

A photographic scale for the assessment of human facial wrinkles

KAZUE TSUKAHARA, YOSHINORI TAKEMA,
HARUHITO KAZAMA, YUKIKO YORIMOTO,
TSUTOMU FUJIMURA, SHIGERU MORIWAKI,
TAKASHI KITAHARA, MICHIO KAWAI, and
GENJI IMOKAWA, *Biological Science Laboratories, Kao Corporation,*
2606 Akabane, Ichikai, Haga, Tochigi, 321-3497 (K.T., Y.T., T.F.,
S.M., T.K., G.I.), and Cosmetic Science Laboratories, 2-1-3 Bunka,
Sumida-ku, Tokyo, 131-8501 (H.K., Y.Y., M.K.), Japan.

Accepted for publication February 15, 2000.

Synopsis

We report the utility of photographic scales to assess the degree of human facial wrinkles. A five-grade photo scale of wrinkles at eye corners was developed using photos obtained from 411 female participants aged 17 to 83 years. Based on this photo scale, scorings (five grades and nine grades) of all photos were performed by two specialists, and a standard photo for each wrinkle grade was obtained. In addition, in order to evaluate the influence of inter-observer differences in grading criteria, another scoring was performed by five general observers. The agreement between the grade standard and the score given by the general observers was evaluated by calculating the chance-corrected coefficient, i.e., the kappa value. The degree of agreement for the five-grade scale was more consistent than that for the nine-grade scale system, in which the kappa values were 0.499 and 0.396, respectively. When scoring was performed using the five-grade photo scale for the eye corners after 15 participants used a wrinkle-improving agent, a significant reduction of the wrinkle scores was confirmed. Taken together, the present findings indicate that the five-grade wrinkle photo scale is valid and useful to assess the degree of facial wrinkles.

INTRODUCTION

In recent years, cosmetic agents for the improvement of wrinkles have been actively developed. Photographs have been used for the clinical panel assessment of photodamaged skin treated with several reagents (1,2). Clinical tests of retinoic acid as a treatment for wrinkles are representative, and the evaluation of its effects is often performed by scoring based on descriptive definitions (3–7). Scoring of photodamage, namely wrinkling, using a photo scale has been performed, and its statistical validity has been

Address correspondence and reprint requests to Genji Imokawa. Tel: 285-68-7467, Fax: 285-68-7305, E-mail: 073733@kastanet.kao.co.jp

reported by dermatologists (8,9). However, these photo scales of wrinkling are standards for Caucasian people, and it is necessary to develop a photo scale for Japanese, because the principal manifestation of photodamage in Far East Asians, including Japanese, is based on pigmentary change rather than wrinkling (8,14). We therefore evaluated the statistical validity of the evaluation of facial wrinkles using photo scales, and we also examined age-related changes in wrinkles at eight facial sites in Japanese females. In addition, these photo scales were applied to an evaluation of a cosmetic agent for the improvement of wrinkles.

MATERIALS AND METHODS

SUBJECTS

The subjects were 613 healthy Japanese females aged 17–83 years living in the Tokyo metropolitan area. Each subject washed her face using a liquid face-washing agent (Kao Corp., Tokyo) and water. The subjects were photographed in a room with a constant temperature of 20°C and a constant relative humidity of 40–50%. Ten minutes after washing her face, each subject was seated on a chair with her eyes lightly closed, and photos were taken as follows. A photo of the entire face was obtained for the scoring at eight sites on the face, and separate close-up photos of the outer eye corners and the forehead were also obtained. The photographs were taken using a Medical Nikkol camera (Nikon, Tokyo) with a diffused-light electric flash, which has a frame to ensure the standardization of lighting, angle, and distance. All film used (Fujichrome for color transparencies, Fuji Film, Tokyo) originated from the same batch. The first frames of each role of film included a standard gray scale and color charts from the film manufacturer to facilitate color correction (10). The subjects were photographed in three groups (202 subjects = first group, 164 = second group, and the remaining 247 = third group).

SELECTION OF PHOTOGRAPHIC SCALE PHOTOS AND ASSIGNMENT OF STANDARD GRADES

Five-grade wrinkle scales for the outer eye corners and forehead. Using a five-point descriptive scale (defined as follows: 1 = none, 2 = mild, 3 = mild/moderate, 4 = moderate, 5 = severe), two specialists separately performed a scoring of wrinkles on photos of the second group (164 subjects) and gave consensus scores. A single photograph with an appropriate score was then chosen to depict each of the five grades. The photo scale for the eye corners is shown in Figure 1.

