Validating the Reliability and Clinical Relevance of an Infraorbital Hollow Photonumeric Scale

Z. Paul Lorenc, MD, FACS; Stacy Smith, MD; Lawrence S. Bass, MD, FACS;
David Bank, MD; Robert Weiss, MD; Doug Canfield, BS;
Brian M. D'Alessandro, PhD; and Lisa M. Cramer, BA

OXFORD
UNIVERSITY PRESS

Abstract

Background: Infraorbital hollowing is a facial aesthetic issue for which a broad age range of patients seek treatment. Expanding treatment options for this region warrants the development of validated tools to objectively assess infraorbital hollow (IOH) severity before and after treatment.

Objectives: To validate a 4-point rating scale to assess depression of IOH, depression relative to the mid-pupillary line, and visibility of the lateral orbital rim.

Methods: The IOH scale described herein was developed and subjected to live validation with a total of 73 patients representing the full range of IOH severities. Scale validation was performed by board-certified plastic surgeons and dermatologists (3 raters) over 2 rounds, 2 weeks apart. Intrarater and interrater reliabilities were used to demonstrate test–retest reliability as quantitated with percentage of agreement, weighted kappa statistic with 95% confidence interval (CI), and intraclass correlation coefficient with 95% CI. The clinical relevance of a 1-grade difference was evaluated by comparing rater assessments of 77 photo pairs with their previously determined designation as "clinically different" or "not clinically different."

Results: The IOH scale demonstrated substantial to near-perfect intrarater and interrater reliabilities when utilized by trained raters to assess a diverse group of live patients. Furthermore, clinically relevant differences between grades were established, and detection of a 1-point difference could be achieved by trained evaluators using the IOH scale.

Conclusions: This highly reliable, clinically relevant, and validated IOH scale provides a user-friendly, standardized grading system to objectively evaluate and track changes in infraorbital hollowing in clinical practice and research.

Level of Evidence: 3

3

Diagnostic

Editorial Decision date: June 5, 2023; online publish-ahead-of-print August 9, 2023.

Dr Lorenc is a plastic surgeon in private practice in New York, NY, USA; and a clinical editor for *Aesthetic Surgery Journal*. Dr Smith is president, California Dermatology and Clinical Research Institute, Encinitas, CA, USA. Dr Bass is a clinical assistant professor of plastic surgery, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA. Dr Bank is a dermatologist in private practice in Mt Kisco, NY, USA. Dr Weiss is an adjunct assistant professor, Department of Dermatology, University of Maryland, Baltimore, MD, USA. Mr Canfield is president, Dr D'Alessandro is principal scientist, and Ms Cramer is director, Canfield Scientific, Inc., Parsippany, NJ, USA.

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

The periorbital area has a prominent role in communication and perception of beauty, and is also one of the first areas of the face where aging is apparent, making it a common area for facial rejuvenation among aesthetic patients.[1-3] Within the periorbital region, infraorbital hollowing is one of the most bothersome facial defects for which patients seek aesthetic treatment in both younger and older populations.[2] Even subtle shadowing and hollowing in the infraorbital region can convey the appearance of sadness or fatigue, prompting concerned patients to present in clinical practice seeking a more evenly contoured, vibrant facial look.[2-4]

The tear trough is the natural depression that extends inferolaterally from the medial canthus toward the mid-pupillary line and the palpebromalar groove, which extends around the lateral half of the inferior orbit.[5,6] Prominent infraorbital hollow (IOH) is primarily caused by tethering or loss of tensile strength of the tear trough ligament, which connects the medial suborbital skin to the maxilla.[7] The appearance of the resulting concave groove is augmented by contrasting tissue quality and quantity above and below it.[7] Above the IOH, subcutaneous fat is absent, and the skin is thinner, softer, and may be relatively darker, whereas, below the IOH, subcutaneous fat is covered by thicker skin.[7] Although not exclusively related to aging and sometimes present in younger patients due to facial anatomy, the IOH becomes more pronounced with time as it extends laterally with the palpebromalar groove, eventually leading to a visible lateral orbital rim.[7-9] Age-related changes contributing to increasing IOH prominence include orbital fat bulging above the ligament opposite loss of mid-facial fat pad volume, tissue descent, maxillary retrusion, ligament laxity, and atrophy below the ligament.[7,9]

