

Development and Validation of a Photonumeric Scale for Evaluation of Volume Deficit of the Temple

JEAN CARRUTHERS, MD,* DEREK JONES, MD,[†] BHUSHAN HARDAS, MD, MBA,[‡]
DIANE K. MURPHY, MBA,[‡] LISA DONOFRIO, 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 temple appearance before and after aesthetic treatment in practice and clinical studies.

OBJECTIVE To describe the development and validation of the 5-point photonumeric Allergan Temple Hollowing 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 = 298) 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, 1.1 [0.94–1.26] for clinically different image pairs and 0.67 [0.51–0.83] for not clinically different pairs). Intrarater agreement between the 2 validation sessions was almost perfect (mean weighted kappa = 0.86). Interrater agreement was almost perfect during the second session (0.81, primary endpoint).

CONCLUSION The Allergan Temple Hollowing Scale is a validated and reliable scale for physician rating of temple volume deficit.

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.

Loss of subdermal adipose tissue, or lipoatrophy, can lead to loss of soft-tissue fullness in multiple areas of the face including the temple area.^{1–3} Lipoatrophy may be associated with genetic disorders (e.g., rare autosomal recessive conditions), treatment of acquired diseases (e.g., antiretroviral therapy), physical trauma, the normal course of aging, and low body fat at any age.^{3,4} Age-related lipoatrophy

happens slowly and symmetrically, whereas disease-related loss of subdermal fat may be more rapid and asymmetric and can be associated with psychological issues including body image distortions or withdrawal from social activities.³

As facial tissue ages, the temporal bone becomes increasingly concave, and the overlying temporalis

*Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, British Columbia, Canada; [†]Division of Dermatology, University of California at Los Angeles, Los Angeles, California; [‡]Allergan plc, Irvine, California; [§]Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut; ^{||}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:S203–S210 • DOI: 10.1097/DSS.0000000000000848

muscle decreases in volume, resulting in loss of the convexity and fullness of the temporal region that is associated with a youthful appearance.^{5,6} Although fullness in the temporal region contributes to overall facial shape and balance, perceptions of ideal face shape differ culturally.⁷⁻⁹ For example, many Asian women prefer an oval facial shape, with fullness in the upper half of the face and tapering from the cheek to the chin.¹⁰

Several aesthetic techniques have been used to treat temple hollowing, including surgical alloplasty, autologous fat transfer, and subdermal filler injections.^{5,6} Knowledge of temporal anatomy is critical to safely achieving optimal temple reflation, particularly with injectable options.^{11,12} Four studies have evaluated the effectiveness of hyaluronic acid fillers for the treatment of temple hollowing.^{5,13-15} However, only 2 of the studies included objective assessments of the severity of temple hollowing before and after treatment. In both studies, the scales used had not been validated and did not include example images, relying solely on verbal descriptors of each grade.^{14,15}

This report describes the development and validation of a new photometric scale designed to rate the severity of volume deficit in the temple (Allergan Temple Hollowing 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 clinically significant difference in scale scores and to establish the interrater and intrarater reliability of the scale for rating volume deficit in the temple in live subjects.

Methods

Scale Development

Figure 1 summarizes key steps in the creation and validation of the Allergan Temple Hollowing 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 collected by