Nine-grade wrinkle scale for the outer eye corners. In the nine-grade evaluation, the five-grade photos selected were given middle scores between grades, resulting in a nine-grade scale. Using the nine-grade photo scale, two specialists separately performed the scoring of wrinkles on photos of the third group of 247 subjects. They discussed the obtained scores until reaching a consensus on the photos to use for each of the nine grades.

ASSESSMENT OF THE USEFULNESS OF THE PHOTOGRAPHIC WRINKLE SCALES

Comparison of the five-grade and nine-grade scales. Using the five- and nine-grade standard

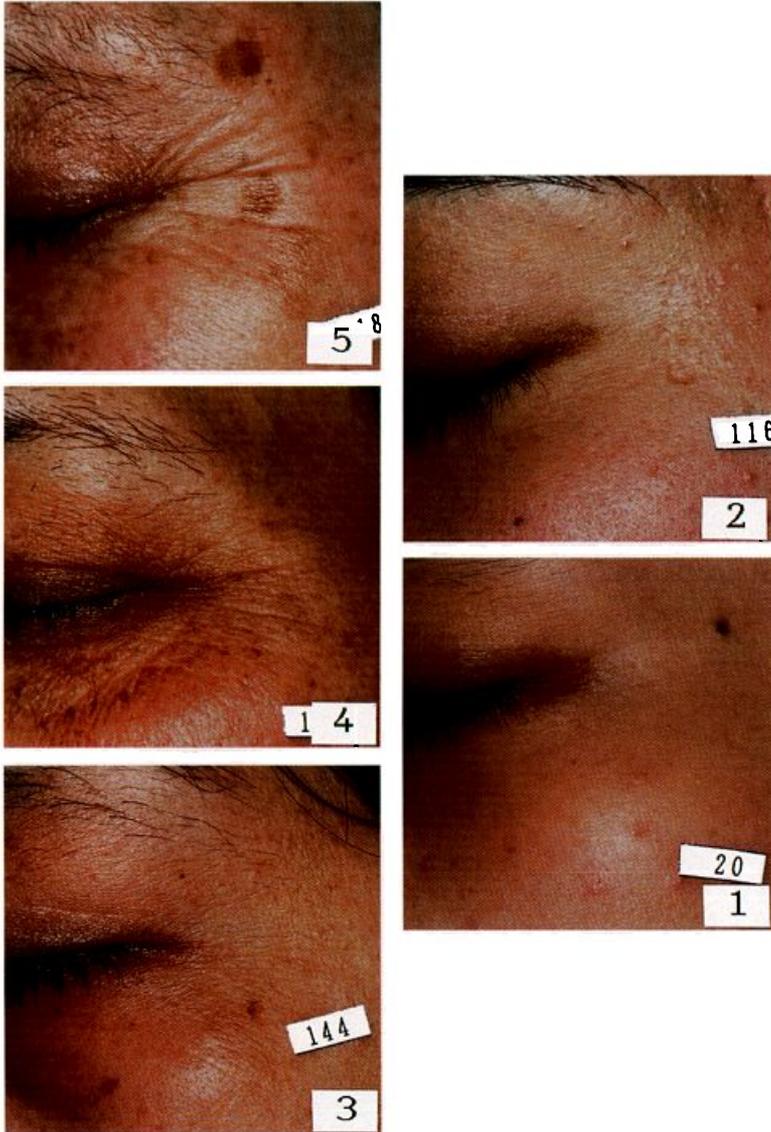


Figure 1. Five-grade photographic scale used to evaluate wrinkles at the outer eye corners. Grade 1 is no wrinkles; grade 2 is mild wrinkles; grade 3 is mild/moderate wrinkles; grade 4 is moderate wrinkles; grade 5 is severe wrinkles.

photos of the outer eye corners selected as described above, five general observers scored the eye corner wrinkles in the photos of the second group (164 subjects), who were used to establish the five-grade scale, and the third group (247 subjects), who were used to establish the nine-grade scale, respectively.

Reproducibility of the nine-grade evaluation. For the determination of the reproducibility of scores according to the photo standards, scoring was done by the same five general observers one week after the initial scoring (to avoid errors resulting from being accustomed to the evaluation), using the nine-grade scales.

Effects of facial site (the eye corners and forehead) on the wrinkle scores. Five general observers scored the wrinkles according to the five-grade scale on photos of the eye corners and forehead of the second group of subjects. These scores were compared with the consensus scores, and the consistency of each photo score was evaluated. The five-grade photo scale for forehead wrinkles is shown in Figure 2.

