

Available online at www.jbr-pub.org

Open Access at PubMed Central

JBR

The Journal of Biomedical Research, 2017 31(2): 108–115

Original Article

Classification of facial wrinkles among Chinese women

Jiechen Zhang^{1,Δ}, Wei Hou^{2,Δ}, Suying Feng², Xiangsheng Chen², Hongwei Wang^{1, ⊠}

¹Department of Dermatology, Huadong Hospital, Fudan University, Shanghai 200040, China;

² Institute of Dermatology, Chinese Academy of Medical Sciences and Peking Union Medical College, Nanjing, Jiangsu 210042, China.

Abstract

It is generally recognized that Caucasians and Asians have different skin aging features. The aim of this study was to develop a facial wrinkle grading scale for Chinese women. Standard photographs were taken of 242 Chinese women. Six sets of 0 to 9 wrinkle scales with reference photographs and descriptions were selected, including grading scales for resting and hyperkinetic crow's feet, frontalis lines, glabellar frown lines, and nasolabial folds. To identify the scale by objective quantitative measurement, skin surface measurements from the Visioscan[®] VC98 were used. To test the reliability and validity of our wrinkle scale, a multi-rater consensus method was used. A double-blind, randomized, vehicle-controlled 12-week study was conducted to use this clinical photo-score to evaluate the efficacy and safety of Centella triterpenes cream[®] in treating crow's feet. A newly developed 10-point photographic and descriptive scale emerged from this study. The final atlas of these photographs contained a total of 6 sets with 10 pictures each. From 0 to 9, surface evaluation of smoothness (SEsm) parametric measurements decreased progressively, indicating that the scale increased inversely. Weighted kappa coefficients for intra-assessor were between 0.75–0.87. The overall Kendall's coefficient is 0.86 on the first rating and 0.87 on the second rating. Thirty-six volunteers were recruited and 35 subjects completed a 12-week trial. Clinical photo-score by investigator showed a significant difference (P < 0.05) between the treatment side and control side after 4 weeks. Use of these scales in clinical settings to evaluate facial wrinkles in Asians individuals is recommended.

Keywords: facial wrinkles, classification, grading scale, photographic scale, Centella Triterpenes cream, evaluation

Introduction

The skin undergoes intrinsic aging (chronological aging), like all other body organs. The skin also undergoes extrinsic aging (photoaging), which is the result of exposure to ultraviolet radiation. Therefore, the aging process of the skin can be divided into two independent, biologically and clinically distinct processes: chronological aging and photoaging. The effects

of both processes overlap on facial skin^[1]. Despite the variety of clinical characteristics of facial skin aging, wrinkles are considered the most representative manifestation and have an important social impact.

As demand for facial wrinkle rejuvenation increases, related research about wrinkle prevention or treatment is increasing which highlights the need for an objective clinical instrument for the evaluation of the effectiveness of therapies. The techniques of evaluating skin

Received 09 December 2015, Revised 18 March 2016, Accepted 06 July 2016, Epub 09 November 2016

CLC number: R75, Document code: A

The authors reported no conflict of interests.

 $^{^{\}Delta}$ These authors contributed equally to this work.

[™]Corresponding author: Dr. Hongwei Wang, Department of Dermatology, Huadong Hospital, Fudan University, Shanghai 200040, China. Tel:+ 86-136-5176-3180, Email: hongweiwang2005@aliyun.com.

^{© 2017} by the Journal of Biomedical Research. All rights reserved

aging can be divided into direct methods (including clinical grading systems and mechanical measurements) and indirect methods (including silicone impression and computer software analysis)^[2]. Among them, the clinical grading system is more widely used because it is the easiest way to perform and therefore more practical in the clinical setting. The variety of scoring and scaling systems for assessing facial wrinkles can be classified as descriptive grading scales^[3–5], photographic grading scales^[6–14] and visual analog scales^[4]. However, there is no "gold standard" among grading scales and almost all of the scales mentioned above are based on Caucasian individuals.

