Development of a Photographic Scale for Consistency and Guidance in Dermatologic Assessment of Forearm Sun Damage

Naja E. McKenzie, PhD, RN; Kathylynn Saboda, MS; Laura D. Duckett, CCRS; Rayna Goldman, MBA; Chengcheng Hu, PhD; Clara N. Curiel-Lewandrowski, MD

Objectives: To develop a photographic sun damage assessment scale for forearm skin and test its feasibility and utility for consistent classification of sun damage.

Design: For a blinded comparison, 96 standardized 8×10 digital photographs of participants' forearms were taken. Photographs were graded by an expert dermatologist using an existing 9-category dermatologic assessment scoring scale until all categories contained photographs representative of each of 4 clinical signs. Triplicate photographs were provided in identical image sets to 5 community dermatologists for blinded rating using the dermatologic assessment scoring scale.

Setting: Academic skin cancer prevention clinic with high-level experience in assessment of sun-damaged skin.

Participants: Volunteer sample including participants from screenings, chemoprevention, and/or biomarker studies.

Main Outcome Measures: Reproducibility and agreement of grading among dermatologists by Spearman correlation coefficient to assess the correlation of scores given for the same photograph, κ statistics for ordinal data, and variability of scoring among dermatologists, using analysis of variance models with evaluating physician and photographs as main effects and interaction effect variables to account for the difference in scoring among dermatologists.

Results: Correlations (73% to >90%) between dermatologists were all statistically significant ($P < .001$). Scores showed good to substantial agreement but were significantly different ($P < .001$) for each of 4 clinical signs and the difference varied significantly ($P < .001$) among photographs.

Conclusions: With good to substantial agreement, we found the development of a photographic forearm sun damage assessment scale highly feasible. In view of significantly different rating scores, a photographic reference for assessment of sun damage is also necessary.

Arch Dermatol. 2011;147(1):31-36

Author Affiliations: College of Nursing (Dr McKenzie), Skin Cancer Prevention Annex (Mss Saboda, Duckett, and Goldman), Mel and Enid Zuckerman College of Public Health (Dr Hu), and College of Medicine (Dr Curiel-Lewandrowski), Arizona Cancer Center, University of Arizona, Tucson.

T HE QUEST FOR CONSISTENCY in clinical assessment of sun damage has led to the development of objective grading methods for characterization and quantification of sun damage. The methods include descriptive,[1,2] visual analog,[3,4] and photographic grading scales.[2,5] Published scales have been for facial assessment only, but when skin biopsies are required, forearms are preferable rather than cosmetically sensitive facial areas.

Weiss and colleagues[2] developed a descriptive scale for the assessment of overall cutaneous photoaging to be used along with facial photographic samples but did not discuss agreement or validity. The R.W. Johnson Pharmaceutical Research Institute descriptive scale[1] achieved a chance-corrected agreement (κ coefficient) of 0.11. Dermatologic research protocols rely on consistent clinical identification, description, and quantification of sun damage in forearm skin. To date, no valid and reliable photographic assessment scale of forearm skin sun damage has been developed.

The clinical assessment of human skin for sun damage is a highly subjective but vital part of evaluating the effectiveness of agents and interventions for their ability to reduce or reverse sun damage. Since histopathologic evaluation is a regulatory requirement along with clinical evaluation to assess safety and efficacy of test articles, biopsied tissue must be obtained. Human subjects considerations suggest that forearm skin, rather than facial skin,

Clinical Sign	Absent	Mild	Moderate	Severe
Fine wrinkling	0	1 2 3	4 5 6	7 8 9
Coarse wrinkling	0	1 2 3	4 5 6	7 8 9
Abnormal pigmentation	0	1 2 3	4 5 6	7 8 9
Global	0	1 2 3	4 5 6	7 8 9

Figure 1. Dermatologic Assessment Form Forearm Photographic Assessment Scale.

should be used for this purpose. This consideration alone makes an objective grading scale for forearm skin essential. Furthermore, a standardized teaching set will be valuable for developing a reproducible method and can support the comparison of findings from a variety of studies. The objective of this study was to begin the development of a consistent photographic assessment scale of sun damage in forearm skin, complemented with a descriptive scale, that can become a criterion standard in dermatologic studies. This study is the first step toward this objective.