Age-related changes in the wrinkle score at each site of the face. The scoring of wrinkles at eight sites of the face (forehead, glabella, eye corners, upper eyelids, lower eyelids, cheeks,

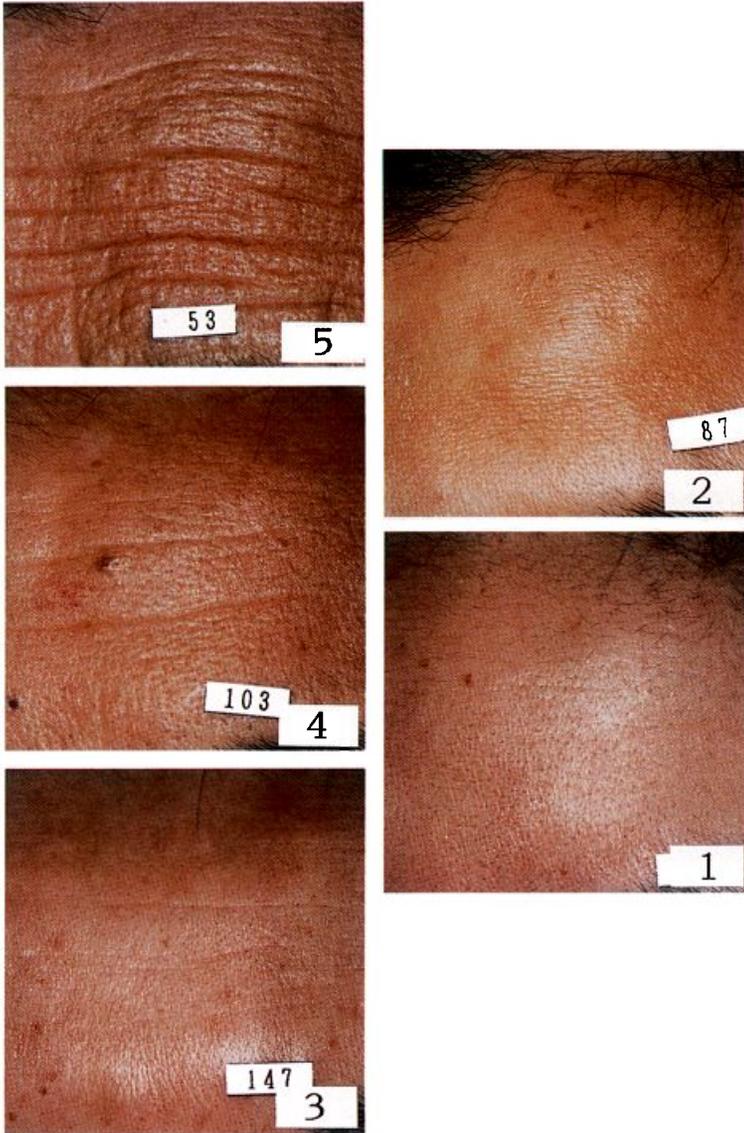


Figure 2. Five-grade photographic scale used to evaluate wrinkles at the forehead. Grade 1 is no wrinkles; grade 2 is mild wrinkles; grade 3 is mild/moderate wrinkles; grade 4 is moderate wrinkles; grade 5 is severe wrinkles.

nasolabial groove, and mouth angles) was done using the five-grade standard on photos of the frontal face of 366 females (the first and second groups) by the two specialists.

Clinical test of wrinkle-improving agents. A clinical test of an agent for the improvement of wrinkles (Sofina Seraty, Kao Corp.) was carried out in 15 Japanese females. This agent and placebo (a base cream) were topically applied by the subjects to both outer eye corners in the half-face manner daily for about eight weeks. Evaluations of the wrinkles of each subject were done before and after the application period by a specialist, in the double-blind manner, using the nine-grade scale.

Statistical analysis. The influence of age on wrinkling was studied using analysis of linear, logarithmic, and exponential correlation coefficients. Inter-observer agreement and intra-observer repeatability were analyzed and quantified by use of the kappa coefficient (11,12), a chance-corrected intraclass correlation coefficient with possible values ranging from -1 (complete disagreement) to +1 (complete agreement). Values above 0.61 are generally interpreted as indicating substantial agreement, values between 0.41 and 0.60 as moderate agreement, and values between 0.21 and 0.40 as fair agreement; in general, an agreement level above fair is preferable (13). Table I shows an illustrative agreement matrix (11).

In the evaluation of the improvement in wrinkles following the use of a cosmetic agent, differences between the before-use and after-use scores were analyzed by the paired Student t-test.