It is generally recognized that Caucasians and Asians have different skin aging features. A pilot skin aging study between Chinese and European individuals showed that for each facial skin area, wrinkle onset is delayed by about 10 years in Chinese women as compared to French women^[15]. Despite the variety of published scoring systems for assessing different parts of facial wrinkles, few have been based on Asian individuals. A research study in Japan^[16] surveyed 87 women in Tokyo (Japan), 100 women in Shanghai (China), and 90 women in Bangkok (Thailand). The result indicates the diversity of Asian skin. For example, Chinese women had significantly more severe wrinkles in the area around the eyes compared to Japanese women, while Thai women had significantly more severe wrinkles in the lower halves of their faces compared to Chinese women. In this study, Japanese researchers developed a 5-point photo scale for facial wrinkles based on Japanese women, but did not test the validity and reliability. To investigate cutaneous photoaging in Koreans including the influence of sex, sun exposure, smoking. and skin color, the researcher also developed new photographic scales for assessing the cutaneous grading of wrinkles and dyspigmentation. This scale did not examine validity and reliability and was for the whole face, not for each facial skin area.

We believe a photographic scale for the nonwhite population is necessary especially because the Caucasian skin type is represented in just a small minority of the world's population. We developed a facial wrinkle scoring system for evaluating the severity of facial wrinkles in Asian individuals.

Subjects and methods

Instrument development

Healthy female volunteers from 15 to 75 years old were included. Exclusion criteria: 1) Pregnant or nursing during the study. 2) Previous cosmetic surgery including laser, chemical peeling, botulinum toxin, injectable fillers, face lift, etc. 3) Severe chronic diseases that affect skin evaluation. 4) Burn history in the previous month. 5) History of chronic medicine intake (more than 10 years). A total of 242 volunteers, ranging in age from 19 to 71 years old, were involved in this study and signed the consent form.

After washing their faces, volunteers were acclimated for 20 minutes in the same condition-controlled room (temperature $20\pm2^{\circ}$ C, humidity 50%-60%). Standardized facial photographs were taken by the Skin Image Analyzer (SIA0612) programmed to the same light source, fixed position, and identical amplification factor settings. Separately taken standardized facial photographs were taken at rest (static) and with expression (dynamic), in both frontal and oblique (45°) positions.

For this study, four of the facial regions were selected: lateral canthus (both static and dynamic), glabellar, forehead (both static and dynamic), and nasolabial folds. The severity of wrinkles was assessed in three stages. The first stage roughly organized the 242 photos into three broad classes: mild, moderate, and severe wrinkles. Rather than length or number of wrinkles, the depth of the midpoint between the wrinkles was used as a reference point for comparisons. In the case of multiple wrinkles, only the deepest wrinkle was assessed. In the second stage, a more refined score was obtained by comparing an individual subject's photograph with photos from each broad class. Then, photographic standards for photos to serve as representative examples of each wrinkle class were selected. In the third stage, two dermatologists who constructed the scales reviewed the scores of the 242 photos to test the feasibility of the newly developed photographic scale.

To quantify the scale using objective quantitative measurement, skin surface measurements from the Visioscan® VC98 were used. The SELS parameter of Visioscan® VC98 consists of four parameters, in which SEsm smoothness is inversely proportional to the width and form of the wrinkles.

Reliability and validity study

Nine dermatologists (2 dermatologic-surgeons, 3 dermatologists with laser expertise, 2 cosmetic dermatologists, and 2 dermatopathologists) were trained to use the final atlas of the photographic grading scales with descriptions. They then rated 48 images which were selected from the 242 subjects based on quality and representative distribution across each four facial regions. To avoid any biases, the images presented only the area of the face to be evaluated, rather than the

Crow' s feet (static)



Fig. 1 The final atlas of the crow's feet grading scale. A: static grading scale. B: dynamic grading scales. Five main classes: 1, 3, 5, 7 and 9 representing yet visible wrinkle, visible fine wrinkle; well-defined moderate wrinkle, deep carven edges wrinkle, and redundant folds. Class 2, 4, 6 and 8 were between the main classes. Class 1: Very shallow or lines yet visible wrinkle; Class 2: Just visible wirnkle, like hazy crease; Class 3: Visible wrinkle, like light clear crease; Class 4: Clearly visible weinkle; Class 5: Clearly visible wrinkle and well-defined edges; Class 6: Moderately deep wrinkle; Class 7: Deep wrinkle and carven edges; Class 8: Deep and prominent wrinkle with furrow; Class 9: Redundant folds.

whole face. The assessments took place over 2 consecutive days and began within 1 hour following completion of the training.

