



Validation of a Chin Retrusion Scale for Chinese Subjects

An Junxue, MD,* Lujia Chen, MD,† Ma Xiaobing, MD,‡ Jiazhi Qu, MD,§
Andrew Schumacher, PhD,|| and Li Dong, MD, PhD¶

Abstract: Chin augmentation procedures are gaining in popularity. The purpose of this study was to validate the China (Allergan) Chin Retrusion Scale (CACRS) and to evaluate the reliability of the scale. A team of 10 physicians based in Beijing, China, consisting of 1 principal investigator and 9 independent raters, assessed a pool of subject images. Using standardized equipment to capture 2-dimensional images, 100 subjects were photographed from a left, lateral, 90° view. Two measures were used to assess the validity of the scale: rater evaluations of clinically significant differences in 52 pairs of images and rater assessment of chin retrusion using the CACRS to grade the 100 images. The CACRS demonstrated almost perfect inter-rater agreement during 2 validation sessions, with intraclass correlation coefficients of 0.89 and 0.90 at sessions 1 and 2, respectively. Overall inter-rater agreement for all 9 raters was almost perfect, with a weighted kappa of 0.82. Based on inter- and intra-rater agreement, the CACRS was validated for physician ratings of chin retrusion in Chinese subjects. Results from clinical significance evaluations indicated that a 1-grade difference between image pairs was considered clinically meaningful for the CACRS. The CACRS is a validated and reliable photonic scale for the assessment of chin retrusion in Chinese men and women. The scale is a clinically significant reference tool for evaluating the effect of chin augmentation.

Key Words: Aesthetics, China, dermal fillers, treatment outcome, validation studies

(*J Craniofac Surg* 2022;33: 48–51)

From the *Peking University International Hospital; †Tsinghua University Affiliated Beijing Tsinghua Changgung Hospital; ‡Beijing Hospital; §Allergan Aesthetics, an AbbVie Company, Beijing, China; ||Allergan Aesthetics, an AbbVie Company, Irvine, CA; and ¶Beijing University Third Hospital, Beijing Shi, China.

Received March 25, 2021.

Accepted for publication April 20, 2021.

Address correspondence and reprint requests to Li Dong, MD, PhD, Beijing University Third Hospital, 49 Huayuan N. Road, Haidian Qu, Beijing, China; E-mail: lidong9@sina.com

This study was sponsored by Allergan Aesthetics, an AbbVie Company, Dublin, Ireland.

JQ was an employee of Allergan Aesthetics, an AbbVie Company, at the time the study was conducted. A. Schumacher is an employee of Allergan Aesthetics, an AbbVie Company. The remaining authors report no conflicts of interest.

Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jcraniofacialsurgery.com).

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of Mutaz B. Habal, MD.

ISSN: 1049-2275

DOI: 10.1097/SCS.00000000000007849

An important determinant of perceived attractiveness in different countries and cultures is the harmonic proportion between the chin and other facial areas.^{1–6} In China, chin projection is a common aesthetic concern because individuals generally consider round, narrow, and pointy chins more favorable than flat and wide chins.^{3,7,8} Individuals may therefore undergo chin augmentation to improve their appearances.^{7,9} This procedure can be performed surgically (eg, insertion of chin implants, chin advancement via osteotomy of the bony mentum) or less invasively using injectable soft-tissue filler treatments, which have become increasingly popular in many geographic regions, including Asia.^{1,7}

To assess chin prominence, Allergan previously developed and validated the photonic Allergan Chin Retrusion Scale (ACRS).¹⁰ The ACRS features images of Western men and women and uses anthropometric landmarks of the lower face to derive objective estimates of the degree of retrusion. The ACRS displayed substantial inter- and intra-rater agreement when tested in a live-subject validation study.¹⁰ Although valid and reliable for Western subjects, the ACRS is not suitable for use with Chinese individuals because the 2 ethnic groups have different craniofacial morphology characteristics (eg, mandibular plane angle) and distinct aesthetic goals (eg, ideal mandible-lip ratio).^{11–19} Furthermore, compared with people of European origin, Chinese individuals more often exhibit lip protrusion and chin retrusion.⁷ Thus, dermatologists and plastic surgeons who treat Chinese men and women seeking aesthetic improvements would be well served by a validated scale that can be uniformly applied to this population.⁷

The objectives of this study were to validate a Chinese version of the ACRS, the China Allergan Chin Retrusion Scale (CACRS), which was developed for measuring chin retrusion in Chinese subjects, and to determine whether there was a clinically significant difference between grades of the scale.