RESULTS

COMPARISON OF THE FIVE-GRADE AND NINE-GRADE PHOTOSCALES FOR EYE-CORNER WRINKLES

Five general examiners performed a five-grade evaluation and a nine-grade evaluation of wrinkles on photos of the eye corners, according to the same photo scales. More than

Table I
An Agreement Matrix of Proportions

Score		Judge A					Σfib
		1	2	3	4	5	
Judge B	1	f11*	f12	f13	f14	f15	Σf1b
	2	f21	f22	f23	f24	f25	Σf2b
	3	f31	f32	f33	f34	f35	Σf3b
	4	f41	f42	f43	f44	f45	Σf4b
	5	f51	f52	f53	f54	f55	Σf5b
Σfai		Σfa1	Σfa2	Σfa3	Σfa4	Σfa5	Σf = N _T

* Numbers of notations for two-observer agreement.

Σfai: the proportion of units in which the judges agreed.

Σfib: the proportion of units for which agreement is expected by chance.

$$f_0 = \Sigma f_{ii} \quad f_c = \frac{\Sigma(\Sigma f_{ai} \times \Sigma f_{ib})}{N_T} \quad \kappa = \frac{f_0 - f_c}{N_T - f_c}$$

60% of the five-grade scores of wrinkles at the eye corners in the second group were grade 4 or 5 (Table II), showing relatively many wrinkles. In the nine-grade evaluation, however, subjects with grades 7–9 occupied 38.9% of the third group, and the distribution of the scores was relatively even (Table III). The degree of consistency between the five-grade and nine-grade evaluation scores determined by each examiner and the consensus score by calculation of kappa values are shown in Table IV. The kappa values in the five-grade evaluation ranged from 0.400 to 0.657 (mean 0.499), indicating that the agreement was fair to substantial. In the nine-grade evaluation, the kappa value ranged from 0.252 to 0.507 (mean 0.396), indicating fair-to-moderate agreement. Since the distribution of grades and the number of subjects differed between these two evaluations, a direct comparison of the results may not be appropriate. However, the nine-grade evaluation scores were less consistent with the standard scores than were the five-grade evaluation scores.

REPRODUCIBILITY OF THE NINE-GRADE SCALE

Table V shows reproducibility when the examiner repeated the evaluation of the wrinkles of the same subjects in the third group by the nine-grade photo scale. The kappa values ranged from 0.252 to 0.507 (mean 0.396) in the first scoring and from 0.327 to 0.474 (mean 0.414) in the second scoring, indicating that this nine-grading procedure is reproducible with fair-to-moderate agreement.

EFFECTS OF SITE (EYE CORNERS VERSUS FOREHEAD) ON THE WRINKLE SCORES

The distribution of the five-grade score for forehead wrinkles is shown in Table VI. In this group, grades 4 and 5 accounted for 31.1%, which was lower than the percentage of grades 4 + 5 at the eye corners (about 60%). A relatively even distribution of forehead wrinkle scores was observed. Table VII shows the degree of consistency between the five-grade scores for the forehead and those for the eye corners determined by each examiner and the consensus score. The kappa values in the evaluation at the eye corner ranged from 0.400 to 0.657 (mean 0.499), indicating fair-to-substantial consistency, and that at the forehead ranged from 0.322 to 0.441 (mean 0.363), indicating fair-to-moderate consistency. The kappa value was slightly lower at the forehead than at the eye corners.

Table II
Consensus Values of Evaluation by Five General Examiners of Severity of Wrinkles at the Outer Eye Corners of 164 Subjects

Wrinkle grade	n (%)
1	34 (20.7)
2	11 (6.7)
3	19 (11.6)
4	66 (40.2)
5	34 (20.7)
Total	164

Table III
Consensus Values of Evaluation by Five General Examiners of Severity of Wrinkles at the Outer Eye Corners of 247 Subjects

Wrinkle grade	n (%)
1	36 (14.6)
2	25 (10.1)
3	21 (8.5)
4	36 (14.6)
5	33 (13.4)
6	35 (14.2)
7	25 (10.1)
8	19 (7.7)
9	17 (6.9)
Total	247

Table IV
Inter-Observer Agreement on Eye Corner Wrinkles Using the Five- and Nine-Grade Photoscales

Examiner	5-Grade photoscale		9-Grade photoscale	
	f_0/N_T^*	κ	f_0/N_T^*	κ
YT	86/164	0.407	130/247	0.464
TF	89/164	0.426	114/247	0.397
KT	114/164	0.604	139/247	0.507
AO	84/164	0.400	108/247	0.361
HO	122/164	0.657	83/247	0.252

* Number of scores in agreement with the consensus panel (f_0) divided by total number (N_T).