While surgical interventions can correct IOH, noninvasive techniques have been used increasingly to treat the defect, yielding a high rate of patient satisfaction in a region where noticeable improvement can arise from minimally invasive treatment, and injection with fillers is relatively durable.[9,10] Expanding treatment options for this region, as well as the dramatic impact of treatment and the need to characterize and/or compare efficacy and durability, warrant the development of validated tools to objectively assess IOH severity before and after treatment. To meet this demand, a 4-point scale assessing IOH was developed and validated by establishing its high test–retest reliability and clinical relevance (Table 1).

METHODS

The IOH scale is a 4-point rating scale to assess the depression of the IOH, depression in relation to the mid-pupillary line, and visibility of the lateral orbital rim. The scale was developed by a team consisting of a board-certified plastic surgeon and a board-certified dermatologist ("developers"), and separately validated using live participants and

Table 1. Descriptors for Infraorbital Hollow Scale

Grade	Term	Descriptor
0	None/minimal	• None to minimal depression of the infraorbital area
1	Mild	• Mild depression of the infraorbital hollow • Does not extend past the mid-pupillary line
2	Moderate	• Moderate depression of the infraorbital hollow • May extend past the mid-pupillary line • Lateral orbital rim not visible
3	Severe	• Severe depression of the infraorbital hollow • Extends past the mid-pupillary line • Visible lateral orbital rim

a separate group of clinicians ("raters") consisting of 2 board-certified dermatologists and 1 plastic surgeon. Raters were trained by grading sample images and achieving consensus rating live patients using the IOH scale prior to validation.

Scale Development

A total of 126 adult males and females representing various races, ethnicities, and Fitzpatrick skin types consented to and participated in image collection to build the scale. While IRB approval was not required for this study, the consent forms signed and the process of image collection conformed to the principles set forth in the Declaration of Helsinki. Participants wore no makeup and/or jewelry and were clean-shaven during image capture. Scale developers independently reviewed right and left oblique and frontal images, 1 patient at a time, and scored the severity of IOHs by designating as none/minimal, mild, moderate, and severe in the absence of descriptors.

The completed scale consists of 3 components: (1) textual descriptors, (2) morphed images, and (3) actual patient images, representing different sexes and Fitzpatrick skin types (Figure). Textual descriptors for each grade were composed by the scale developers. Developers selected 2 representative images for each grade to form a diverse set of actual patient images for the scale.

The morphed images were developed through artificial intelligence (AI) facial averaging technology.[11] The goal of this component was to model grade-based differences across the 4 points of the scale through statistical model-based "morphing" of a single individual "base image."

