Global Assessment of Psoriasis Severity and Change from Photographs: A Valid and Consistent Method

David Farhi[1,2,3], Bruno Falissard[1,4,5] and Alain Dupuy[6,7]

Five raters tested the validity and consistency of global assessments of severity and change from standardized photographs in 30 consecutive patients with plaque psoriasis. The main outcome measures were physician global assessment (PGA) scores for change between baseline and follow-up visits ("dynamic PGA") and for severity at the baseline visit ("static PGA"). These photographic evaluation scores were compared with in-person clinical ratings. Panel ratings were obtained using the mean of the five raters' independent evaluations from photographs. Validity and consistency were assessed with intra-class coefficients (ICCs; 95% confidence interval). Intra-rater and intra-panel consistencies for photographic dynamic PGA scores were 0.85 (0.74–0.92) and 0.95 (0.92–0.99), respectively. As an evaluation of validity, agreement between photographic and clinical static PGA scores was 0.87 (0.75–0.93). We concluded that global assessment of psoriasis severity and change from photographs by a panel of experts was accurate and consistent. The generalizability of the results requires further studies. The intrinsic limitations of photographic assessment of individual characteristics such as plaque thickness and their effect on global photographic assessment should be further evaluated.

Journal of Investigative Dermatology (2008) **128,** 2198–2203; doi:10.1038/jid.2008.68; published online 17 April 2008

INTRODUCTION

A therapeutic effect is generally assessed by evaluating change in severity between a baseline and a final state. In everyday practice, dermatologists assess change by recalling the baseline state while assessing the final state on clinical criteria during the follow-up visit. This quick global assessment of change cannot be used in rigorous clinical research, however, because the baseline and final states are not assessed from the same material (memory versus actual patient examination). To circumvent this drawback, clinical researchers have developed severity scores. In this case, change is measured by computing the difference between the final and baseline scores. One of the most widely used scores in Dermatology is the psoriasis area severity index (PASI). Its validity and reproducibility have, however, been challenged (Ashcroft *et al.*, 1999; Langley and Ellis, 2004; Feldman and Krueger, 2005; Finlay, 2005), and the clinical meaning of score values is obscure for many physicians (Feldman and Krueger, 2005).

Using photographs to assess change is appealing because of two theoretical advantages: both states are evaluated in the same way, and change can be directly assessed rather than being calculated, and thus could be more meaningful for clinicians. In this article, a pilot study on validity and reproducibility of a global assessment of severity and change conducted on standardized photographs is presented.

RESULTS

The clinical description of the patients is presented in Table 1. Baseline and follow-up clinical PASI scores are presented for each patient in Figure 1.

Consistency among experts for photographic assessment of change

Intra-rater consistency was assessed for each of the five experts, using their ''photographic dynamic physician global assessment (PGA)'' scores at the ''test'' and ''retest'' sessions. Four of the five experts (80%) had an intra-class coefficient (ICC) >0.80 (range: 0.71–0.93). Overall, the mean intra-rater consistency was 0.85 (95% CI: 0.74–0.92).

Inter-rater consistency was assessed using the five ''photographic dynamic PGA'' scores. The inter-rater consistency was 0.73 (95% CI: 0.56–0.87).

Consistency for photographic assessment of change by the panel

When using the mean of the five experts' scores as the synthetic assessment from the panel, that is, ''panel photographic dynamic PGA'' scores, intra-panel consistency ICC between ''test'' and ''retest'' sessions was 0.95 (95% CI: 0.92–0.99).

[1]*Inserm, U669, Paris, France;* [2]*Université Paris Descartes, Paris, France;* [3]*Département de Dermatologie, AP-HP, Hôpital Cochin, Paris, France;* [4]*Université Paris-Sud and Université Paris Descartes, UMR-S0669, Paris, France;* [5]*Département de Santé Publique, AP-HP, Hôpital Paul Brousse, Villejuif, France;* [6]*Université Paris 7 Denis-Diderot, Paris, France and* [7]*Département de Dermatologie, AP-HP, Hôpital Saint-Louis, Paris, France*