METHODS

Scale Validation Process

Similar to the ACRS,¹⁰ the CACRS has 5 categories, each of which is associated with a numeric score and a descriptive grade (Supplemental Digital Content, Table 1, <http://links.lww.com/SCS/C881>). Numeric scores of the CACRS range from 0 to 4, with 0 representing no retrusion and 4 representing severe retrusion. The CACRS contains morphed subject photographs for all 5 severities of chin retrusion; the base image is a photograph of a subject with grade 2 retrusion. In addition, the scale contains real example images for all 5 severities including both sexes. An English translation of the scale is shown in Figure 1. The Chinese language version of the CACRS was used to train raters and for validation assessments in this study (Supplemental Digital Content, Figure 1, <http://links.lww.com/SCS/C882>).

Figure 2 presents the lateral view of the anatomic markers of the lower face used to assess chin retrusion using the CACRS. Using this assessment guide, chin retrusion is judged by the location of the chin midpoint in relation to the cheilion, labrale inferius, and suprumentale.

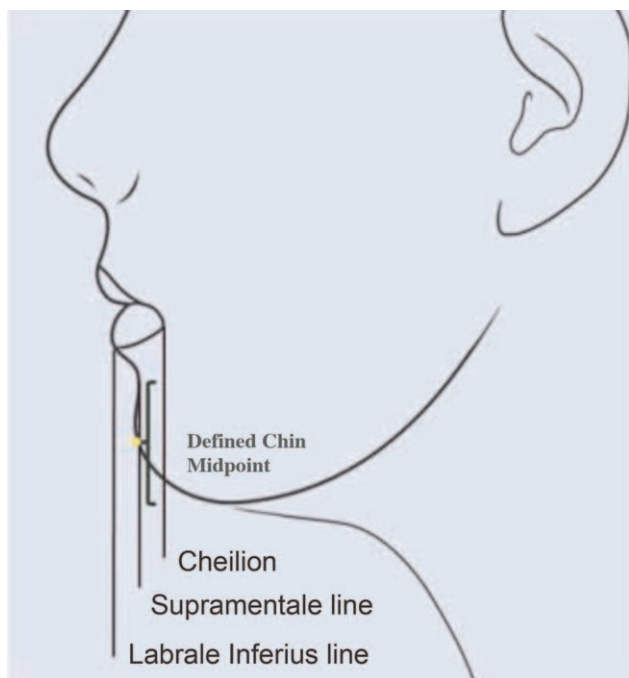


FIGURE 1. Assessment guide for the China (Allergan) Chin Retrusion Scale.

To validate the CACRS, a team of 10 physicians based in China, consisting of 1 principal investigator and 9 independent raters experienced in facial aesthetics, assessed a pool of subject images using image rating and electronic data capture tools provided by Canfield Scientific, Inc. (Canfield, Fairfield, NJ). Figure 3 depicts the key steps in the process of validating the CACRS. Two effectiveness measures were used for scale validation: rater evaluations of clinically significant differences in select image pairs and rater assessment of chin retrusion using the CACRS.

This study was conducted in conformance with the principles of the Declaration of Helsinki or the laws and regulations of the country in which the research was conducted, whichever afforded greater protection for the individual.

Subjects

A total of 300 subjects were enrolled to have their photographs taken, with images of 100 subjects planned for inclusion in the assessment. The subjects were required to be men and women of Han nationality from Beijing aged 18 to 50 years, in good general health, with varying degrees of chin retrusion. All subjects provided written informed consent. Subjects were excluded if they appeared in training images for the CACRS, had substantial congenital or traumatic cephalofacial deformity, scarring, or defects that interfere with visual assessment, or had a history of cephalofacial surgery.

Subject Images and Validation Sessions

Using standardized equipment to capture 2-dimensional images (ie, Canon PowerShot G16, 12.1 MP camera), subjects were photographed, without smiling, from a left, lateral, 90° view in natural light, with a relaxed, natural posture and with tragus points exposed and the Frankfort horizontal plane parallel to the floor.