Table V
Reproducibility of the Nine-Grade Photoscale Scores of Eye-Corner Wrinkles of 247 Subjects

Examiner	Exp. 1		Exp. 2	
	f_0/N_T^*	κ	f_0/N_T^*	κ
YT	130/247	0.464	122/247	0.425
TF	114/247	0.397	113/247	0.392
KT	139/247	0.507	127/247	0.452
AO	108/247	0.361	132/247	0.474
HO	83/247	0.252	99/247	0.327

* Number of scores in agreement with the consensus panel (f_0) divided by total number (N_T).

AGE-RELATED CHANGES IN THE WRINKLE SCORES AT EACH FACIAL SITE

The age-related changes in the five-grade scores at each facial site are shown in Table VIII. The correlation coefficient and the slope of the linear approximate expression were calculated. The sites showing a high correlation with age were as follows, in the order shown: eye corners > lower eyelids > upper eyelids > cheeks > forehead > mouth angles > nasolabial grooves > glabella. The relationships between age and the five grades at representative sites (eye corners and cheeks) are shown in Figure 3 and Figure 4, respectively. The slope of the linear approximate equation at each site is shown in Figure

Table VI
Distribution of Forehead Wrinkle Scores According to the Five-Grade Scale

Wrinkle grade	n (%)
1	54 (32.9)
2	24 (14.6)
3	35 (21.3)
4	34 (20.7)
5	17 (10.4)
Total	164

Table VII
Consistency of Agreement Regarding the Eye-Corner Wrinkles and Forehead Wrinkles of 164 Subjects, Scored by Five General Examiners

Examiner	Eye corners		Forehead	
	f_0/N_T^*	κ	f_0/N_T^*	κ
YT	86/164	0.407	79/164	0.328
TF	89/164	0.426	79/164	0.322
KT	114/164	0.604	93/164	0.441
AO	84/164	0.400	87/164	0.391
HO	122/164	0.657	79/164	0.332

* Number of scores in agreement with the consensus panel (f_0) divided by total number (N_T).

5, and the ages at which the wrinkle score obtained from this slope became 3 are shown in Figure 6. The slope was high, in the order of: eye corners > forehead > lower eyelids > upper eyelids > mouth angles > nasolabial grooves > cheeks > glabella. The ages at which the score became 3 were the early 30s for the eye corners and lower eyelids, the late 40s for the upper eyelids, nasolabial grooves, and forehead, the late 40s for the glabella and the mouth angles, and the middle 50s for the cheeks.

CLINICAL TEST OF WRINKLE-IMPROVING AGENT

The demographic details of the subjects in the clinical test are shown in Table IX. Their mean age was 42.4 years, and scores >6 on the nine-grade scale were observed in 33%. The results of the clinical test are shown in Table X. The wrinkle scores at the eye corners were significantly improved after the use of the agent for eight weeks.

DISCUSSION

In a preliminary study, the results using the photo scales were compared with those using descriptive standards, and the usefulness of the photo scales was confirmed (data not shown). Griffiths *et al.* (8) evaluated photodamage according to photo scales and descriptive standards and found the kappa value significantly higher for the former (0.31) compared to the latter (0.11), suggesting the usefulness of photo scales. The results of our preliminary evaluation were consistent with their findings. Here, a five-grade photo scale was compared with a nine-grade scale, and the former scale was

Table VIII
Mean Wrinkle Grade Variation With Age (decade) at Several Facial Sites

Age-group (years)	No. in group	Mean wrinkle standard grade ± SD							
		Eye corners	Glabella	Mouth angles	Nasolabial groove	Upper eyelid	Forehead	Lower eyelid	Cheek
17-20	42	1.33 ± 0.62	1.67 ± 1.26	1.14 ± 0.47	1.71 ± 0.38	1.50 ± 0.59	1.36 ± 0.62	1.48 ± 0.59	1.05 ± 0.22
21-30	61	1.60 ± 0.89	1.95 ± 1.18	1.15 ± 0.40	1.75 ± 0.77	1.72 ± 0.76	1.43 ± 0.69***	1.78 ± 0.74*	1.02 ± 0.29
31-40	79	3.34 ± 0.78***	2.56 ± 1.22***	2.06 ± 0.91***	2.85 ± 0.85***	2.68 ± 0.67***	2.37 ± 0.98***	2.86 ± 0.78***	1.82 ± 0.76***
41-50	83	4.00 ± 0.62***	3.11 ± 1.07***	2.53 ± 0.97***	3.25 ± 1.05***	3.31 ± 0.90***	3.23 ± 1.02***	3.58 ± 0.83***	2.35 ± 0.77***
51-60	72	4.33 ± 0.53***	3.50 ± 1.11***	3.38 ± 0.90***	3.85 ± 0.85***	4.11 ± 0.59***	3.96 ± 0.86***	4.13 ± 0.58***	3.00 ± 0.95***
61-76	29	4.55 ± 0.69***	2.97 ± 1.50***	4.34 ± 0.72***	4.48 ± 0.63***	4.66 ± 0.81***	4.59 ± 0.63***	4.72 ± 0.53***	3.76 ± 0.87***
Average	366	3.26 ± 1.36	2.69 ± 1.33	2.35 ± 1.26	2.95 ± 1.23	2.96 ± 1.25	2.77 ± 1.37	3.08 ± 1.25	2.10 ± 1.11
Correlation coefficient									
Linear	366	0.825	0.413	0.796	0.738	0.844	0.802	0.852	0.795
Logarithmic	366	0.859	0.434	0.764	0.732	0.834	0.788	0.854	0.764
Exponential	366	0.794	0.409	0.782	0.703	0.794	0.777	0.803	0.8.03
Slope†	366	0.0784	0.0380	0.0692	0.0628	0.0722	0.0758	0.0734	0.0610