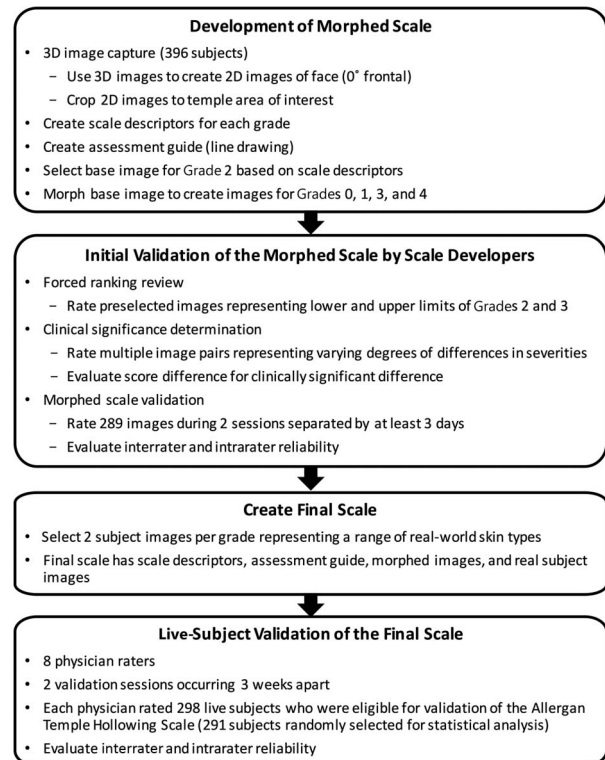


Figure 1. Development and validation processes for the Allergan Temple Hollowing Scale.

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 3-dimensional (3D) images of the face were obtained using a VECTRA M3 Camera with 3D Capture Software. The 3D images were used to create 2D images of the face (0° frontal), which were then cropped from the face midline to the ear margin and from the anterior hairline to the subnasale to ensure the left temple area was the primary focus and fully visible.

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

TABLE 1. Descriptors for the Allergan Temple Hollowing Scale

<i>Grade</i>	<i>Term</i>	<i>Descriptor</i>
0	Convex	Rounded temple
1	Flat	Flat temple; temporal fusion line may be visible
2	Minimal	Shallow depression or concavity with minimal volume loss; temporal fusion line may be visible
3	Moderate	Moderate depression or concavity with moderate volume loss; moderate prominence of temporal fusion line
4	Severe	Deeply recessed, sunken appearance; marked prominence of temporal fusion line and zygomatic arch

wording for each grade was then finalized by the Allergan team.

An assessment guide with a line drawing of anatomic markers demarcating the temple 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 after careful review by the Allergan team. The temple area of assessment was defined as the

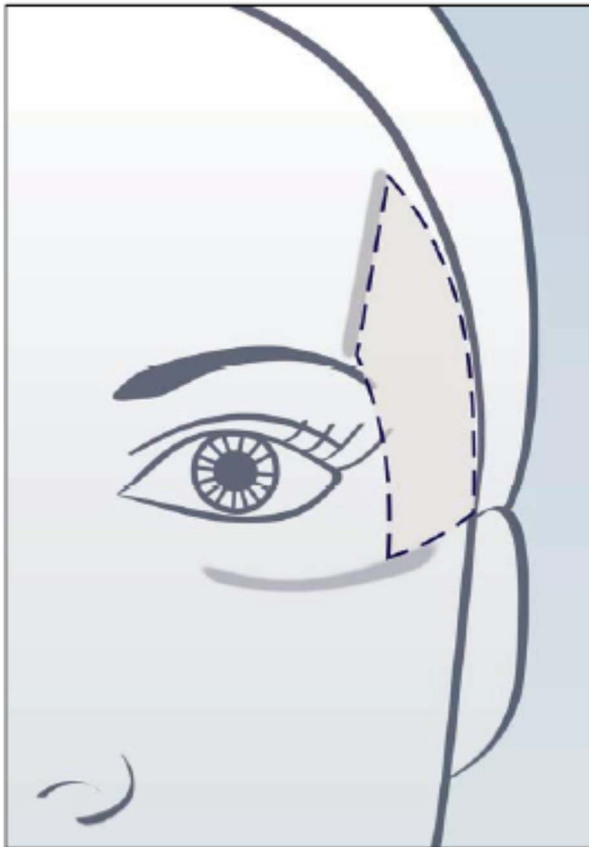


Figure 2. Assessment guide for the Allergan Temple Hollowing Scale.

area between the temporal fusion line, the zygomatic arch, the lateral orbital rim, and the hairline.

A base image to demonstrate Grade 2 temple hollowing was selected, and this image was morphed to represent all 5 grades of the scale. Convex temples were defined as the lower limit of temple hollowing (Grade 0) so that the scale may be used with Asian patients. This results in Caucasian subjects generally being limited to Grades 1 to 4 on the scale, as convex temples are not a sought-after appearance for this racial group. 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.

