Ann Dermatol Vol. 27, No. 5, 2015 http://dx.doi.org/10.5021/ad.2015.27.5.557

ORIGINAL ARTICLE

Reduction of Inter-Rater and Intra-Rater Variability in Psoriasis Area and Severity Index Assessment by Photographic Training

Sang Woong Youn, Chong Won Choi, Bo Ri Kim, Je Byeong Chae

Department of Dermatology, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea

Background: Severity grading is important for the assessment of psoriasis treatment efficacy. This is most commonly achieved by using the psoriasis area and severity index (PASI), a subjective tool with inherent inter-rater and intra-rater variability. PASI-naive dermatologists require training to properly conduct a PASI assessment. **Objective:** In the present study, we aimed to investigate whether photographic training improves inter-rater and intra-rater variabilities. We also determined which PASI component has the greatest impact on variability. **Methods:** Twenty-one dermatologists received 1 hour of PASI training. They were tested before and after the training to evaluate intra-rater variability. The physicians were further tested after training by using a reference photograph. **Results:** The mean of each PASI component was underevaluated compared with scoring by a PASI expert. The concordance rate with the expert's grading was highest for thickness followed by erythema, scaling, and area. The scaling score showed the greatest improvement after training. After training, the distribution of deviation from the expert's grading, which signifies inter-rater variability, improved only for the PASI area component. The deviation of scaling grading improved upon retesting by using a reference photograph. **Conclusion:** PASI assessment training improved variabilities to some degree but not for every PASI component. The development of an objective psoriasis severity assessment tool will help overcome the subjective variabilities in PASI assessment, which can never be completely eliminated via training. **(Ann Dermatol 27(5) 557~562, 2015)**

-Keywords-
Body surface area, Education, Psoriasis, Severity of illness index

INTRODUCTION

Psoriasis is a chronic, relapsing inflammatory skin disorder treated with various modalities according to individual disease severity. It is important to use standardized severity assessment tools to evaluate static and dynamic severity before and after treatment. In addition, the recent introduction of biologics for the treatment of psoriasis strengthens the need for standardized methods of assessing the severity of psoriasis[1,2]. The psoriasis area and severity index (PASI) is the most commonly used assessment tool for psoriasis[3,4]. The PASI is calculated by evaluating the following four components: erythema, induration, scaling, and the involved area of psoriatic lesions from four sections of the body, namely the head, upper extremities, lower extremities, and trunk[3]. Although the PASI is a widely used and convenient method that can be performed without instruments, its inherent subjectivity, which results in inter-rater and even intra-rater variability, is the main drawback of this method[1,3]. Furthermore, the PASI is calculated by using a complex equation without the aid of an automated calculating tool, which leads to calculation errors. The PASI also does not take into account the psychological burdens of psoriatic lesions located on exposed areas such as the face, hands, and nails[1]. Mainly, how-

Received May 27, 2015, Revised June 17, 2015, Accepted for publication June 22, 2015

Corresponding author: Sang Woong Youn, Department of Dermatology, Seoul National University Bundang Hospital, Seoul National University College of Medicine, 82 Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Korea. Tel: 82-31-787-7319, Fax: 82-31-787-4058, E-mail: swyoun@snu.ac.kr

ever, each of the PASI components is measured according to an evaluator's subjective assessment. Thus, the PASI can be inaccurate inter-individually and even show time-to-time variability in a single rater[3,5]. Therefore, expert training is required to minimize inter-rater and intra-rater variability in PASI assessment[6]. However, in actuality, most clinics do not provide residents with PASI assessment training.
In this study, we elucidated the effects of PASI training on variability by recruiting dermatology residents and young board-certified dermatologists with no PASI assessment experience. We evaluated the efficacy of an in-person training course on PASI assessment accuracy before training, after training without reference photographs, and after training by using reference photographs. In addition, each PASI component score was analyzed to determine the effects of individual components on variability.

MATERIALS AND METHODS

Subjects and study design

Young board-certified dermatologists (<5 years' experience) and residents who were naive to PASI assessment received a 1-hour training lecture. The training course consisted of a severity grading method of each PASI component and a practical PASI scoring system. During the lecture, a questionnaire with three successive visual photographic questions regarding PASI elements was given to the audience. The institutional review board of the Seoul National University Bundang Hospital approved this study (IRB No. X-1503-292-901).