METHODS

A criterion standard is a performance standard with which experts or peers agree and with which individual practice can be compared.[6] Establishing such a criterion standard requires a strong empirical relationship between the scale and the variable it represents.[7]

Forearm photodamage assessment in current studies[8-10] is performed using a subjective 10-point scale for each of 4 clinical signs of UV-induced skin damage: fine wrinkling, coarse wrinkling, abnormal pigmentation, and a global assessment. The global assessment is used to give an overall impression of sun damage. Each clinical sign is ranked and subdivided as follows: absent (0), mild (1-3), moderate (4-6), and severe (7-9). This approach is similar to the R.W. Johnson Pharmaceutical Research Institute descriptive scale[1,11] which is used for assessment of photodamage in facial skin. Our scale, the Dermatologic Assessment Form Forearm Photographic Assessment Scale, is presented in **Figure 1**.

PARTICIPANTS

In the spring and summer of 2007, a total of 48 adults (26 women [54.2%] with a mean age of 52 years and 22 men [45.8%] with a mean age of 63 years) were recruited for this study. Participants identified themselves as white (n=47) or African American (n=1). Participants further identified themselves as Hispanic (n=6) or non-Hispanic (n=36); 6 did not provide any ethnic identification. The sample included community volunteers and participants taking part in screenings and clinical studies. Individuals whose dorsal forearms were unsuitable for use in a photographic scale, including those with significant inflammation or irritation, tattoos, or other markings, were not eligible. Individuals on the extremes—almost no sun damage and very severe sun damage—had to be sought by referral and invited to participate.

One academic physician (C.C.-L.) and 5 community dermatologists agreed to assist with the study as raters. Of these, the academic physician was designated as the project's expert

Table 1. Distribution of Reference Standard Initial Grading by Category and Clinical Sign

Category	Level	No.			
		Fine Wrinkling	Coarse Wrinkling	Abnormal Pigmentation	Global Assessment
None	0	2	9	5	6
Low	1	12	5	12	10
	2	7	11	7	8
	3	11	8	10	9
Moderate	4	6	5	4	5
	5	10	14	13	14
	6	21	10	19	17
Severe	7	15	19	13	15
	8	9	7	10	6
	9	3	8	3	6

dermatologist and reference standard. This dermatologist is the primary study physician leading our clinical trials and therefore has the most experience assessing skin photodamage involving the forearm. This physician's initial grading was designated as the reference standard for subsequent gradings using the photographic scale.

ETHICAL APPROVAL AND INFORMED CONSENT

This study was approved by the institutional review board of the University of Arizona, which has a Federalwide Assurance with the US Office of Human Research Protections and functions under a Statement of Compliance. All participants provided signed informed consent.

DIGITAL PHOTOGRAPHY

Digital photographs were taken of the dorsal forearms from knuckle (metacarpal-phalangeal) to elbow to avoid personal identification. Both forearms of each of the 48 participants were photographed, for a total of 96 unique photographs. A Nikon COOLPIX 4300 digital camera (Nikon, Tokyo, Japan) was used with standardized methods to ensure consistency. Standardized lighting consisted of available overhead lighting in a windowless studio with no separate skin illumination. The Anytime Flash setting was used with maximum aperture (preset between 2.8 and 7.6), and all photographs were taken on a uniform blue background. Additional settings included image size, 2272 × 1704; image quality, fine; focus, macro close-up automatic single mode; and sensitivity, 100 ISO. The focal length of the COOLPIX lens system is 8 to 24 mm.

The expert dermatologist scored the photographs by clinical sign using our existing clinical sun damage assessment scale until all score categories were saturated for each clinical sign.

RANDOMIZATION AND GRADING

Each photograph was printed unedited in triplicate, coded, and paired with a blank dermatologic assessment scale form (Figure 1). The expert dermatologist performed the initial grading of the photographs, thus establishing our reference standard for comparison (**Table 1**). The triplicate image sets, consisting of 288 photographic pages, were randomly ordered in binders and delivered to the 5 evaluating dermatologists. They each blindly evaluated the 96 unique photographs 3 times. Finally, the dermatologist designated as the reference standard repeated evaluation of the randomized set of photographs.