Morphing models were developed from participant images. Briefly, images were bucketed according to their respective grades. Next, all images were annotated with anatomical landmarks and then aligned to a common base image of Grade 0 using thin plate spline warping. Statistical color and topography models were then built using the IOH grade of each image as the independent variable.[11-13] The completed models allowed for accurate
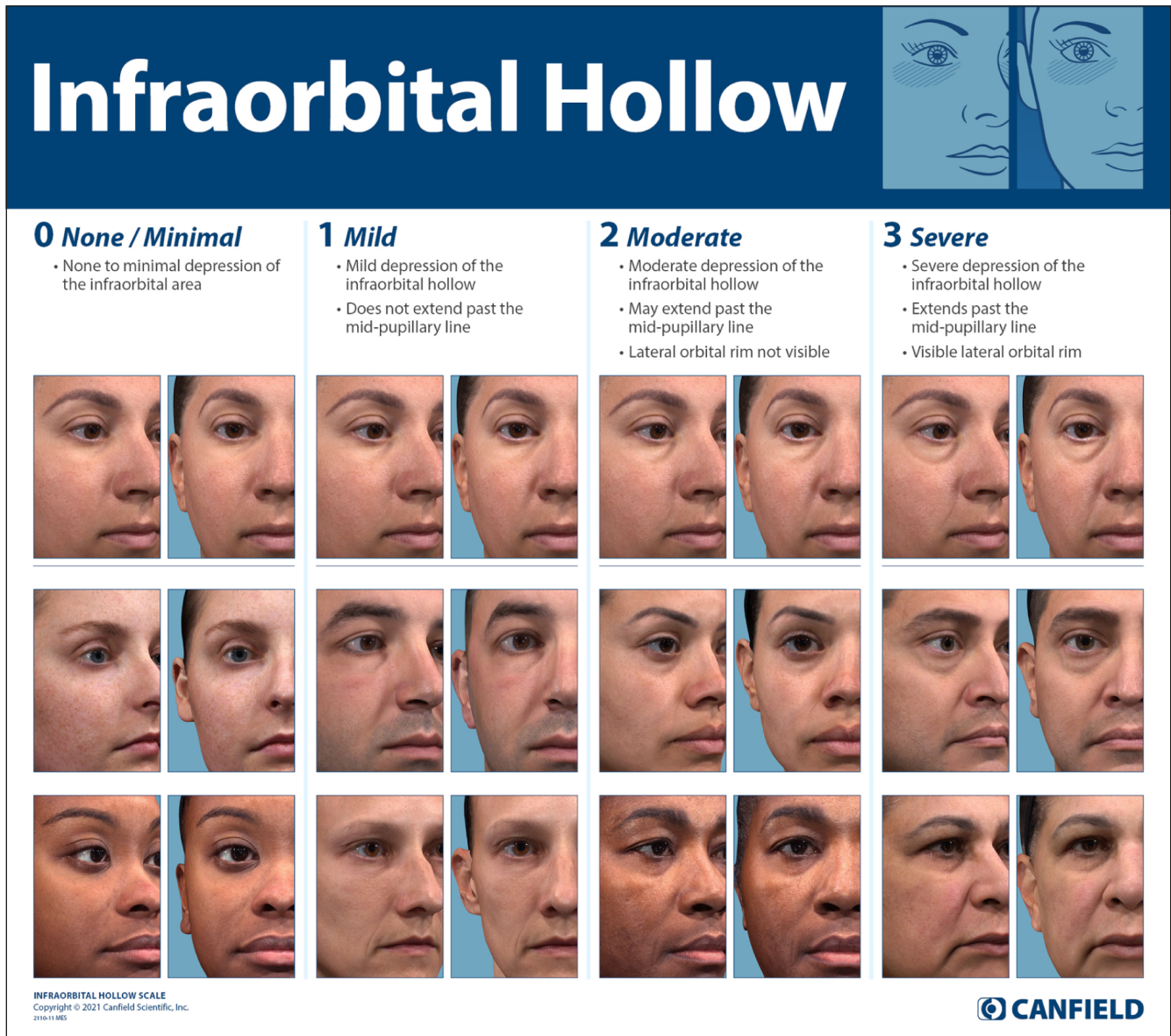
Figure. The Infraorbital Hollow Scale illustrates each severity grade with 3 sets of vivid images framed as cropped right and left oblique and corresponding detailed descriptions. The top line photographs were morphed from a base image to represent each grade using facial averaging, whereas the rest of the scale was populated with unmorphed, actual patient images selected for each grade of IOH severity. This scale is owned and licensed for use by Canfield Scientific, Parsippany, NJ, USA.

image-based prediction of facial appearance for a desired target grade on the 4-point IOH scale. Note that the models were not just limited to the IOH region but predict the full face including any other facial feature correlated with IOH and accounted for every pixel in the image given. Relative difference of the predictions for Grade 0 and each of the Grades 1 through 3 was applied to the base image of Grade 0 to produce simulated, morphed images for Grades 1, 2, and 3, along with unmorphed images of patients across all gradations. Morphed images maintain the identity of the base participant but include the appearance

of the higher IOH grades based entirely on the real graded data set and the statistical models they produced. The result is a realistic, data-based IOH scale on a single base individual which serves as a guide, alongside actual images of real patients in each grade of IOH.