PASI assessment training and evaluation of training efficacy

The subjects were tested before and after the in-person training course to evaluate the impact of training on PASI assessment accuracy. First, the subjects were required to assess erythema, induration, scaling, and the involved area of psoriatic lesions on a photograph (pretest). The subjects then received 1 hour of training by a psoriasis specialist regarding the basic concepts of PASI assessment, including grading the erythema, thickness, and scaling components and the method of estimating the affected area. To enhance the proficiency of the in-person training course, representative photographs of each grade of erythema, induration, scaling, and involved area of psoriatic lesions were presented. After completing the in-person training course, the subjects were required to assess the PASI components of psoriatic lesions by using the same photograph (retest). In addition, the subjects were required to assess the PASI components in reference photographs provided by the lecturer for the same psoriasis case (retest with reference photographs). The scores of the board-certified dermatologists and dermatology residents were compared with those of the PASI assessment expert to determine their concordance rate. The distribution of deviations between the expert's grading and the subjects' grading was analyzed by using the following equation:

Deviation from the expert's grading = the absolute value of [the expert's grading-the subject's grading]

For example, if the expert's erythema grading is 3, the deviation is 1 if the resident's grading is 4 or 2. This is another approach to the evaluation of inter-rater variability and the quality of training. Finally, additional questionnaires covering PASI assessment behavior were administered (e.g., what is the most difficult component to grade in the PASI assessment?).

Statistical analyses

Differences in the severity grading of each PASI component in the dermatologist and resident groups were compared by using the Mann-Whitney U test. Post-lecture changes were analyzed by using the Wilcoxon signed-rank test. For inter-rater variability, concordance rates with the PASI expert before and after training were compared simply without using a statistical method, but the expert's grading distribution for each test was analyzed by using the Wilcoxon signed-rank test. Statistical analyses were performed by using SPSS Statistics ver. 17.0 statistical software (SPSS Inc., Chicago, IL, USA). A $p<0.05$ was considered statistically significant.

RESULTS

Subjects' characteristics

The work experience of the 21 participants in dermatologic practice was distributed as follows: nine participants (42.9%) had 1 to 2 years of experience, three (14.3%) had 3 to 4 years of experience, six (28.6%) had 5 to 8 years of experience, and three (14.3%) had 9 years of experience or more. Twelve participants (57.1%) were residents, and nine (42.9%) were board-certified dermatologists.

Grading of PASI components according to physician experience

The expert graded the photographs as follows: erythema, 3; thickness, 3; and scaling, 3. Before training, the residents graded the photographs as 2.58 ± 0.67 for erythema, 2.33 ± 0.65 for thickness, and 2.25 ± 0.62 for scaling. The dermatologists' pretest grades were a little closer to the expert's grading in terms of mean value (2.89 ± 0.60,

2.33±0.50, and 2.22±0.44 for erythema, thickness, and scaling, respectively). However, no statistically significant differences were observed between the groups (Table 1). The results of retesting or testing with reference photographs also showed no statistical difference between the two groups (Table 1). This indicates that in the PASI-naive

Table 1. Comparisons of psoriasis severity grading among dermatology residents and board-certified dermatologists before and after photographic psoriasis area and severity index (PASI) training

Performer	Erythema	Thickness	Scaling	Area (%)
Reference severity				
PASI expert	3	3	3	15
Before training				
Residents	2.58±0.67	2.17±0.58	2.17±0.58	20.83±9.73
Dermatologists	2.89±0.60	2.33±0.50	2.22±0.44	23.67±9.41
p-value	0.31	0.602	0.602	0.554
After training				
Residents	3.00±0.43	2.33±0.65	2.25±0.62	15.42±5.82
Dermatologists	2.67±0.71	2.44±0.53	2.00±0.71	20.00±7.07
p-value	0.247	0.554	0.464	0.169
With reference photograph				
Residents	3.33±0.49	2.25±0.62	2.92±0.29	Not checked*
Dermatologists	2.89±0.33	2.44±0.53	2.67±0.71	
p-value	0.129	0.382	0.602	

The severity grades of erythema, thickness, and scaling range from 0 to 4. *Area was not rechecked because exact reference photographs for area are not available in the training lecture slides.