Statistical analyses were performed using SPSS 17.0. To test the agreement between two ratings of the same 48 images by the same assessors, the result for weighted kappa for intra-assessor was calculated for the 9 dermatologist raters. To test the reliability among multiple observers, the Kendell's coefficients for interassessor were calculated for the 9 dermatologist raters. They range from 0 to 1, where 0 represents poor agreement and 1 represents strong agreement.

Clinical use

Centella Asiatica (an herb) has been used hundreds of years for wound healing and as a traditional medicine in Asiatic countries. It has been reported that a preparation containing asiaticoside can significantly improve the periorbital wrinkles^[17]. To test our newly developed scales, we design a randomized, double-blind vehicle-controlled 12 week study of the anti-wrinkle effects of the centella triterpenes cream[®] on crow's feet of female volunteers. Centella triterpenes cream[®] was applied three times daily to one side of the canthus and vehicle-

controlled cream was applied to the other side. Efficacy was based on a investigator-blinded assessment by the newly developed crow's feet wrinkle scale, subject selfblinded assessment, and Visioscan VC98[®] quantitative analysis every 4 weeks.

Results

Classification of Chinese women's facial wrinkles

Newly developed 10-point photographic and descriptive scale comprised of five main classes: class 1, class 3, class 5, class 7 and class 9 representing yet to be formed visible wrinkles, visible fine wrinkles, welldefined moderate wrinkles, deeply etched wrinkles, and

Mild

redundant folds. Class 2, 4, 6 and 8 were between the main classes. The final atlas of these photographs contained a total of 6 sets, including lateral canthus (both static and dynamic), glabellar, forehead (both static and dynamic), and nasolabial folds. Each set with 10 pictures (*Fig. 1-4*).

Reliability and validity of the scale system

Weighted kappa coefficients for intra-assessor were between 0.75 and 0.87 (0.75-0.79 for male and 0.81-0.87 for female) (*Table 1*). Among the first rating, the Kendall's coefficient for inter-assessor of the motion forehead wrinkle and nasolabial wrinkle were the highest (0.94), while the static forehead wrinkle was

Forehead lines (static)



Fig. 2 The final atlas of the forehead lines grading scale. A: static grading scale. B: dynamic grading scales. Five main classes: 1, 3, 5, 7 and 9 representing yet visible wrinkle, visible fine wrinkle; well-defined moderate wrinkle, deep carven edges wrinkle, and redundant folds. Class 2, 4, 6 and 8 were between the main classes. Class 1: Very shallow or lines yet visible wrinkle; Class 2: Just visible wirkle, like hazy crease; Class 3: Visible wrinkle, like light clear crease; Class 4: Clearly visible weinkle; Class 5: Clearly visible wrinkle and well-defined edges; Class 6: Moderately deep wrinkle; Class 7: Deep wrinkle and carven edges; Class 8: Deep and prominent wrinkle with furrow; Class 9: Redundant folds.

Moderate

Severe

Glabellar frown lines



Fig. 3 The final atlas of the glabellar frown lines grading scale. Five main classes: 1, 3, 5, 7 and 9 representing yet visible wrinkle, visible fine wrinkle; well-defined moderate wrinkle, deep carven edges wrinkle, and redundant folds. Class 2, 4, 6 and 8 were between the main classes. Class 1: Very shallow or lines yet visible wrinkle; Class 2: Just visible wrinkle, like hazy crease; Class 3: Visible wrinkle, like light clear crease; Class 4: Clearly visible weinkle; Class 5: Clearly visible wrinkle and well-defined edges; Class 6: Moderately deep wrinkle; Class 7: Deep wrinkle and carven edges; Class 8: Deep and prominent wrinkle with furrow; Class 9: Redundant folds.

the lowest (0.72). Overall Kendall's coefficient is 0.86 on first rating, 0.87 on second rating, indicating a high level of inter-assessor consistency of all assessors (*Table 2*).

Parameters measurement

SELS parameters were used to measure the width and form of each class. From 0 to 9, the SEsm parametric measurements decreased progressively, indicating that the scale increased inversely(*Table 3*).