***, Significantly different from 18-20-yr-old age group ($p < 0.005, 0.05$).

† Slope of linear regression line between mean wrinkle standard grade and age.

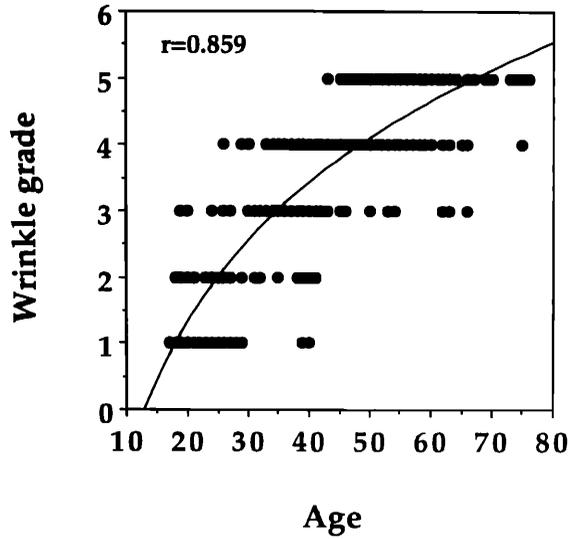


Figure 3. Individual plots of standard wrinkle grades for the eye corners according to age. The standard wrinkle grade for the eye corners was strongly correlated with age ($r = 0.859$, $p < 0.001$).

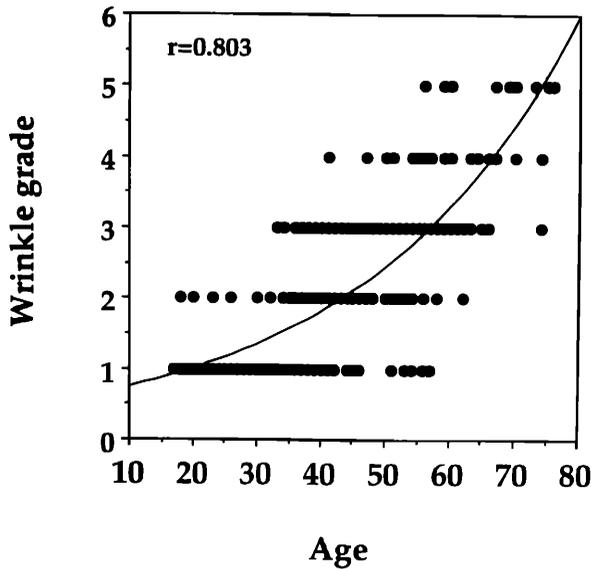


Figure 4. Individual plots of standard wrinkle grade for the cheek according to age. The standard wrinkle grade for the cheek was well-correlated with age ($r = 0.803$, $p < 0.001$).

slightly more consistent with the standard grades. This may be because the probability of consistency was decreased due merely to the greater number of scoring grades in the nine-grade scale. In addition, in the subject group used for the five-grade evaluation, many grade 4 or 5 cases were observed, and their scoring was therefore easy. In our experience, it is desirable to use a scale appropriate to the degree of wrinkles in the group to be evaluated.