Scale Validation

Trained raters performed live validation of the IOH scale over 2 live validation rounds, conducted 2 weeks apart. A total of 73 patients were selected to represent the full range of

Table 2. Demographics for Live-Patient Scale Validation

Characteristic	N = 73
Age (years)	
Mean (SD)	44.6 (15.1)
Median	45
Minimum, maximum	18, 71
Height (cm)	
Mean (SD)	166.7 (8.7)
Median	165.1
Minimum, maximum	150, 188
Weight (kg)	
Mean (SD)	69.1 (14.0)
Median	67.6
Minimum, maximum	46, 104
Sex, *n* (%)	
Male	20 (27.4)
Female	53 (72.6)
Ethnicity, *n* (%)	
Hispanic or Latino	5 (6.9)
Not Hispanic or Latino	68 (93.2)
Race, *n* (%)	
American Indian or Alaska Native	1 (1.4)
Asian Indian	2 (2.7)
Black or African American	5 (6.9)
Asian	5 (6.9)
Native Hawaiian or Other Pacific Islander	1 (1.4)
White	59 (80.8)
Fitzpatrick skin type, *n* (%)	
I	0 (0)
II	11 (15.1)
III	34 (46.6)
IV	23 (31.5)
V	1 (1.4)
VI	4 (5.5)

SD, standard deviation.

different depression levels of the IOH. Participants were instructed to arrive to the sessions clean-shaven, without make-up and/or jewelry, and maintain their usual routine (eg, facial care, sleep, and hydration routines), abstaining from tanning sessions or extensive sun exposure between sessions. During the live validation session, participants presented themselves at a rating station where a scale validator used a printed copy of the photo numeric scale to assign an integer rating of 0 to 3 to each patient for right and left IOHs, separately, and recorded the score through electronic data capture. Each scale validator proceeded from 1 rating station to the next until each participant had been evaluated by all 3 scale validators. Patients were asked to maintain the same self-care routines and avoid facial treatments or environmental exposures that could change their facial appearance in the intervening period. These same patients were assessed at the second round, but in a different random sequence.

Evaluation of Clinical Relevance

To determine if clinically relevant differences between grades on the IOH scale can be detected when utilized by trained raters, a set of photographs was selected to represent all grades on the scale. Rating scores were determined by the majority of scores assigned during scale development. The photographs were then used to generate 77 photo pairs covering all scale grades that might be considered "not clinically different" (29 pairs) or "clinically different" (30 pairs with 1-point difference, 12 pairs with 2-point difference, and 6 pairs with 3-point difference).

The 3 scale validators were asked to perform a side-by-side evaluation of the photo pairs. During the session, the raters were presented with all 77 photo pairs and asked whether there was a clinically significant difference in the IOH depression of the 2 patients. After the session, each validator used the scale to assign a score to randomly sequenced individual photographs from all photo pairs.