Table 2. Correlation of Reference Standard to Repeated Screening by Expert Dermatologist at 3 Months

Dermatologic Assessment Clinical Sign	Spearman ρ Correlation Coefficients		
	Image Set 1	Image Set 2	Image Set 3
Fine wrinkling	0.87	0.92	0.91
Coarse wrinkling	0.92	0.91	0.91
Abnormal pigmentation	0.91	0.90	0.91
Global assessment	0.92	0.93	0.93

Table 3. Correlation of Reference Standard to Community Dermatologists[a]

Dermatologic Assessment Form Criteria	Spearman Correlation Coefficients		
	Set 1	Set 2	Set 3
Fine wrinkling			
Dermatologist			
B	0.79	0.87	0.87
C	0.88	0.89	0.88
D	0.71	0.69	0.74
E	0.81	0.81	0.83
F	0.74	0.86	0.86
Coarse wrinkling			
Dermatologist			
B	0.90	0.89	0.91
C	0.92	0.93	0.91
D	0.82	0.83	0.85
E	0.82	0.83	0.87
F	0.86	0.85	0.88
Abnormal pigmentation			
Dermatologist			
B	0.88	0.86	0.92
C	0.91	0.91	0.89
D	0.89	0.89	0.89
E	0.86	0.92	0.90
F	0.89	0.85	0.91
Global assessment			
Dermatologist			
B	0.90	0.90	0.93
C	0.92	0.92	0.92
D	0.90	0.90	0.91
E	0.86	0.85	0.89
F	0.90	0.88	0.92

[a] All correlations are statistically significant at P<.001.

Table 4. κ Statistics by Clinical Sign for Average Rater Specific Agreement vs Reference Standard[a]

Dermatologists	Agreement	κ (95% Confidence Interval)
Fine Wrinkling		
A	92.1	0.76 (0.68-0.79)
B	90.3	0.69 (0.65-0.70)
C	90.7	0.71 (0.67-0.75)
D	71.4	0.28 (0.26-0.35)
E	77.3	0.37 (0.31-0.43)
F	90.2	0.68 (0.63-0.71)
Coarse Wrinkling		
A	91.7	0.76 (0.70-0.80)
B	88.4	0.70 (0.64-0.72)
C	91.0	0.76 (0.74-0.80)
D	71.1	0.29 (0.24-0.35)
E	86.3	0.61 (0.58-0.68)
F	89.5	0.72 (0.62-0.75)
Abnormal Pigmentation		
A	92.2	0.76 (0.74-0.79)
B	90.0	0.71 (0.68-0.76)
C	91.7	0.76 (0.73-0.79)
D	81.0	0.47 (0.41-0.55)
E	88.6	0.66 (0.63-0.69)
F	89.3	0.70 (0.63-0.75)
Global Assessment		
A	92.4	0.77 (0.72-0.78)
B	91.2	0.75 (0.69-0.76)
C	91.9	0.76 (0.74-0.80)
D	81.8	0.50 (0.44-0.54)
E	87.1	0.64 (0.55-0.69)
F	90.2	0.70 (0.69-0.75)

[a] A is the reference standard dermatologist; B through F, the community dermatologists.

Two dermatologists were able to review only 287 photographs due to a missing image at the time of the evaluation. The remaining dermatologists evaluated the complete set of 288 photographs. The 2 missed evaluations were treated as missing data and imputed using an average of available data for the same reviewer and photograph.

ANALYSIS AND RESULTS

The nonparametric Spearman ρ was used to study the correlation of all scores given for each photograph. Analysis of variance models with random effects were used to study the difference in scores by different dermatologists. All analyses, random ordering, and graphs were carried out in Stata version 10 (StataCorp LP, College Station, Texas).

We first analyzed the relationship among the 4 scores given to each photograph by the expert dermatologist (when setting the reference standard and as assessments 3 months later). **Table 2** summarizes the Spearman ρ correlation coefficients. The expert dermatologist's assessments of the same photographs over time were highly and significantly correlated near or above 90% for all 4 clinical signs.

The correlation between the expert dermatologist's assessment and the scores given by the 5 community dermatologists ranged from 73% to above 90% (**Table 3**) and were all statistically significant (P< .001). These results show that assessments by all dermatologists had a strong linear relationship with the reference standard scores. However, strongly correlated scores can be quite different in magnitude and ultimately fail to show agreement.[12] Therefore, to quantify agreement among the community dermatologists and the reference standard, we calculated the κ statistic for ordinal data. Calculation of κ statistic is based on the ratio of the observed to the expected (ie, by chance) agreement. All κ statistics (**Table 4**) fell between 0.28 and 0.76. Guidelines for interpretation of κ vary. Landis and Koch[13] would categorize 0.28 as "fair" and 0.76 as "substantial." Percentage of agreement among raters, calculated as part of the κ statistic (Table 4), showed that raters agreed with the reference standard 71% to 92% of the time. The highest percent agreement was between the original and final, blinded rating session of the expert dermatologist.