Clinical use

Thirty-six volunteers were recruited and 35 subjects completed a 12-week trial to test Centella Triterpenes cream® in treating crow's feet. One volunteer dropped out in view of a business trip. Clinical photo-score by investigator using this newly developed 10-point photographic and descriptive scale showed a significant difference (P < 0.05) between the treatment side and control side after 4 weeks. The significant difference of the score between the two sides was shown after 8 weeks (*Table 4*). The improvement of wrinkles was more obvious on the treatment side than on the control side. Measurements by Visioscan® VC98 demonstrated a significant increase (P < 0.05) of the SEw value in the treatment side, whereas in the control side, a decrease was observed. Subjects' assessments showed no significant difference in the change of coarse wrinkles, whereas in the fine-wrinkle assessment, a significant difference was observed (P < 0.05).

Rater number	Kappa	Weighted Kappa		
		K _w	95%CI	
1	0.50	0.82	0.75-0.89	
2	0.46	0.75	0.65-0.85	
3	0.61	0.83	0.76-0.91	
4	0.59	0.85	0.79-0.92	
5	0.71	0.87	0.79-0.95	
6	0.46	0.78	0.69-0.87	
7	0.42	0.79	0.71-0.86	
8	0.48	0.81	0.74-0.88	
9	0.49	0.78	0.71-0.87	



Nasolabial folds

Fig. 4 The final atlas of the nasolabial folds grading scale. Five main classes: 1, 3, 5, 7 and 9 representing yet visible wrinkle, visible fine wrinkle; well-defined moderate wrinkle, deep carven edges wrinkle, and redundant folds. Class 2, 4, 6 and 8 were between the main classes. Class 1: Very shallow or lines yet visible wrinkle; Class 2: Just visible wirkle, like hazy crease; Class 3: Visible wrinkle, like light clear crease; Class 4: Clearly visible weinkle; Class 5: Clearly visible wrinkle and well-defined edges; Class 6: Moderately deep wrinkle; Class 7: Deep wrinkle and carven edges; Class 8: Deep and prominent wrinkle with furrow; Class 9: Redundant folds.

Discussion

The increasing interest in surgical and nonsurgical (e.g., laser, BoNT, cosmetic procedures) methods to improve the appearance of facial wrinkles requires the development of techniques to measure the severity of facial wrinkles. A variety of noninvasive and invasive techniques have been developed to assess skin wrinkles. However, according to our clinical experiences and publications^[4], such techniques are more suitable for laboratory research use rather than clinical purposes. Facial wrinkles can be treated in various ways, such as through the use of topical cosmetic agents, injectable derma fillers, botulinum toxin-A, laser and surgery. Thus, a validated tool to objectively evaluate the effects

of specific therapies is valuable in the hands of dermatologists and aesthetic surgeons. Clinical scoring systems are generally considered an easy, consistent, reliable and practical tool in assessments. Recent studies in this field are increasingly more focused on developing a standard grading system instead of a variety of published systems^[18]. A standardized grading system of skin aging should take into account reliability and validation, as well as the differences between Asian, Caucasian, and African skin aging conditions. As Kappes emphasized^[4], special photographic scales for the nonwhite population are necessary, especially since the Caucasian skin type represents just a small minority of the world's population.

There were no public research publications about a

	Kendall's	coefficient	
Facial wrinkle	Rating		
	First	Second	
Crow's feet (static)	0.91	0.93	
Crow's feet (dynamic)	0.76	0.79	
Forehead wrinkle (static)	0.72	0.75	
Forehead wrinkle (dynamic)	0.94	0.98	
Nasolabial wrinkle	0.94	0.88	
Glabellar frowns	0.92	0.91	
Overall	0.86	0.87	