Data Analysis and Statistical Methods

All statistical analyses were performed using SAS Version 9.4 (SAS Institute; Cary, NC). Test–retest reliability was quantitated through measuring intrarater and interrater reliabilities. Intrarater reliability between Rounds 1 and 2 was evaluated for each rater, median of all raters, and all raters combined by calculating the percentage of agreement (exact and ≥1-grade difference), weighted kappa statistic with 95% confidence interval (CI), and intraclass correlation coefficient (ICC) with 95% CI. Interrater reliability was determined for each pair of raters and for each rater against the median score of all 3 raters using the same 3 calculations. Weighted kappa statistics and ICC were calculated using established methods, where >0 and ≤0.2 indicate slight agreement, >0.2 and ≤0.4 indicate fair agreement, >0.4 and ≤0.6 indicate moderate agreement, >0.6 and ≤0.8 indicate substantial agreement, and >0.8 and ≤1.0 indicate almost perfect agreement.[14-16]

Table 3. Intrarater Reliability

Round 1 vs Round 2	Percentage exact agreement	Percentage within 1 grade	Weighted kappa coefficient (95% CI)	ICC (95% CI)
Right face				
Rater 1	68.5	98.6	0.840 (0.769, 0.910)	0.841 (0.759, 0.897)
Rater 2	75.3	100	0.897 (0.850, 0.943)	0.898 (0.842, 0.935)
Rater 3	65.8	100	0.793 (0.718, 0.869)	0.796 (0.689, 0.868)
Combined	—	—	0.849 (0.812, 0.886)	0.923 (0.880, 0.951)
Left face				
Rater 1	67.1	100	0.842 (0.783, 0.902)	0.844 (0.763, 0.899)
Rater 2	79.5	100	0.889 (0.832, 0.946)	0.890 (0.826, 0.931)
Rater 3	71.2	100	0.812 (0.734, 0.890)	0.814 (0.713, 0.881)
Combined	—	—	0.850 (0.813, 0.887)	0.920 (0.875, 0.949)

CI, confidence interval; ICC, intraclass correlation coefficient.

Clinical relevance was evaluated via absolute differences in rating scores between each paired photograph. The absolute differences were calculated from the actual ratings assigned by the 3 independent raters and were summarized using descriptive statistics for the photo pairs originally deemed as clinically different and not clinically different. The mean, standard deviation, and 95% CI of the mean were reported. Further analysis of clinical relevance determined the proportion of agreement between the rater's assessments vs the original assessments of the photo pairs, with agreement being defined as at least 2 out of 3 raters giving the same assessment (ie, clinically different vs not clinically different). The frequency counts and percentages of agreement were summarized for all photo pairs, clinically different photo pairs, and not clinically different photo pairs.

RESULTS

Live-Patient Scale Validation

Patients participating in the live validation population represented both sexes and a wide range of age, self-reported race and ethnicities, and clinical-rater-assessed Fitzpatrick skin types, with most participants identifying as white, non-Hispanic females (Table 2). The age range of evaluated patients represented most of the patient population, with a mean of middle age (44.6 years), representing patients frequently seen in practice. Fitzpatrick Skin Types III and IV were most prevalent in the evaluated population. Participants had a broad range of heights and weights, with the mean values being representative of a healthy body mass index.

The IOH scale intrarater reliability, agreement between the first rating and second rating 2 weeks later from the same rater (Rounds 1 and 2), was assessed for right and left IOHs of 73 patients. Results for Rounds 1 and 2 were evaluated through the percentage of exact matches, reproducibility within 1 grade, and weighted kappa coefficients (95% CI) for the 3 raters (Table 3). The ratings of the right face and left IOHs showed comparable results, with combined weighted kappa coefficients and ICC for the 3 raters indicating almost perfect agreement of 0.849 and 0.923 for the right side and 0.850 and 0.920 for the left side, respectively. The percentage of exact agreement of the assessed grade for a single image between the 2 rounds for each rater ranged from 66% to 80%, whereas the percentage of assigned grades within 1 grade was nearly 100% for each of the raters.

Similarly, interrater analysis (agreement between raters) for evaluation of right and left sides of the IOH using the IOH scale indicated substantial to almost perfect agreement. The weighted kappa coefficients across all rater pairs for the right and left face ranged from 0.746 to 0.785 for Round 1 and 0.750 to 0.829 for Round 2. Similarly, the ICC ranged from 0.748 to 0.787 for Round 1 and 0.752 to 0.831 for Round 2.