Figure 2 shows the distribution of maximum deviation from the reference standard for each dermatologist and each clinical sign. *Deviation* is defined as the difference between a given score and the reference standard, and the maximum deviation is the one with the greatest magnitude (positive or negative)
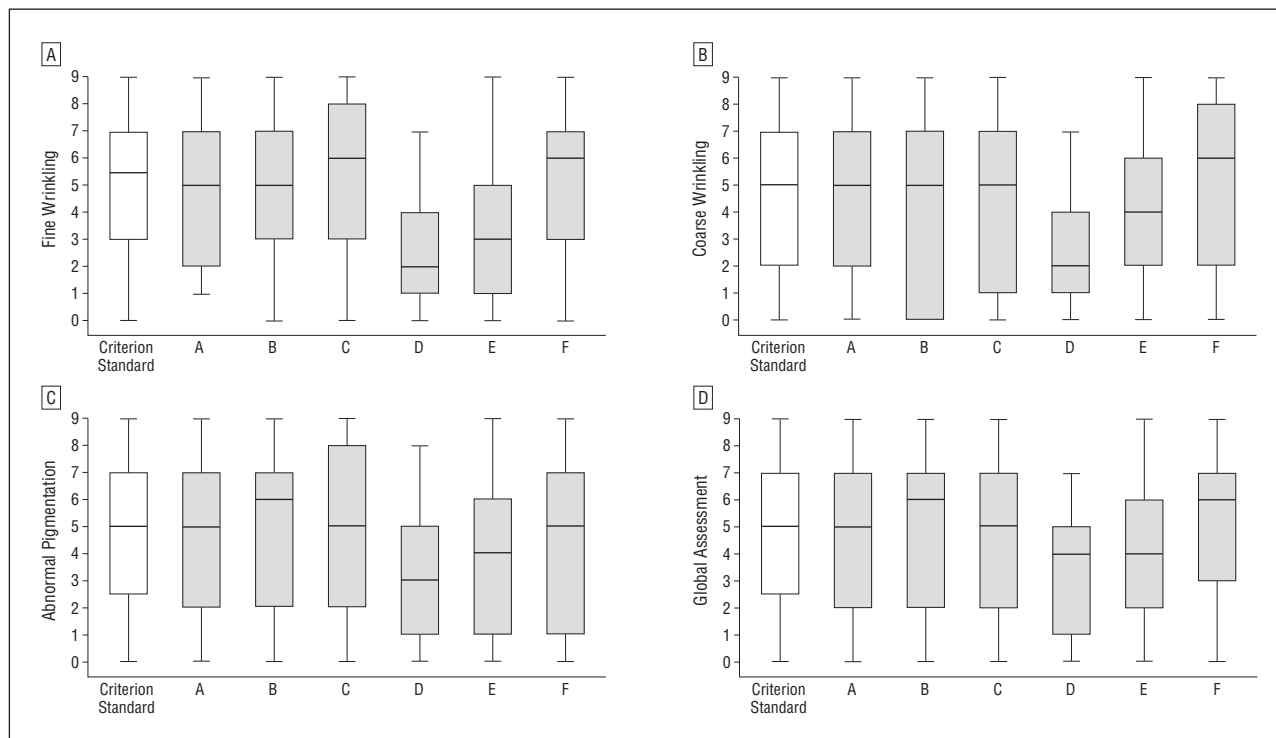
Figure 2. A, Distribution of maximum fine wrinkling scoring deviation from the reference standard for each dermatologist (A is the reference standard dermatologist; B-F are the community dermatologists). B, Distribution of maximum coarse wrinkling scoring deviation from the reference standard for each dermatologist. C, Distribution of maximum abnormal pigmentation scoring deviation from the reference standard for each dermatologist. D, Distribution of maximum global assessment scoring deviation from the reference standard for each dermatologist.

among the 3 scores for each photograph. Here, a deviation of ±3 was not rare and could exceed 5.

We used 2-way analysis of variance to examine the dermatologist effect and the photograph effect. All of the expert dermatologist's assessments were excluded from the data to avoid potential bias. Analysis of variance indicated that the scores given by the 5 remaining dermatologists were significantly different ($P < .001$) for each of the 4 clinical signs, and the differences tended to vary among photographs ($P < .001$).