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 online 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 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 289 images were rated (120 images in session 1 and 169 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 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 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 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 to select the grade that represents the most

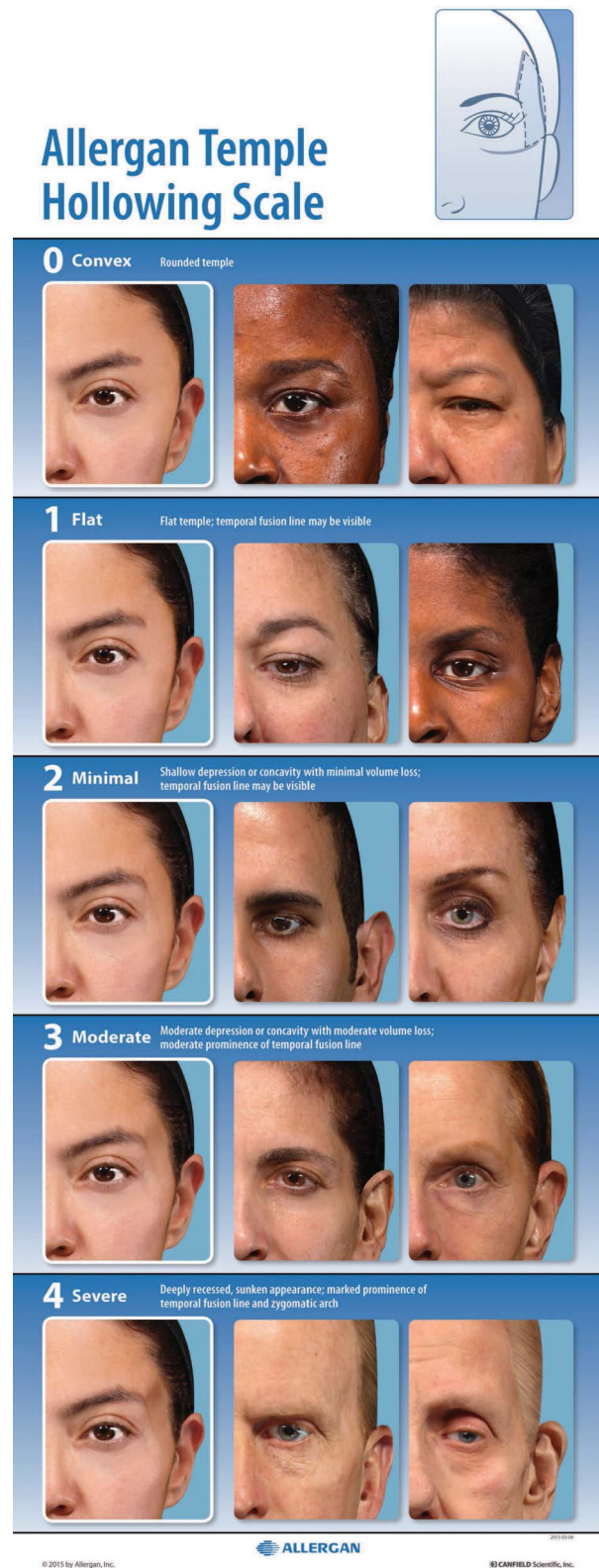


Figure 3. The Allergan Temple Hollowing Scale assigns a grade from convex (0) to severe (4) that describes the degree of volume deficit within the area delineated in the diagram shown in the upper right corner.

affected area of the temple, which may have included either the superior or inferior area or the entire temple. Only left temples were rated to align with the temple shown in the scale. Left temples were selected for the scale because the left side usually looks worse than the right side because of sun exposure while driving.

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 temple area; any treatment with toxin/fillers, dental procedures, or surgery that would alter the temple area 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. Three-dimensional 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 separately^{16–22}). 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 in temple volume deficit, 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.²³ 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.²⁴ 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.²⁵ The a priori primary endpoint 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.²⁶ 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 298 subjects were eligible for validation of the temple hollowing scale, the number of subjects evaluated using each scale was substantially larger than the preplanned sample size of 80, and the overall number of assessments for some grades of this scale were 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 80 subjects with minimum imbalance across the grades (~16 subjects

per each of the 5 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 1.1 (0.94–1.26) for image pairs deemed clinically different and 0.67 (0.51–0.83) for image pairs deemed not clinically different (Table 2). The 95% CIs for the pairs deemed to be clinically different did not overlap with those of pairs deemed not clinically different, confirming that a 1-point difference in scores is clinically significant.