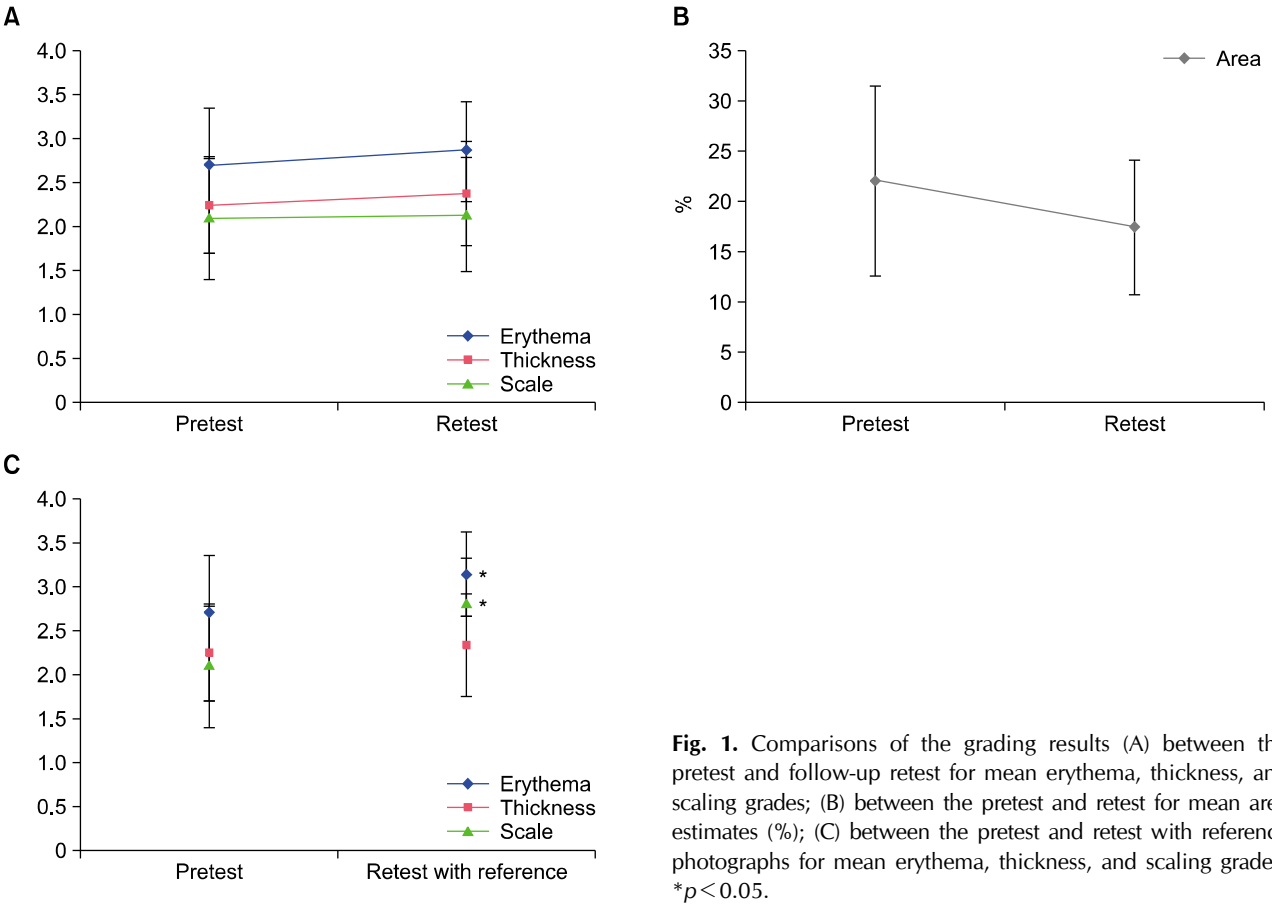

Fig. 1. Comparisons of the grading results (A) between the pretest and follow-up retest for mean erythema, thickness, and scaling grades; (B) between the pretest and retest for mean area estimates (%); (C) between the pretest and retest with reference photographs for mean erythema, thickness, and scaling grades. *$p<0.05$.

physicians, PASI assessment skill does not depend on clinical dermatology experience.