<div align="center">

COMMENT

</div>

Current clinical protocols rely on consistent clinical assessment of sun damage in forearm skin to evaluate baseline and efficacy. To date, no valid and reliable photographic assessment scale of forearm skin sun damage has been developed. The purpose of this study was to develop and test a forearm photographic assessment scale that can be used to ensure such consistency when adopted by study dermatologists who are required to clinically assess photodamage. We plan to subject the scale to expanded testing in order to propose this scale as a criterion standard for general use in dermatologic studies. Weiss and colleagues,[2] in studying the effect of topical tretinoin, used a paper scale that included clinical signs for the assessment of overall improvement in cutaneous photoaging of the face to be used along with photographic samples, but they did not discuss agreement or validity. Griffiths and colleagues[1] developed a photonumeric scale that included the most common features of interest in the evaluation of photodamage of facial skin. The R.W. Johnson Pharmaceutical Research Institute de-

scriptive scale[1] included a detailed description of the manifestations of sun damage with a chance-corrected agreement (κ coefficient) of 0.11 without, and 0.31 with, accompanying facial photographs. Chance-corrected agreement ranges from −1 to +1, with scores of 0.40 to 0.75 considered fair and greater than 0.75 considered excellent or substantial.[13,14] This scale is similar to our clinical assessment scale, but for facial skin. On our scale, hyperpigmentation and mottling have been combined into a single clinical sign and renamed abnormal pigmentation because, in the opinion of all of our principal investigators, pigmentation is difficult to separate into 2 different features.

Visual analog scales rely on health care practioners to estimate features visually on a metrically defined horizontal line. Developers of such scales for assessment of sun damage[3,4] have described them as more sensitive than descriptive scales and highly reproducible, but they have not reported chance-corrected agreement or repeatability. Our 10-step clinical assessment scale consists of 3 levels of severity: mild, moderate, and severe. Each of these is subdivided into 3 numerical grades, allowing for a more nuanced scale not unlike a visual analog scale.

Photographic scales have the advantage of providing a consistent visual frame of reference, thus minimizing variability in perception and subjectivity. The photographic scale of Larnier et al[5] consists of a set of 3 standardized photographs to represent each of 6 grades of sun damage, ranging from mild to very severe. The photographs were taken in a standard manner, from the same angle and of the same side (left) and region of the

Figure 3. Global assessment: severity score 0 (A); 1 (B); 2 (C); 3 (D); 4 (E); 5 (F); 6 (G); 7 (H); 8 (I); and 9 (J).

face. On assessment of interobserver agreement, chance-corrected κ scores ranged from 0.44 to 0.76 on the first and second occasions. In addition, dermatologists with and without experience with sun-damaged skin scored similarly, supporting the notion that a photographic scale increases objectivity and standardization. Testing of our scale achieved similar or better interobserver agreement using blinded image sets. **Figure 3** shows the global assessment photographs with the best agreement.

An upper-extremity photonumeric scale was developed to assess skin aging in smokers and nonsmokers on the protected upper inner arm.[15] The scale was effective in showing greater skin aging in smokers than nonsmokers. Efficacy and safety of a topical agent were evaluated using a photographic method consisting of baseline and repeated side-by-side projection of before-and-after images during 36 weeks of treatment,[16] but the standard was relative and relevant only to that study. The quality of digital photography has improved greatly since the original description of the photographic method,[17] justifying the establishment of an absolute standard for photodamage in forearm skin.

Photographic evaluation of photodamage improvement has also been used in laser resurfacing and remodeling[18]; however, the photographs were facial and therefore not applicable to our scale. Forearm skin is also used to establish combination laser procedures before clinical use and the availability of a forearm skin scale may be useful in nonpharmaceutical approaches to photodamage.

Shoshani and colleagues[19] made a case for a clinically validated scale for the assessment of facial wrinkling. We

propose that a forearm scale is equally necessary. Our findings support the ability of blinded, independent dermatologists to achieve good to excellent agreement and strong linear correlation among their scores as well as internal consistency of ratings, all at a level of high statistical significance. Nevertheless, there were differences in how the dermatologists rated the photographs. All dermatologists in this study have similar years of experience and we cannot immediately explain the differences in how the community dermatologists rated the photographs, although one of them sees primarily a retiree population and did rate the photographs less severely. The size of maximum differences may be related to the type of patients typically seen in the practices of the community dermatologists. However, even without training, our dermatologists achieved high agreement and significant correlation in how they rated the photographs. The high percentage of agreement testifies to the potential for improvement in consistency with training among dermatologists for whom agreement is vital.

The inability of our photographic scale to account directly for hyperkeratotic features, for both extension of skin surface involvement and thickness, must be acknowledged as a limitation of our study. The next phase of scale development will include an objective form of categorical validation, such as optical coherence tomography[20] or microscopy. We also acknowledge that the composition of our sample with regard to race and ethnicity, being mainly white, may limit generalization across all populations. Our sample heterogeneity is representative of our US and local populations; however, it will be expanded in the next phase of scale development.