Clinical Relevance Determination

The absolute differences in scores assigned by 3 independent raters between pairs originally deemed clinically different vs not clinically different for photo-pair selections were evaluated (Table 4). The mean absolute difference in scores between clinically different photo pairs was over 1 grade (median [95% CI]; 1.17 [1.02, 1.31]), whereas the mean difference for not clinically different photo pairs was less than half a grade (median [95% CI]; 0.48 [0.36, 0.60]). Furthermore, the 95% CIs for clinically different vs

Table 4. Evaluation of Clinical Relevance

Assessment	Absolute difference in scores between paired photographs			
Original assessments used for photo-pair selections	n	Mean (SD)	Minimum, maximum	95% CI of mean
Clinically different pairs	144	1.17 (0.88)	0, 3	1.02, 1.31
Not clinically different pairs	87	0.48 (0.57)	0, 2	0.36, 0.60

CI, confidence interval; SD, standard deviation.

not clinically different pairs do not overlap. Taken together, these results suggest that the raters were able to accurately rate photographs by utilizing the IOH scale, and a 1-point difference on the IOH scale is clinically relevant.

Evaluating the proportion of agreement of photo-pair assessments revealed that the raters gave the same assessments for the photo pairs when compared against original assessments (ie, clinically different vs not clinically different). Relative to the original assessments, at least 2 out of 3 raters assigned the same assessments for 64.9% (50 of 77) of the total photo pairs, 75% (36 of 48) of the clinically different pairs, and 48.3% (14 of 29) of the not clinically different pairs. The high proportion of agreement indicates that clinically relevant differences on the IOH scale can be detected by raters when evaluating random side-by-side photographs with ≥1 grade difference.

DISCUSSION

Live validation of the IOH scale was performed using board-certified plastic surgeons and dermatologists with high test–retest reliability as demonstrated by substantial to almost perfect values for weighted kappa coefficients and ICC values as measures of intrarater and interrater reliabilities. Comparable results were observed for the right and left IOH when evaluated separately by raters. The high intrarater and interrater reliabilities indicate that the IOH scale can be used for dependable evaluation by the same rater multiple times and for individual raters at different times, respectively. Furthermore, raters were able to accurately rate photographs by utilizing the IOH scale, demonstrating clinically relevant differences between grades and that a 1-point difference can be detected when utilized by trained evaluators.

Evidence-based grading systems are necessary to objectively assess aesthetic defects to aid in planning corrective procedures and evaluating outcomes.[17] Utilizing a validated scale may lessen subjective perceptions of success for both the physician and patient while providing patients with realistic expectations of nonsurgical aesthetic procedures.[18,19] Validation of the presented IOH scale supports its implementation in clinical practice for

preprocedure assessment and evaluation of outcomes. Although the IOH scale reported herein is not the first published scale rating infraorbital hollowing,[20-23] the scale is unique in that it was developed and validated with live grading for the sole purpose of providing a scale with proven reliability for broad industry access. Two commercially published IOH photonumeric scales were designed with a similar approach and also validated in a live, diverse patient population.[20,21] However, the scale presented herein has added merit as it captures IOH at different angles with ample lighting and a bright background to avoid shadowing, resulting in vibrant imagery and possibly contributing to its slightly higher intrarater and interrater reliabilities scores relative to these previously published scales. The features of the presented scale give enhanced visual context to the IOH to facilitate more accurate grading of the live patient. The proven high reliability, user-friendly design, and suitability to assess real-life populations intrinsic to this scale will likely benefit the facial rejuvenation field and prove useful for preprocedure and postprocedure evaluation of infraorbital hollowing.