Scale	SEsm parametric (mean)						
	Crow's feet (static)	Crow's feet (dynamic)	Forehead wrinkle (static)	Forehead wrinkle (dynamic)	Nasolabial wrinkle	Glabellar frowns	
0	177.400	282.900	262.200	92.360	283.200	286.200	
1	170.800	163.300	186.500	59.920	177.400	258.600	
2	142.400	146.300	163.000	56.330	165.200	223.700	
3	125.100	123.000	119.000	49.170	161.700	191.870	
4	123.000	88.110	96.240	46.260	144.100	172.400	
5	111.600	86.960	92.540	40.900	124.000	164.900	
6	110.700	81.600	91.550	35.120	115.700	157.610	
7	96.500	77.410	82.400	32.970	106.300	137.400	
8	92.570	66.200	74.810	29.120	105.200	122.400	
9	90.900	62.970	70.860	27.530	101.500	107.800	

facial wrinkle scaling system in China, despite the nation boasting the largest population in Asia. Related research in China almost always used published wrinkle classifications based on Caucasian skin aging or created a scoring scale for temporary purposes. Some large pharmaceutical companies have developed scales for evaluating their cosmetic products, but these scales are proprietary commercial industry secrets which cannot be used to judge newly developed procedures and are not available for general use.

Related research on skin aging assessments either used self-developed scales^[19-20] for temporary purposes or cited publications using scales based on other races^[21-22]. Lin *et al.* ^[21] compared the differences between Chung photographic scales and Glogau photoaging classification through the evaluation of 303 Chinese female faces. The former is an Asianbased photographic scale that is designed for assessing wrinkles and dyspigmentation, while the latter is a Caucasian-based descriptive scale. Overall, the authors concluded that the Chung photographic scales are more suitable for Asians than the Glogau scale. However, this scale evaluated photoaging by only including wrinkles and dyspigmentation based on male and female individuals. The authors felt it is difficult to evaluate telangiectasies, which are more common in photoaging skin^[22]. In addition, the Chung photographic scale was for the whole face and not for each facial skin area.

To improve epidemiologic quality and make our grading system more standardized, we collected 242 healthy female volunteers, ranging from 15 to 75 years old, including urban and rural subjects, city people and village folk. This classification assessed skin aging, including chronological aging and photoaging. Our tenpoint facial wrinkle assessment scale is a photographic grading scale with descriptions.

Rated on a 0-9 scale, the wrinkle scale can be used in research with different types of aesthetic procedures. For example, in surgery, injectable dermal fillers, or botulinum toxin-A injections, the improvement in wrinkles is distinct. Thus, none (0), mild (1-3), moderate (4-6), and severe (7-9) can be used. To ensure a high quality in clinical practice the 0 to 9 scale can be used.

Reliability and validation of the wrinkle scale was tested. The weighted Kappa result shows that the agreement between same raters was high. Female raters had more intra-rater agreement than male raters. This suggests that it may be more difficult for males to rate mild wrinkles. The high Kendall's coefficient result shows good inter-rater reliability.

We developed a valid facial wrinkle scoring system

Week	Score (treatment side)	Δ	Score (control side)	Δ	P-value	Δp -value
0	3.4±1.5		3.3±1.4		0.685	
4	3.0±1.5	$0.40{\pm}0.6$	3.3±1.4	$0.0{\pm}0.2$	0.41	0.02*
8	2.6±1.2	-0.8 ± 0.6	3.2±1.3	-0.1±0.4	0.03*	0.00*
12	2.5±1.2	$-0.9 {\pm} 0.8$	3.3±1.4	-0.1 ± 0.4	0.00*	0.00*

not only for use with daily purposes, but also for an objective, quantitative grading scale to be used as a clinical guideline for evaluating the severity of facial wrinkles in Asian patients. Validation studies show that this scale has good inter-and intra-assessor reliability. This scale is now in clinical use in China. We recommended that esthetic doctors in other countries use this scale to evaluate Asian individuals. In the future, a "gold standard" scale should consider the difference in races and account for those differences.

Acknowledgments

We thank Hong-Sheng Wang, Tong Lin, MD Yi, Liu MD; Yan Wang, MD; Shu-Xian Shang, MD; Vu-Le Wu, MD; Li-Ming Huang, MD, for their contribution as assessors of the measurement tool validation. We also thank Adnan Nasir, MD (Chapel Hill, USA) and Crystal Shen, MD candidate (Mayo Medical School, USA) for language edits.