Changes in proficiency after PASI training

Two paired t-tests were performed. When the pretest PASI component grading was compared with retest grading, only the area component showed a significant change (Fig. 1A, B). The estimated area was meaningfully changed from 22.05%±9.46% to 17.38%±6.64%. Estimating the area of guttate or small-plaque psoriasis is always difficult for naive physicians. Training improved the area estimation because the initial grading was far removed from the correct answer. Comparing the results of pretesting and retesting with photographs, we observed a meaningful change in the erythema and scaling gradings, both of which neared the expert's grading (Fig. 1C).

Concordance rate with the expert's grading

In the pretest, the concordance rate between the subjects' and expert's gradings was highest for thickness (0.67), followed by scaling (0.29) and erythema (0.38), and lowest for the average area (0.10; Table 2). After 1 hour of training, the concordance rates of scaling and area improved to almost 100%. However, the erythema concordance rate decreased from 0.38 to 0.24. The concordance rate for thickness did not change after training but remained highest in the retest. In the retest with reference photographs, the concordance rate improved in the three categories of erythema, thickness, and scaling. The concordance rate for scaling was the highest (0.86), and the greatest improvement in the concordance rate was achieved in scaling (296.6%; from 0.29 to 0.86), followed by erythema (200%; from 0.38 to 0.76). The concordance rate for thickness did not change even with the use of reference photographs.

Distribution of deviations from the expert's grading: inter-rater variability

As shown in Fig. 2, the deviations from the expert's grading for erythema gradually decreased. However, the erythema deviation decreased significantly only when the reference photographs were used and not after training without the reference. In the retest, only the area component showed meaningful improvement after training. Even with the reference photographs, the inter-rater variability of thickness grading did not change after training. It is paradoxical that thickness grading showed the highest concordance rate, along with the highest deviation from expert grading. We can conjecture that the training and testing were performed by using two-dimensional photographs, in contrast to the three-dimensional clinical setting. Scaling grading dramatically improved with the reference photographs, indicating that the amount of scaling expressed in common two-dimensional photographs conveys sufficient information for proper scaling grading.

DISCUSSION

The PASI assessment, body surface area (BSA) measurement, investigator's global assessment, and dermatology life quality index are the main subjective tools used in the assessment of psoriasis severity. Among these, the PASI is the most commonly used assessment method[3,4]. However, the PASI has some limitations such as its high degree of variability in assessing BSA, limited sensitivity for evaluating mild psoriasis, and difficulty of use in clinical practice

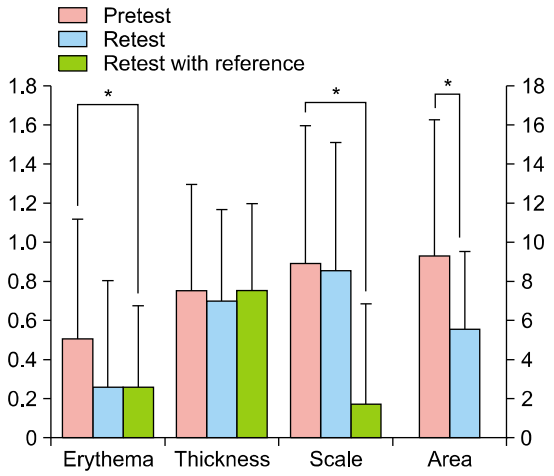

Fig. 2. Deviations from the expert's grading for the psoriasis area and severity index assessment components in three tests. The left side scale is the y-axis of erythema, thickness, and scaling (grading), and the right side scale is the y-axis of the area (%). *$p<0.05$.

Table 2. Concordance rates with the expert's grading for the pre-test, re-test after one hour of training, and re-test using psoriasis area and severity index score guiding photographs

	Erythema	Thickness	Scaling	Area
Pretest	0.38	0.67	0.29	0.10
Retest	0.24	0.67	0.57	0.19
Retest with guiding photograph	0.76	0.71	0.86	Not checked

Table 3. The most-difficult-to-grade psoriasis area and severity index component

	Erythema	Thickness	Scaling	Area
Study participants	0	2 (9.5)	0	19 (90.5)

Values are presented as number (%).

due to the complexity involved in calculating the PASI[1,3]. PASI scores by untrained evaluators have been reported to show high variability, especially in terms of BSA overestimation, because the assessments were based on visual evaluation[3,7,8]. In our study, nearly all of the participants agreed that the area component is the hardest to evaluate accurately (Table 3).