Study personnel developed 2 randomization schemes for subject images to be used in 2 validation sessions that occurred at least 2 weeks apart. For validation session 1, 52 pairs of images were selected for the clinical significance assessment. The principal investigator reviewed and approved the image pairs



FIGURE 2. China (Allergan) Chin Retrusion Scale subject photographs (training images; English language version). The scale assigns a grade from 0 (no retrusion) to 4 (severe retrusion).

represented 1-, 2-, 3-, and 4-grade differences in retrusion scores on the CACRS and represented all levels of difference among CACRS grades. Validation session 1 took place in a group setting at a single location, and the clinical significance assessment was conducted before raters received training on the CACRS. For each of the 52

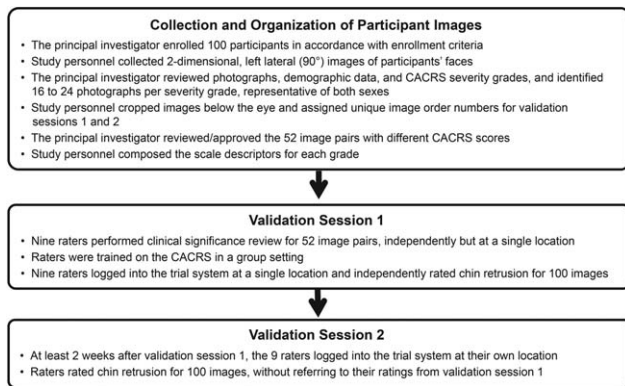


FIGURE 3. Validation process for the China (Allergan) Chin Retrusion Scale.

pairs of images assembled by the principal investigator, raters were asked to designate the images as either clinically significantly different or not different.

The raters then received training on how to assess chin retrusion using the CACRS and participated in group discussion and consensus-building. The raters next viewed the 100 images of the Chinese subjects, and independently graded the category of chin retrusion for each subject.

Validation session 2 took place in an individual setting at the rater's own location. Each rater logged into a secure website, viewed the 100 images of the Chinese subjects, and graded the category of chin retrusion for each subject.

Data Analysis and Statistics

The sample size was calculated using the method described by Bonett et al.²⁰ All statistical analyses were performed using SAS version 9.2 or higher (Cary, NC, USA). With at least 8 raters and an intraclass correlation coefficient (ICC) of 0.60, a total of 100 subjects provided 95% CIs with a width of at least 0.16 for interrater reliability. A minimum of 16 subject images and a maximum of 24 subject images were selected for each category of the CACRS.

The SAS macro "INTRACC" was used to determine interrater reliability by calculating the ICC and 95% CI for validation sessions 1 and 2.²¹ For each rater, the SAS procedure "FREQ" was used to determine intra-rater reliability by calculating a weighted kappa with Cicchetti-Allison weights (SAS default weights) and 95% CI.²² The overall mean weighted kappa was computed by averaging the overall weighted kappa for each rater from validation sessions 1 and 2.

The agreement was evaluated as defined by Landis and Koch.²³ Five degrees of agreement were characterized: slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.00).

To assess the clinical significance of paired images, the percentage of image pairs identified as clinically significantly different, or not different, were presented for all absolute differences in CACRS scores.

RESULTS

Subject Demographics

Nearly two-thirds of the population (62%) was female. The mean (SD) age of the subjects was 27.2 (7.8) years.

Scale Validation

The CACRS demonstrated almost perfect inter-rater agreement during sessions 1 and 2, with ICCs of 0.89 and 0.90, respectively

(Supplemental Digital Content, Table 2, <http://links.lww.com/SCS/C881>). Overall intra-rater agreement for all 9 raters was almost perfect, with a weighted kappa of 0.82 (Supplemental Digital Content, Table 2, <http://links.lww.com/SCS/C881>). CACRS agreement was almost perfect for 6 of 9 raters and substantial for the remaining 3 raters. The difference in weighted kappas between the highest- and lowest-performing raters was 0.21.

Clinical Significance

The 9 raters evaluated all 52 image pairs, for a total of 468 assessments (Supplemental Digital Content, Table 3, <http://links.lww.com/SCS/C881>). In total, 154, 210, 64, 27, and 13 assessments represented pairs with no difference in CACRS category score and pairs with an absolute difference of 0, 1, 2, 3, and 4, respectively. Additionally, 67% of pairs of images with an absolute difference score of 1 and 56% of pairs of images with an absolute difference score of 0 had a clinically significant difference rating (Supplemental Digital Content, Table 3, <http://links.lww.com/SCS/C881>). The percentages of raters who identified clinically significant differences between images increased with the absolute difference between pairs (92.2% for image pairs with an absolute difference score of 2, and 100% for pairs with an absolute difference of 3 or 4; Supplemental Digital Content, Table 3, <http://links.lww.com/SCS/C881>). Based on these data, a 1-grade difference was determined to be clinically significant.