References

- Fisher GJ, Kang S, Varani J, et al. Mechanisms of photoaging and chronological skin aging[J]. *Arch Dermatol*, 2002, 138(11): 1462–1470.
- [2] Hatzis J. The wrinkle and its measurement—a skin surface Profilometric method[J]. *Micron*, 2004, 35(3): 201–219.
- [3] Glogau RG. Aesthetic and anatomic analysis of the aging skin[J]. Semin Cutan Med Surg, 1996, 15(3): 134–138.
- [4] Kappes UP, Elsner P. Clinical and photographic scoring of skin aging[J]. Skin Pharmacol Appl Skin Physiol, 2003, 16(2): 100– 107.
- [5] Fitzpatrick RE, Tope WD, Goldman MP, et al. Pulsed carbon dioxide laser, trichloroacetic acid, Baker-Gordon phenol, and dermabrasion: a comparative clinical and histologic study of cutaneous resurfacing in a porcine model[J]. *Arch Dermatol*, 1996, 132(4): 469–471.
- [6] Weiss JS, Ellis CN, Goldfarb MT, et al. Tretinoin therapy: practical aspects of evaluation and treatment[J]. *J Int Med Res*, 1990, 18(Suppl 3): 41C–48C.
- [7] Larnier C, Ortonne JP, Venot A, et al. Evaluation of cutaneous photodamage using a photographic scale[J]. *Br J Dermatol*, 1994, 130(2): 167–173.
- [8] Griffiths CE, Wang TS, Hamilton TA, et al. A photonumeric scale for the assessment of cutaneous photodamage[J]. Arch

Dermatol, 1992, 128(3): 347-351.

- [9] Lemperle G, Holmes R E, Cohen S R, et al. A classification of facial wrinkles[J]. *Plast Reconstr Surg*, 2001, 108(6):1735– 1750, 1751–1752.
- [10] Carruthers A, Carruthers J, Hardas B, et al. A validated grading scale for forehead lines[J]. *Dermatol Surg*, 2008, 34(Suppl 2): S155–S160.
- [11] Carruthers A, Carruthers J, Hardas B, et al. A validated grading scale for marionette lines[J]. *Dermatol Surg*, 2008, 34(Suppl 2): S167–S172.
- [12] Kim EJ, Reeck JB, Maas CS. A validated rating scale for hyperkinetic facial lines[J]. *Arch Facial Plast Surg*, 2004, 6(4): 253–256.
- [13] Chung JH, Lee SH, Youn CS, et al. Cutaneous photodamage in Koreans: influence of sex, sun exposure, smoking, and skin color[J]. Arch Dermatol, 2001, 137(8): 1043–1051.
- [14] Carruthers A, Carruthers J, Hardas B, et al. A validated grading scale for crow's feet[J]. *Dermatol Surg*, 2008, 34(Suppl 2): S173–S178.
- [15] Nouveau-Richard S, Yang Z, Mac-Mary S, et al. Skin ageing: a comparison between Chinese and European populations. A pilot study[J]. *J Dermatol Sci*, 2005, 40(3): 187–193.
- [16] Tsukahara K, Sugata K, Osanai O, et al. Comparison of agerelated changes in facial wrinkles and sagging in the skin of Japanese, Chinese and Thai women[J]. *J Dermatol Sci*, 2007, 47(1): 19–28.
- [17] Lee J, Jung E, Lee H, et al. Evaluation of the effects of a preparation containing asiaticoside on periocular wrinkles of human volunteers[J]. *Int J Cosmet Sci*, 2008, 30(3): 167– 173.
- [18] Carruthers A, Carruthers J. A validated facial grading scale: the future of facial ageing measurement tools?[J]. J Cosmet Laser Ther, 2010, 12(5): 235–241.
- [19] Yuan Chao X W. Comparison of three methods for evaluation of wrinkles[Z].2005, 92–96.
- [20] Hong-Hua Y. Treatment of upper-facial wrinkles with botulinum toxin type A in 1 000 cases[J]. J Pract Aesthetic Plast Surg, 2001, 12(4): 179–181.
- [21] Tong L, Zhan-Chao Z. Gongxiang-Dong. Comparison of two different methods on evaluation of female face photoaging[J]. *Chin J Med*, 2009, 18(11): 1648–1649.
- [22] Li L, Xi W, Wei L, et al. Analysis of factors related to female facial wrinkles:a survey of 1004 Chinese women of Han nationality[J]. *Chin J Pract Aesthetic Plast Surg*, 2004, 15(3): 126–128.