Live-Subject Scale Validation

Of the 298 subjects eligible for scale validation analysis, 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 (27%) 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 almost perfect (mean weighted kappa = 0.86) (Table 4). Interrater agreement was substantial

(ICC = 0.79) during the first rating session and almost perfect (ICC = 0.81) during the second rating session (primary endpoint) (Table 4).

Discussion

This study demonstrated substantial to almost perfect interrater and intrarater agreement for the Allergan Temple Hollowing Scale, indicating 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 volume deficit in the temple area. The scale's standardized ratings may be uniformly applied in day-to-day clinical practice and potentially in clinical trials, because of its validation in live subjects and use of both morphed and unaltered images.

The scale includes verbal descriptors for each grade and a facial diagram clearly defining the temple area to be assessed; 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 constant across scale grades. The inclusion of real-world images representing a diverse range of skin types across sexes and races is also important, because morphed images may not always translate clinically to the broad array of physical appearances or physical changes observed in the aging face. When the scale grades were created, convex temples were defined as the lower limit of temple hollowing so that the scale may be used with

TABLE 2. Difference in Scores for Image Pairs Deemed Clinically Different or Not Clinically Different Using the Allergan Temple Hollowing 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	158	1.1 (1.02)	0–4	0.94–1.26
Not clinically different pairs	82	0.67 (0.74)	0–3	0.51–0.83

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

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

Characteristic	Number (%) of Subjects N = 291
Sex, n (%)	
Female	195 (67.0)
Male	96 (33.0)
Age, yrs	
Median	48
Range (min, max)	(18–83)
Fitzpatrick skin type, n (%)	
I	23 (7.9)
II	61 (21.0)
III	79 (27.1)
IV	89 (30.6)
V	25 (8.6)
VI	14 (4.8)
Race, n (%)	
Caucasian	229 (78.7)
Hispanic or Latino	31 (10.7)
African American	15 (5.2)
Asian	14 (4.8)
American Indian or Alaska Native	1 (0.3)
Caucasian/Hispanic or Latino	1 (0.3)

Asian patients, who may exhibit or prefer a convexly shaped temple area.¹⁰

Clinician ratings of attractiveness may vary significantly from those of patients, because clinicians may or may not be as critical as patients.²⁷ However, in the authors' experience, aesthetically concerned subjects have positive perceptions of a 1-point change in appearance after filler treatment of the temples, because it not only rejuvenates the proportions of the upper face to make the eyes seem wider, but helps

TABLE 4. Physician Intrarater and Interrater Agreement on the Allergan Temple Hollowing Scale (Validation Testing With Live Subjects)

Intrarater agreement	
Mean weighted kappa (95% CI)	0.86 (0.790–0.928)
Interrater agreement	
Round 1, ICC (95% CI)	0.79 (0.720–0.855)
Round 2,* ICC (95% CI)	0.81 (0.758–0.871)
*Primary endpoint.	
CI, confidence interval; ICC, intraclass correlation coefficient.	

elevate the tail of the eyebrow. Cultural differences may further widen the gap between the clinician's and the patient's perceptions of aesthetic ideals.^{7–9} Use of a validated scale for formalized and reproducible consultation procedures may allow for more informed treatment decisions²⁸ and potentially lead to overall improvements in patient satisfaction.

Study Limitations

The clinical significance of temple hollowing scale scores was determined solely by the scale developers. Although a 1-point change on this scale was considered significant to the scale developers, it may or may not be meaningful for 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 use of a validated patient satisfaction instrument, such as the FACE-Q, may be helpful for capturing the patient's perspective on appearance before and after treatment.²⁹ The verbal descriptors for each grade on the scale are subjective; however, the descriptors were developed and refined by extensive feedback among 9 experts, minimizing inherent subjectivity.