Overcoming the limitations of PASI assessment necessitates education and training to decrease inter-rater and intra-rater variability, especially in clinical trials. Some attempts have been made to improve PASI assessment standardization, and in-person training courses by a PASI-expert physician were provided regularly. Recently, PASI training using videos was introduced as well[3]. In the video training study by Armstrong et al.[3], training was shown to improve the accuracy of PASI scoring. However, in our study, only minute improvement was observed after the lecture on PASI scoring. We could not rule out the effect of the difference in training materials. The participants in our training were mostly young residents or young board-certified dermatologists who were naive to PASI assessment. Even considerable dermatologic practice experience without PASI training could not improve the PASI assessment concordance rate. This indicates that PASI training is essential to regular PASI assessment practice for any dermatologist who has no or only limited experience with PASI assessment.

We analyzed training efficacy according to the following four PASI assessment components: erythema, thickness, scaling, and area. Before the study, we hypothesized that these four components would differentially affect inter-rater and intra-rater variability. In our results, erythema showed a low concordance rate with little improvement after training but showed marked improvement when using the reference photographs. This means that in order to standardize the erythema component, a reference photograph is needed when conducting multicenter clinical trials with a large number of investigators. The thickness component showed the highest concordance rate, without change in concordance even after training or using reference photographs. We can conjecture that this may have been due to differences in the training environment, which could not simulate three-dimensional psoriatic lesions. The scaling component showed a low initial concordance rate with a slight improvement after simple training, but dramatic improvement with the use of reference photographs. Like erythema, the evaluation of scaling also requires reference photographs to improve evaluator concordance rates in multicenter trials. Area is always the most difficult PASI component to assess, even for highly trained experts. In Far East Asia, small-plaque psoriasis predominates the large-plaque psoriasis that is common in Caucasians. The numerous small papules or plaques exhibited by patients with small-plaque psoriasis are responsible for the low concordance rate of the area assessment. With the aid of the PASI BSA calculator application for iPAD (http://itunes.apple.com/app/id514524967, Janssen Korea, Seoul, Korea), the area can be estimated according

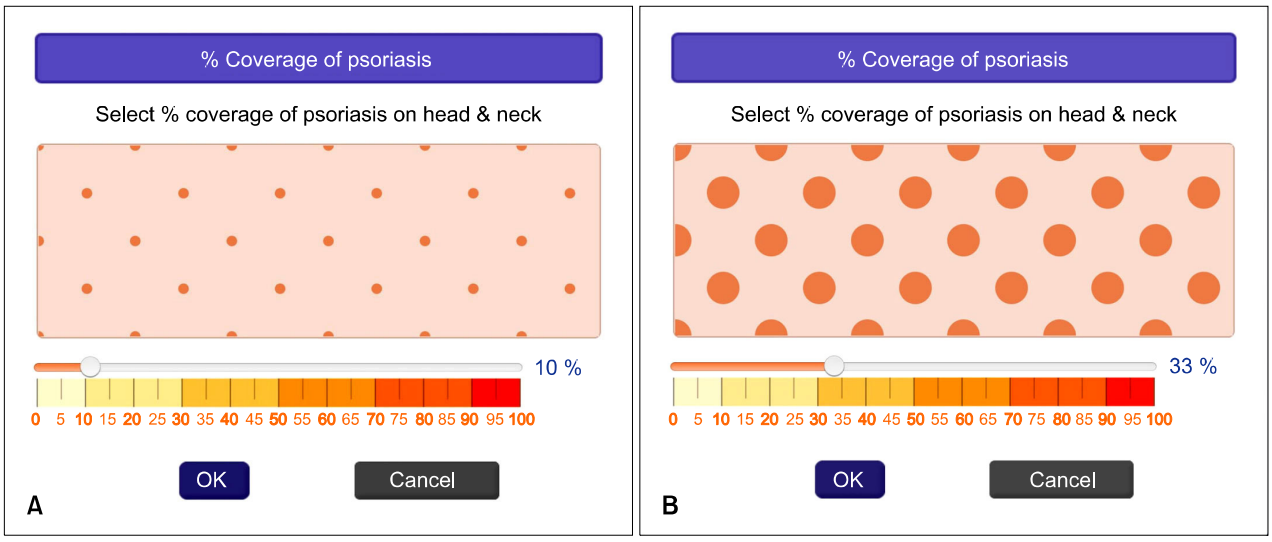

Fig. 3. Coverage area (%) estimation module for psoriasis area and severity index body surface area calculation. When we slide the control bar under the rectangle, the area (%) covered by the circles in the main rectangle changes. The sum of the circle areas represents (A) 10% and (B) 33% coverage of the rectangle. To estimate the area covered by small plaques distributed across a body region, the approximate area (%) can be obtained by sliding the control bar to display a size similar to that of the patient's lesion in the main rectangle.