CONCLUSIONS

Based on these results, the expanded Dermatologic Assessment Form Forearm Photographic Assessment Scale has great potential to yield highly consistent scoring of forearm sun damage in study participants. Further steps are needed to create a training image set that can be considered the criterion standard for forearm sun damage.

Correspondence: Naja E. McKenzie, PhD, RN, Arizona Cancer Center, University of Arizona, PO Box 245024, 1515 N Campbell Ave, Tucson, AZ 85724-5024 (nmckenzie@azcc.arizona.edu).

REFERENCES

1. Griffiths CE, Wang TS, Hamilton TA, Voorhees JJ, Ellis CN. A photonumeric scale for the assessment of cutaneous photodamage. *Arch Dermatol*. 1992;128(3):347-351.
2. Weiss JS, Ellis CN, Goldfarb MT, Voorhees JJ. Tretinoin therapy: practical aspects of evaluation and treatment. *J Int Med Res*. 1990;18(3)(suppl 3):41C-48C.
3. Lever L, Kumar P, Marks R. Topical retinoic acid for treatment of solar damage. *Br J Dermatol*. 1990;122(1):91-98.
4. Marks R, Edwards C. The measurement of photodamage. *Br J Dermatol*. 1992;127(suppl 41):7-13.
5. Larnier C, Ortonne JP, Venot A, et al. Evaluation of cutaneous photodamage using a photographic scale. *Br J Dermatol*. 1994;130(2):167-173.
6. *JAMA* Instructions for Authors. *JAMA*. 1995;273(1):27-28.
7. DeVellis RF. *Scale Development*. Newbury Park, CA: Sage; 1991.
8. Einspahr JG, Bowden GT, Alberts DS, et al. Cross-validation of murine UV signal transduction pathways in human skin. *Photochem Photobiol*. 2008;84(2):463-476.
9. Stratton SP, Alberts DS, Einspahr JG, et al. A phase 2a study of topical perillyl alcohol cream for chemoprevention of skin cancer. *Cancer Prev Res (Phila)*. 2010;3(2):160-169.
10. Stratton SP, Saboda KL, Myrdal PB, et al. Phase 1 study of topical perillyl alcohol cream for chemoprevention of skin cancer. *Nutr Cancer*. 2008;60(3):325-330.
11. Olsen EA, Katz HI, Levine N, et al. Sustained improvement in photodamaged skin with reduced tretinoin emollient cream treatment regimen: effect of once-weekly and three-times-weekly applications. *J Am Acad Dermatol*. 1997;37(2, pt 1):227-230.
12. Jakobsson U, Westergren A. Statistical methods for assessing agreement for ordinal data. *Scand J Caring Sci*. 2005;19(4):427-431.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
14. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977;33(2):363-374.
15. Helfrich YR, Yu L, Ofori A, et al. Effect of smoking on aging of photoprotected skin: evidence gathered using a new photonumeric scale [published correction appears in Arch Dermatol. 2007;143(5):633]. *Arch Dermatol*. 2007;143(3):397-402.
16. Maddin S, Lauharanta J, Agache P, Burrows L, Zultak M, Bulger L. Isotretinoin improves the appearance of photodamaged skin: results of a 36-week, multicenter, double-blind, placebo-controlled trial. *J Am Acad Dermatol*. 2000;42(1, pt 1):56-63.
17. Armstrong RB, Lesiewicz J, Harvey G, Lee LF, Spoehr KT, Zultak M. Clinical panel assessment of photodamaged skin treated with isotretinoin using photographs. *Arch Dermatol*. 1992;128(3):352-356.
18. Marini L. SPF-RR sequential photothermal fractional resurfacing and remodeling with the variable pulse Er:YAG laser and scanner-assisted Nd:YAG laser. *J Cosmet Laser Ther*. 2009;11(4):202-211.
19. Shoshani D, Markovitz E, Monstrey SJ, Narins DJ. The modified Fitzpatrick Wrinkle Scale: a clinical validated measurement tool for nasolabial wrinkle severity assessment. *Dermatol Surg*. 2008;34(suppl 1):S85-S91.
20. Korde VR, Bonnema GT, Xu W, et al. Using optical coherence tomography to evaluate skin sun damage and precancer. *Lasers Surg Med*. 2007;39(9):687-695.