sci* 2008;30:323–32.
5. Fisher GJ, Varani J, Voorhees JJ. Looking older: fibroblast collapse and therapeutic implications. *Arch Dermatol* 2008;144:666–72.
6. Le Louarn C, Buthiau D, Buis J. Structural aging: the facial recurve concept. *Aesthetic Plast Surg* 2007;31:213–8.
7. Ho ET, Trookman NS, Sperber BR, Rizer RL, et al. A randomized, double-blind, controlled comparative trial of the anti-aging properties of non-prescription tri-retinol 1.1% vs. prescription tretinoin 0.025%. *J Drugs Dermatol* 2012;11:64–9.
8. Ogden S, Samuel M, Griffiths CE. A review of tazarotene in the treatment of photodamaged skin. *Clin Interv Aging* 2008;3:71–8.
9. Lee BM, Han DG, Choi WS. Rejuvenating effects of facial hydrofilling using restylane vital. *Arch Plast Surg* 2015;42:282–7.
10. Kerscher M, Bayrhammer J, Reuther T. Rejuvenating influence of a stabilized hyaluronic acid-based gel of nonanimal origin on facial skin aging. *Dermatol Surg* 2008;34:720–6.
11. Savoia A, Landi S, Baldi A. A new minimally invasive mesotherapy technique for facial rejuvenation. *Dermatol Ther (Heidelb)* 2013;3:83–93.
12. Prikhnenko S. Polycomponent mesotherapy formulations for the treatment of skin aging and improvement of skin quality. *Clin Cosmet Investig Dermatol* 2015;8:151–7.
13. Karrer S, Kohl E, Feise K, Hiepe-Wegener D, et al. Photodynamic therapy for skin rejuvenation: review and summary of the literature—results of

- a consensus conference of an expert group for aesthetic photodynamic therapy. *J Dtsch Dermatol Ges* 2013;11:137–48.
14. Ghersetich I, Brazzini B, Peris K, Cotellessa C, et al. Pyruvic acid peels for the treatment of photoaging. *Dermatol Surg* 2004;30:32–6.
  15. Berardesca E, Cameli N, Primavera G, Carrera M. Clinical and instrumental evaluation of skin improvement after treatment with a new 50% pyruvic acid peel. *Dermatol Surg* 2006;32:526–31.
  16. Rhie JW, Shim JS, Choi WS. A pilot study of skin resurfacing using the 2,790-nm erbium:YSGG laser system. *Arch Plast Surg* 2015;42:52–8.
  17. American Society of Plastic Surgeons. *2014 Plastic Surgery Statistics Report*. 2015. Available from: <http://www.plasticsurgery.org/Documents/news-resources/statistics/2014-statistics/plastic-surgery-statistics-full-report.pdf>. Accessed July 21, 2016.
  18. Ryu JH, Seo YK, Boo YC, Chang MY, et al. A quantitative evaluation method of skin texture affected by skin ageing using replica images of the cheek. *Int J Cosmet Sci* 2014;36:247–52.
  19. Tchivaleva L, Zeng H, Markhvida I, McLean DI, et al. Skin roughness assessment. In: Campolo D, ed. *New Developments in Biomedical Engineering*. Rijeka, Croatia: InTech; 2010:341–58.
  20. Wunsch A, Matuschka K. A controlled trial to determine the efficacy of red and near-infrared light treatment in patient satisfaction, reduction of fine lines, wrinkles, skin roughness, and intradermal collagen density increase. *Photomed Laser Surg* 2014;32:93–100.
  21. 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.
  22. Jones D, Donofrio L, Hardas B, Murphy DK, et al. Development and validation of a photonumeric scale for evaluation of volume deficit of the hand. *Dermatol Surg* 2016;42(Suppl 10):S195–202.
  23. Carruthers J, Jones D, Hardas B, Murphy DK, et al. Development and validation of a photonumeric scale for evaluation of volume deficit of the temple. *Dermatol Surg* 2016;42(Suppl 10):S203–10.
  24. Sykes JM, Carruthers A, Hardas B, Murphy DK, et al. Development and validation of a photonumeric scale for assessment of chin retrusion. *Dermatol Surg* 2016;42(Suppl 10):S211–18.
  25. Carruthers J, Donofrio L, Hardas B, Murphy DK, et al. Development and validation of a photonumeric scale for evaluation of facial fine lines. *Dermatol Surg* 2016;42(Suppl 10):S227–34.
  26. Jones D, Carruthers A, Hardas B, Murphy DK, et al. Development and validation of a photonumeric scale for evaluation of transverse neck lines. *Dermatol Surg* 2016;42(Suppl 10):S235–42.
  27. Carruthers A, Donofrio L, Hardas B, Murphy DK, et al. Development and validation of a photonumeric scale for evaluation of static horizontal forehead lines. *Dermatol Surg* 2016;42(Suppl 10):S243–50.
  28. Donofrio L, Carruthers J, Hardas B, Murphy DK, et al. Development and validation of a photonumeric scale for evaluation of infraorbital hollows. *Dermatol Surg* 2016;42(Suppl 10):S251–58.
  29. 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.
  30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
  31. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
  32. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21:1331–5.
  33. 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.
  34. 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.
  35. 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.
  36. Klassen AF, Cano SJ, Schwitzer JA, Baker SB, et al. Development and psychometric validation of FACE-Q skin, lips and facial rhytides appearance scales and adverse effect checklists for cosmetic procedures. *JAMA Dermatol* 2016;152:443–51.

---

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