to the coverage of circles with controllable diameters in a rectangle (Fig. 3). The area estimation module in this application is especially useful for countries where most psoriasis patients have small plaque psoriasis instead of large plaques that can be measured with relative ease. This application also presents reference photographs for the other three PASI assessment elements.

Another confounding factor in PASI assessment is the absence of a standard for the grading of the severity of lesions with variable characteristics mixed in a single region. This factor also affects the variability of PASI assessment. A generally accepted rule for selecting a lesion for grading should be made to minimize variability. In actuality, even with vigorous PASI training, we could not eliminate the variability inherent to subjective grading. Objective assessment tools for psoriasis could be an alternative[2]. Bioengineering devices would allow anyone to measure psoriasis severity without inter-rater or intra-rater variability. In the past, more than three devices for the measurement of color, elasticity, and scaling were used to determine bioengineering parameters that describe psoriasis severity[9]. However, based on the linear regression analysis, only a colorimeter with an integrated equation made objective assessment possible without expensive equipment[10]. Attempts have been made to correlate histopathological severity with PASI score[2].

In conclusion, PASI assessment training is essential for decreasing scoring variabilities to some extent. Reference photographs can improve the concordance rate with the ideal PASI score. The use of tablet PC applications with reference photographs would help decrease the variability of PASI assessment. Developing and introducing an objective psoriasis severity assessment tool will circumvent the limitations of the subjective PASI assessment tool. Although the PASI score has many innate drawbacks, these efforts to minimize variability can diminish the limitations of subjective grading.

REFERENCES

1. Robinson A, Kardos M, Kimball AB. Physician Global Assessment (PGA) and Psoriasis Area and Severity Index (PASI): why do both? A systematic analysis of randomized controlled trials of biologic agents for moderate to severe plaque psoriasis. J Am Acad Dermatol 2012;66:369-375.
2. Kim BY, Choi JW, Kim BR, Youn SW. Histopathological findings are associated with the clinical types of psoriasis but not with the corresponding lesional psoriasis severity index. Ann Dermatol 2015;27:26-31.
3. Armstrong AW, Parsi K, Schupp CW, Mease PJ, Duffin KC. Standardizing training for psoriasis measures: effectiveness of an online training video on Psoriasis Area and Severity Index assessment by physician and patient raters. JAMA Dermatol 2013;149:577-582.
4. van de Kerkhof PC. On the limitations of the psoriasis area and severity index (PASI). Br J Dermatol 1992;126:205.
5. Gelfand JM, Weinstein R, Porter SB, Neimann AL, Berlin JA, Margolis DJ. Prevalence and treatment of psoriasis in the United Kingdom: a population-based study. Arch Dermatol 2005;141:1537-1541.
6. Faria JR, Aarão AR, Jimenez LM, Silva OH, Avelleira JC. Inter-rater concordance study of the PASI (Psoriasis Area and Severity Index). An Bras Dermatol 2010;85:625-629.
7. Ramsay B, Lawrence CM. Measurement of involved surface area in patients with psoriasis. Br J Dermatol 1991;124:565-570.
8. Speight EL, Essex TJ, Farr PM. The study of plaques of psoriasis using a scanning laser-Doppler velocimeter. Br J Dermatol 1993;128:519-524.
9. Choi JW, Kwon SH, Youn JI, Youn SW. Objective measurements of erythema, elasticity and scale could overcome the inter- and intra-observer variations of subjective evaluations for psoriasis severity. Eur J Dermatol 2013;23:224-229.
10. Choi JW, Kim BR, Choi CW, Youn SW. One device, one equation: the simplest way to objectively evaluate psoriasis severity. J Dermatol 2015;42:154-158.