Patient participating in the live validation population represented both sexes and a wide range of age, self-reported race and ethnicities, and clinical-rater-assessed Fitzpatrick skin types. The reported high test–retest reliability of the IOH scale may be attributed to its representation of multiple Fitzpatrick skin types of both sexes at each severity grade, giving raters several visuals on which to base their assessments. Additionally, each severity grade is illustrated with 3 sets of vivid images, framed as cropped right oblique and frontal, and corresponding detailed descriptions, effectively representing the IOH characteristics typical of each grade. As previously mentioned, the top line-scale photographs were morphed from a base image to represent each grade using statistical facial averaging models, whereas most of the scale was populated with unmorphed, actual patient images selected for each grade of IOH severity. This strategy allows the evaluator to isolate the IOH changes that occur with each progressing grade in the morphed image while referencing multiple real-world images of patients representing each grade.

Although multiple Fitzpatrick skin types are represented in the IOH scale, one possible limitation is that facial features from White patients are most often depicted. Future studies could be strengthened by including greater proportions of participant from different racial backgrounds. Additionally, the number of reviewers selected for this study was based on the number of patients: with 3 raters and 73 patients, the 95% CI is predicted to be 0.1 (when the expected ICC is 0.7**) or 0.2 (when the expected ICC is 0.6). However, we do recognize that the validity of study outcomes is affected not only by the number of raters, but also by their qualification and training. In this study, the raters were both experienced clinicians, well-practiced in

using scales for clinical research, and were trained regarding the validation process and its implementation. While future studies may be further strengthened by including additional raters, we believe this study to be of sufficient strength to warrant use of the presented scale in clinical studies and/or clinical practice. Finally, within this study, confirmation of the clinical relevance of a 1-grade difference was determined by the reviewers by comparing rater assessments of photo pairs with previously determined designations of pairs as clinically different or not clinically different; however, any future clinical studies in which 1-grade differences on this scale are achieved through aesthetic interventions performed on patients and are achieved alongside other validated and/or commonly used patient and physician measures, such as the Global Aesthetic Improvement Scale or will be of interest and informative for confirming that 1-point changes are associated with global improvement, satisfaction, or other outcomes.

CONCLUSIONS

The presented IOH scale demonstrated substantial to near-perfect intrarater and interrater agreement among board-certified physicians and clinically relevant differences between each scale grade. By using both morphed and real-world images representing males and females with diverse Fitzpatrick skin types, the IOH scale facilitates accurate patient assessment by allowing the evaluator to reference isolated progressive IOH changes. Clear and vibrant photographs showcasing 2 views of the IOH defect with corresponding text descriptions enhance the scale's utility. The resulting clinically relevant, easy-to-use, highly reliable and validated IOH scale provides a standardized grading system for aesthetics clinicians to objectively assess infraorbital hollowing in clinical practice.