Concerning the effects of the facial site (forehead and eye corners) on the scoring, the

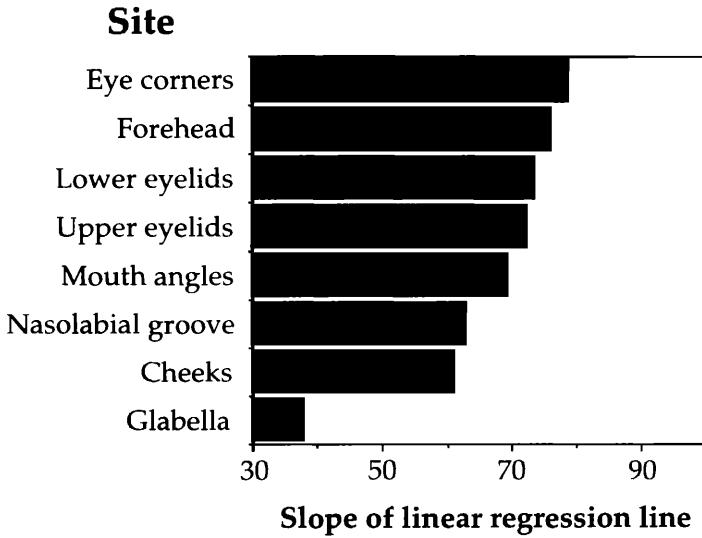


Figure 5. Slope of linear regression line (standard wrinkle grade vs. age) variation for several facial sites.

agreement with the grade standards differed between these sites. This may be because scoring was more difficult at the forehead than at the eye corners, and because there were many subjects with a high score at the eye corners and many with a low score at the forehead even in the same group. As was observed with the nine-grade scale, the scoring of wrinkles of a low degree was difficult.

Griffiths *et al.* (8) and Larnier *et al.* (9) performed a scoring of photoaging using whole-face photos and suggested that photo standards cannot be used for such a purpose for Mongoloid populations in the Far East because pigmentation rather than wrinkles is

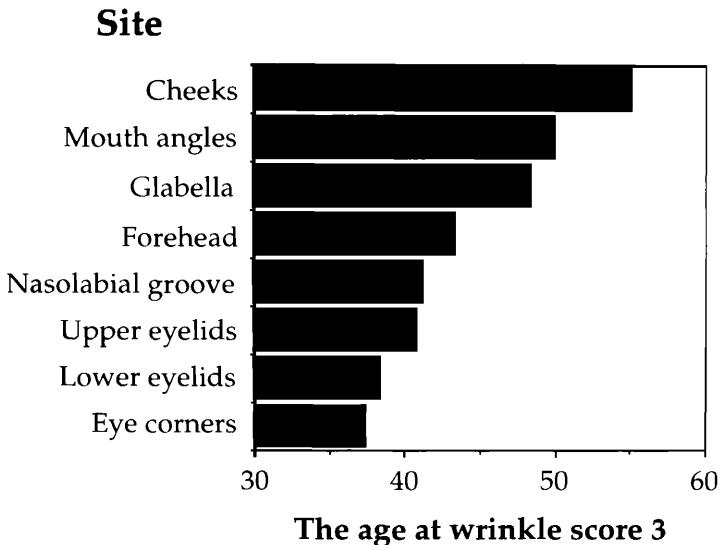


Figure 6. The age at wrinkle score 3 (obtained from the slope of the linear approximate equation).

Table IX
Demographic Characteristics of the Subject Population for the Clinical Test of a
Wrinkle-Improving Agent

Number	15
Age (yr)	
Mean	42.5
Range	35–54
Sex	female
Overall severity of wrinkle*	
Mild	10.0 (67%)
Moderate	5.0 (33%)

* Mild = grade 1–5; moderate = grade 6–8.

Table X
Improvement in Wrinkle Score After an Eight-Week Use of a Wrinkle-Improving Agent

	Eye corners	
	Agent	Placebo
Mean baseline score	5.00	4.93
Mean week-8 score	4.40	5.06
Mean change from baseline	-0.60	0.07
Percent change from baseline	12.0	-1.42
p Value*	<0.01	

The data are means of 15 subjects.

* Significant (t-test) compared with placebo.

more readily affected by photoaging compared with Caucasians. Griffiths *et al.* demonstrated this based on a profile of a 72-year-old obese Japanese and the results of clinical tests of tretinoin in Mongoloids in Singapore, Indonesia, and Malaysia (14). In the present study, we showed the usefulness of enlarged photo scales for evaluating photoaging, especially wrinkles, using enlarged photos of the eye corners. However, the decrement of the kappa value of the photo standards at the forehead compared with the kappa value at the eye corners may support the contention of Griffiths *et al.* that photo scales cannot be used in Japanese.

Concerning age-related changes in the scores at each facial site, wrinkles around the eyes were highly correlated with age. A similar high correlation between changes at the eye corners and age was also observed in a three-dimensional (3-D) analysis of skin replicas (15). However, our previous replica analysis showed higher correlations of age with wrinkles at the glabella, nasolabial grooves, and mouth angles compared to wrinkles at other sites (excluding the eye corners) (15). Thus, the wrinkle score at the eye corners visually differed from the 3-D surface morphology analyzed using replicas.