DISCUSSION

Using the ACRS as a foundation, a chin retrusion scale, the CACRS, was developed for use by Chinese physicians and their patients.¹⁰ Both the ACRS and the CACRS instruments assess chin retrusion on a 5-grade scale from 0 (none) to 4 (extreme/severe). The CACRS displayed excellent inter- and intra-rater agreement among the 9 Chinese physicians who participated in this study. Results from clinical significance evaluations indicated that a 1-grade difference between image pairs can be considered clinically meaningful for the CACRS, as demonstrated previously with the ACRS.¹⁰

However, a relatively large percentage (56%) of subjects with a 0-grade difference was judged to be clinically different by the raters. Thus, a 0-grade change could indicate no change or a somewhat smaller, but still clinically meaningful, change. This finding potentially indicates that the scale may not be optimally sensitive to clinical change. Because the raters assessed the image pairs before being trained on the scale, ratings of clinically meaningful difference may have incorporated factors of the chin (eg, height, contour) that are not evaluated by the CACRS.

Suwanchinda et al recently published a study that validated a chin scale for use in female Asian subjects.²⁴ The inter- and intra-rater reliability of the Asian Chin Projection Scale (ACPS) was similar to that of the CACRS, although the clinical significance of the ACPS was not reported.

An increasing number of young Chinese individuals are undergoing surgical and nonsurgical aesthetic procedures, driven by factors such as greater selection and accessibility of aesthetic products.⁷ China is currently ranked third in the world for the estimated number of plastic surgeons, behind the United States and Brazil.²⁵ A need exists for standardized ratings that medical practitioners who treat Chinese men and women seeking facial aesthetic improvements can uniformly apply to these subjects.

The present study demonstrated the validity of a photonic scale for assessing chin retrusion in this population. Comparison of other validation studies for aesthetic procedures indicates that photonic scales may be superior to other types of rating tools in evaluating target areas for aesthetic improvement. For example, photonic scales developed to evaluate the upper, middle, and

lower portions of subjects' faces displayed substantial or almost perfect interrater reliability.^{26,27} For assessing facial photodamage, a photonumeric scale displayed better interrater reliability than a descriptive scale.²⁸

Surgeons, not subjects, assessed the clinical significance of the CACRS; subjects may not agree or be able to discern that their appearance has changed. Further, the CACRS allows subjects to judge the appearance of their chin only in a lateral view, and many subjects may be more concerned with the frontal view. Additionally, the CACRS has not yet been used before and after treatment for chin retrusion. Whereas a 1-grade change on the CACRS is clinically meaningful, it may not represent the minimum change that would be considered meaningful.

CONCLUSIONS

Based on inter- and intra-rater agreement, the CACRS was validated for physician ratings of chin retrusion in Chinese subjects. Intra- and interrater agreement of the CACRS was substantial to almost perfect. A 1-grade difference was determined to be clinically significant for the CACRS. The CACRS offers medical practitioners validated standardized ratings uniformly applicable to Chinese men and women seeking aesthetic improvement for chin retrusion.

ACKNOWLEDGMENTS

The authors thank Diane Gao for administrative assistance. Writing and editorial assistance was provided to the authors by Regina Kelly, MA, of Peloton Advantage, LLC, an OPEN Health company, Parsippany, NJ, and was funded by Allergan Aesthetics, an AbbVie Company, Dublin, Ireland. All authors meet the ICMJE authorship criteria. Neither honoraria nor payments were made for authorship.