REFERENCES

1. DeFatta RJ, Williams EF. Evolution of midface rejuvenation. *Arch Facial Plast Surg*. 2009;11(1):6-12. doi: 10.1001/archfaci.11.1.6
2. Narurkar V, Shamban A, Sissins P, Stonehouse A, Gallagher C. Facial treatment preferences in aesthetically aware women. *Dermatol Surg*. 2015;41(Supplement 1): S153-S160. doi: 10.1097/DSS.0000000000000293
3. Siperstein R. Infraorbital hyaluronic acid filler: common aesthetic side effects with treatment and prevention options. *Aesthet Surg J Open Forum*. 2022;4:ojac001. doi: 10.1093/asjof/ojac001
4. Michaud T, Gassia V, Belhaouari L. Facial dynamics and emotional expressions in facial aging treatments. *J Cosmet Dermatol*. 2015;14(1):9-21. doi: 10.1111/jocd.12128
5. Park KY, Kwon HJ, Youn CS, Seo SJ, Kim MN. Treatments of infra-orbital dark circles by various etiologies. *Ann Dermatol*. 2018;30(5):522-528. doi: 10.5021/ad.2018.30.5.522
6. Luo S, Zhang X, Dong H, Wen C, Hao L. Correction of the tear trough deformity and concomitant infraorbital hollows with extracellular matrix/stromal vascular fraction gel. *Dermatol Surg*. 2020;46(12):e118-e125. doi: 10.1097/DSS.0000000000002359
7. Wong CH, Hsieh MKH, Mendelson B. The tear trough ligament: anatomical basis for the tear trough deformity. *Plast Reconstr Surg*. 2012;129(6):1392-1402. doi: 10.1097/PRS.0b013e31824ecd77
8. Yang C, Zhang P, Xing X. Tear trough and palpebromalar groove in young versus elderly adults: a sectional anatomy study. *Plast Reconstr Surg*. 2013;132(4):796-808. doi: 10.1097/PRS.0b0133182a0539e
9. Fabi S, Zoumalan C, Fagien S, Yoelin S, Sartor M, Chawla S. A prospective, multicenter, single-blind, randomized, controlled study of VYC-15L, a hyaluronic acid filler, in adults for correction of infraorbital hollowing. *Aesthet Surg J*. 2021;41(11):NP1675-NP1685. doi: 10.1093/asj/sjab308
10. Sharad J. Treatment of the tear trough and infraorbital hollow with hyaluronic acid fillers using both needle and cannula. *Dermatol Ther*. 2020;33(3):e13353. doi: 10.1111/dth.13353

11. Matts PJ, Canfield D, D'Alessandro B. A new model to simulate human facial appearance, accurately, and realistically, across age and ethnicity. *Plast Reconstr Surg*. 2021; 148(6S):14S-20S. doi: 10.1097/PRS.0000000000008781

12. D'Alessandro BM, Matts PJ. Methods and apparatuses for age appearance simulation. US patent 290257B2. October 16, 2012.

13. Matts PJ, D'Alessandro BM. Methods for age appearance simulation. Japanese Patent JP6985403B2. December 22, 2021.

14. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas*. 1973;33(3):613-619. doi: 10.1177/001316447303300309

15. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428. doi: 10.1037/0033-2909.86.2.420

16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174. doi: 10.2307/2529310

17. Gupta S, Biskup N, Mattison G, Leis A. Development and validation of a clinical assessment tool for platysmal banding in cervicomental aesthetics of the female neck. *Aesthet Surg J*. 2015;35(6):NP141-NP146. doi: 10.1093/asj/sju160

18. Lemperle G, Holmes RE, Cohen SR, Lemperle SM. A classification of facial wrinkles. *Plast Reconstr Surg*. 2001; 108(6):1735-1750; discussion 1751-1752. doi: 10.1097/00006534-200111000-00048

19. Jandhyala R. Improving consent procedures and evaluation of treatment success in cosmetic use of incobotulinumtoxin A: an assessment of the treat-to-goal approach. *J Drugs Dermatol*. 2013;12(1):72-78.

20. Moradi A, Lin X, Allen S, Fagien S, Norberg M, Smith S. Validation of photonumeric assessment scales for temple volume deficit, infraorbital hollows, and chin retrusion. *Dermatol Surg*. 2020;46(9):1148-1154. doi: 10.1097/DSS.0000000000002269

21. Donofrio L, Carruthers J, Hardas B, et al. Development and validation of a photonumeric scale for evaluation of infraorbital hollows. *Dermatol Surg*. 2016;42(Supplement 1): S251-S258. doi: 10.1097/DSS.0000000000000856

22. Carruthers J, Flynn TC, Geister TL, et al. Validated assessment scales for the mid face. *Dermatol Surg*. 2012;38(2 Spec No.):320-332. doi: 10.1111/j.1524-4725.2011.02251.x

23. Pavicic T, Pooth R, Prinz V, et al. Validated 5-point photonumeric scales for the assessment of the periorbital region. *J Cosmet Dermatol*. 2022;21(1):158-166. doi: 10.1111/jocd.14643