Since the slope of the linear approximate expression represents the rate of development of wrinkles, wrinkles more rapidly develop in the order of: eye corners > forehead > lower eyelids > upper eyelids > mouth angles > nasolabial grooves > cheeks > glabella. Wrinkles around the eyes that are highly correlated with age also showed a high wrinkle development rate. A wrinkle score of 3 is considered to represent the initiation of the development of wrinkles. According to the ages at which the present subjects showed score 3, wrinkles appear to develop early and rapidly at the eye corners.

We evaluated an agent for wrinkle improvement using a photo scale and obtained good results and found that an evaluation of pre- and post-treatment wrinkles was possible using a wrinkle photo scale, suggesting the validity and usefulness of this method.

REFERENCES

- (1) R. B. Armstrong, J. Lesiewicz, G. Harvey, L. F. Lee, K. T. Spoehr, and M. Zultak, Clinical panel assessment of photodamaged skin treated with isotretinoin using photographs, *Arch. Dermatol.*, **128**, 352–356 (1992).
- (2) M. J. Stiller, J. Bartolone, R. Stern, S. Smith, N. Kollias, R. Gillies, and L. A. Drake, Topical 8% glycolic acid and 8% L-lactic acid creams for the treatment of photodamaged skin, *Arch. Dermatol.*, **132**, 631–636 (1996).
- (3) J. S. Weiss, C. N. Ellis, J. T. Headington, T. Tincoff, T. A. Hamilton, and J. J. Voorhees, Topical tretinoin improves photoaged skin, *JAMA*, **259**, 527–532 (1988).
- (4) J. J. Leyden, G. L. Grove, M. J. Grove, E. G. Thorne, and L. Lufano, Treatment of photodamaged facial skin with topical tretinoin, *J. Am. Acad. Dermatol.*, **21**, 638–644 (1989).
- (5) L. Lever, P. Kumar, and R. Marks, Topical retinoic acid for treatment of solar damage, *Br. J. Dermatol.*, **122**, 91–98 (1990).
- (6) G. L. Grove, M. J. Grove, J. J. Leyden, L. Lufano, B. Schwab, B. H. Perry, and E. G. Thorne, Skin replica analysis of photodamaged skin after therapy with tretinoin emollient cream, *J. Am. Acad. Dermatol.*, **25**, 231–237 (1991).
- (7) C. E. M. Griffiths, S. Kang, C. N. Ellis, K. J. Kim, L. J. Finkel, L. C. Ortiz-Ferrer, G. M. White, T. A. Hamilton, and J. J. Voorhees, Two concentrations of topical tretinoin (retinoic acid) cause similar improvement of photoaging but different degrees of irritation, *Arch. Dermatol.*, **131**, 1037–1044 (1995).
- (8) C. E. M. Griffiths, T. S. Wang, T. A. Hamilton, J. J. Voorhees, and C. N. Ellis, A photometric scale for the assessment of cutaneous photodamage, *Arch. Dermatol.*, **128**, 347–351 (1992).
- (9) C. Larnier, J. P. Ortonne, A. Venot, B. Faivre, J. C. Beani, P. Thomas, T. Brown, and E. Sendagorta, Evaluation of cutaneous photodamage using a photographic scale, *Br. J. Dermatol.*, **130**, 167–173 (1994).
- (10) Y. Takema, Y. Yorimoto, and M. Kawai, The relationship between age-related changes in the physical properties and development of wrinkles in human facial skin, *J. Soc. Cosmet. Chem.*, **46**, 163–173 (1995).
- (11) J. Cohen, A coefficient of agreement for nominal scales, *Educ. Psychol. Measurement*, **20**, 37–46 (1960).
- (12) J. D. Fleiss and J. Cohen, Large sample standard errors of kappa and weighted kappa, *Psychol. Bull.*, **72**, 323–327 (1969).
- (13) J. R. Landis and G. G. Koch, The measurement of observer agreement for categorical data, *Biometrics*, **33**, 159–174 (1977).
- (14) S. H. Goh, The treatment of visible signs of senescence: The Asian experience, *Br. J. Dermatol.*, **122**(Suppl. 35), 105–109 (1990).
- (15) Y. Takema, K. Tsukahara, T. Fujimura, and M. Hattori, Age-related changes of 3-dimensional morphological changes in the human facial skin, *Skin Res. Tech.*, **3**, 95–100 (1997).