REFERENCES

- Bertossi D, Galzignato PF, Albanese M, et al. Chin microgenia: a clinical comparative study. *Aesthetic Plast Surg* 2015;39:651–658
- Hoeflin SM. *The Beautiful Face: The First Mathematical Definition, Classification, and Creation of True Facial Beauty*. Santa Monica, CA: Steven M. Hoeflin, MD; 2002
- Soh J, Chew MT, Wong HB. An Asian community's perspective on facial profile attractiveness. *Community Dent Oral Epidemiol* 2007;35:18–24
- Khosravanifard B, Rakhshan V, Raeesi E. Factors influencing attractiveness of soft tissue profile. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2013;115:29–37
- Torsello F, Graci M, Grande NM, et al. Relationships between facial features in the perception of profile attractiveness. *Prog Orthod* 2010;11:92–97
- Naini FB, Donaldson AN, McDonald F, et al. Assessing the influence of chin prominence on perceived attractiveness in the orthognathic patient, clinician and layperson. *Int J Oral Maxillofac Surg* 2012;41:839–846
- Liew S, Wu WT, Chan HH, et al. Consensus on changing trends, attitudes, and concepts of Asian beauty. *Aesthetic Plast Surg* 2016; 40:193–201
- Samizadeh S, Wu W. Ideals of facial beauty amongst the Chinese population: results from a large national survey. *Aesthetic Plast Surg* 2018;42:1540–1550
- Reilly MJ, Tomsic JA, Fernandez SJ, et al. Effect of facial rejuvenation surgery on perceived attractiveness, femininity, and personality. *JAMA Facial Plast Surg* 2015;17:202–207
- Sykes JM, Carruthers A, Hardas B, et al. Development and validation of a photonumeric scale for assessment of chin retrusion. *Dermatol Surg* 2016;42(Suppl 1):S211–S218
- Wu J, Hagg U, Rabie AB. Chinese norms of McNamara's cephalometric analysis. *Angle Orthod* 2007;77:12–20
- Lew KK, Ho KK, Keng SB, et al. Soft-tissue cephalometric norms in Chinese adults with esthetic facial profiles. *J Oral Maxillofac Surg* 1992;50:1184–1189, discussion 1189–1190
- Baillie LJ, Mirijali SA, Niven BE, et al. Ancestry and BMI influences on facial soft tissue depths for a cohort of Chinese and Caucosoid women in Dunedin, New Zealand. *J Forensic Sci* 2015;60:1146–1154
- Liu Y, Kau CH, Pan F, Zhou H, Zhang Q, Zacharopoulos GV. A 3-dimensional anthropometric evaluation of facial morphology among Chinese and Greek population. *J Craniofac Surg* 2013;24:e353–e358
- Farkas LG, Katic MJ, Forrest CR, et al. International anthropometric study of facial morphology in various ethnic groups/races. *J Craniofac Surg* 2005;16:615–646
- Gu Y, McNamara JA Jr, Sigler LM, Baccetti T. Comparison of craniofacial characteristics of typical Chinese and Caucasian young adults. *Eur J Orthod* 2011;33:205–211
- Sim RS, Smith JD, Chan AS. Comparison of the aesthetic facial proportions of southern Chinese and white women. *Arch Facial Plast Surg* 2000;2:113–120
- Wirthlin J, Kau CH, English JD, et al. Comparison of facial morphologies between adult Chinese and Houstonian Caucasian populations using three-dimensional imaging. *Int J Oral Maxillofac Surg* 2013;42:1100–1107
- Moate SJ, Darendeliler MA. Cephalometric norms for the Chinese: a compilation of existing data. *Aust Orthod J* 2002;18:19–26
- Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21:1331–1335
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428
- Cicchetti DV, Allison T. A new procedure for assessing reliability of scoring EEG sleep recordings. *Amer J EEG Tech* 1971;11:101–110
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174
- Suwanchinda A, Seo K, Geister TL, et al. A validated assessment scale for Asian chin projection. *Dermatol Surg* 2019;45(Suppl 1):S30–S37
- ISAPS international survey on aesthetic/cosmetic procedures performed in 2017. 2018; Available at: https://www.isaps.org/wp-content/uploads/2018/10/ISAPS_2017_International_Study_Cosmetic_Procedures.pdf. Accessed: January 22, 2019
- Yu Chao YY, Chhabra C, Corduff N, et al. Validation of a photonumeric assessment scale for grading the slope of the Asian forehead. *Dermatol Surg* 2019;45(Suppl 1):S38–S45
- Rzany B, Carruthers A, Carruthers J, et al. Validated composite assessment scales for the global face. *Dermatol Surg* 2012;38 (2 Spec No.):294–308
- Ayer J, Ahmed A, Duncan-Parry E, et al. A photonumeric scale for the assessment of atrophic facial photodamage. *Br J Dermatol* 2018; 178:1190–1195