Conclusion

The Allergan Temple Hollowing Scale demonstrated almost perfect interrater and intrarater agreement among physicians, and 1-point score differences were shown to reflect clinically significant differences in temple volume deficit. This unique scale includes user-friendly diagrams, 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 that treat men and women seeking enhancement of the temple.

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

- Coleman SR, Grover R. The anatomy of the aging face: volume loss and changes in 3-dimensional topography. *Aesthet Surg J* 2006;26(1 Suppl):S4-S9.
- Lambros V. Observations on periorbital and midface aging. *Plast Reconstr Surg* 2007;120:1367-76.
- Szczerkowska-Dobosz A, Olszewska B, Lemanska M, Purzycka-Bohdan D, et al. Acquired facial lipoatrophy: pathogenesis and therapeutic options. *Postepy Dermatol Alergol* 2015;32:127-33.
- Pavicic T, Ruzicka T, Korting HC, Gauglitz G. Monophasic, cohesive-polydensified-matrix crosslinking-technology-based hyaluronic acid filler for the treatment of facial lipoatrophy in HIV-infected patients. *J Drugs Dermatol* 2010;9:690-5.
- Moradi A, Shirazi A, Perez V. A guide to temporal fossa augmentation with small gel particle hyaluronic acid dermal filler. *J Drugs Dermatol* 2011;10:673-6.
- Rose AE, Day D. Esthetic rejuvenation of the temple. *Clin Plast Surg* 2013;40:77-89.
- Rowe-Jones JM. Facial aesthetic surgical goals in patients of different cultures. *Facial Plast Surg Clin North Am* 2014;22:343-8.
- Weeks DM, Thomas JR. Beauty in a multicultural world. *Facial Plast Surg Clin North Am* 2014;22:337-41.
- Broer PN, Juran S, Liu YJ, Weichman K, et al. The impact of geographic, ethnic, and demographic dynamics on the perception of beauty. *J Craniofac Surg* 2014;25:e157-e61.
- Liew S, Wu WT, Chan HH, Ho WW, et al. Consensus on changing trends, attitudes, and concepts of Asian beauty. *Aesthet Plast Surg* 2016;40:193-201.
- Breithaupt AD, Jones DH, Braz A, Narins R, et al. Anatomical basis for safe and effective volumization of the temple. *Dermatol Surg* 2015;41(Suppl 1):S278-S83.
- Sykes JM, Cotozana S, Trevidic P, Solish N, et al. Upper face: clinical anatomy and regional approaches with injectable fillers. *Plast Reconstr Surg* 2015;136:204S-18S.
- Lambros V. A technique for filling the temples with highly diluted hyaluronic acid: the "dilution solution". *Aesthet Surg J* 2011;31:89-94.
- Ross JJ, Malhotra R. Orbitofacial rejuvenation of temple hollowing with Perlane injectable filler. *Aesthet Surg J* 2010;30:428-33.
- Moradi A, Shirazi A, Moradi J. A 12-month, prospective, evaluator-blinded study of small gel particle hyaluronic acid filler in the correction of temporal fossa volume loss. *J Drugs Dermatol* 2013;12:470-7.
- 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.
- 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.
- Donofrio L, Carruthers A, Hardas B, Murphy DK, et al. Development and validation of a photonumeric 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 photonumeric 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 photonumeric 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 photonumeric 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 photonumeric 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.
- Torsello F, Graci M, Grande NM, Deli R. Relationships between facial features in the perception of profile attractiveness. *Prog Orthod* 2010;11:92-7.
- 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.
- Klassen AF, Cano SJ, Schwitzer JA, Scott AM, et al. FACE-Q scales for health-related quality of life, early life impact, satisfaction with outcomes, and decision to have treatment: development and validation. *Plast Reconstr Surg* 2015;135:375-86.

Address correspondence and reprint requests to: Jean Carruthers, MD, Carruthers Clinical Research, 943 West Broadway, Suite 820, Vancouver, BC, Canada V5Z 4E1, or e-mail: drjean@carruthers.net