This study was performed in Paris, France

Correspondence: Dr Alain Dupuy, Department of Dermatology, Hôpital Saint Louis, AP-HP, 1 avenue Claude Vellefaux, Paris 75010, France. E-mail: alain.dupuy@sls.aphp.fr

Abbreviations: CI, confidence interval; PASI, psoriasis activity and severity index; PGA, physician global assessment; ICC, intra-class coefficient

Received 29 July 2007; revised 20 January 2008; accepted 7 February 2008; published online 17 April 2008

Table 1. Clinical description of 30 patients with chronic plaque psoriasis

Characteristics	*n* (%)[1]
Female gender	12 (40)
Median age (range)	42 (19–74)
Past treatments	
Hospitalization	7 (23)
Classic systemic treatment[2]	17 (57)
Biologics	8 (27)
Present treatments	
Classic systemic treatment[2]	6 (20)
Phototherapy	15 (50)
Biologics	1 (3)
Topical treatments[3]	8 (27)
Skin color	
White	26 (86)
Black	2 (7)
Asian	2 (7)
Median baseline PASI (range)	6.9 (2.1–32.7)

[1]Unless otherwise specified.
[2]Retinoid, methotrexate, cyclosporin.
[3]Steroids, vitamin D derivatives, and emollients.

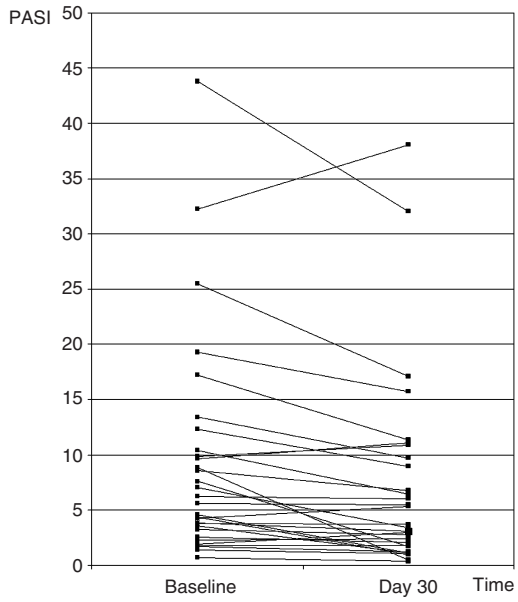

Figure 1. Change in severity during study. Baseline and follow-up (day 30) clinical PASI scores are presented for each patient.

Using the Spearman–Brown formula, the predicted inter-panel consistency was above 0.90 for a panel composed of four or more experts: inter-panel ICCs with 2, 3, 4, and 5 experts in the panel were respectively 0.84 (95% CI: 0.71–0.91), 0.88 (95% CI: 0.79–0.94), 0.91 (95% CI: 80–95), and 0.93 (95% CI: 0.86–0.97). This result suggests that the increase of inter-panel consistency gained by increasing the number of experts in the panel is marginal for panels of more than four experts.

Consistency for photographic assessment of severity

In the same way as for assessment of change, we tested consistency for severity scores, using each single rating by the experts (intra-rater and inter-rater consistency of photographic static PGA) or the mean rating by the panel (intra-panel consistency of photographic static PGA).

Intra-rater consistency was assessed for each of the five experts, using their ''photographic static PGA'' scores at the ''test'' and ''retest'' sessions. Four of the five experts (80%) had intra-rater ICC > 0.80 (range: 0.66–0.94). The mean intra-rater consistency ICC was 0.84 (95% CI: 0.78–0.90).

Inter-rater consistency was assessed using the five ''photographic static PGA'' scores. The inter-rater consistency ICC was 0.80 (95% CI: 0.68–0.89).

When using the mean of the five experts' scores as the synthetic assessment from the panel, intra-panel consistency was excellent: ICC = 0.95 (95% CI: 0.92–0.99).

Validity of photographic assessment of severity

Agreement ICC between panel photographic static PGA scores and clinical static PGA scores was 0.87 (95% CI: 0.75–0.93). Agreement ICC between clinical and photographic Delta-static PGAs was 0.64 (95% CI: 0.51–0.79).

DISCUSSION

Our study has shown that global assessment of psoriasis severity and change from photographs by a panel of experts is an accurate and consistent method. First, there is the excellent agreement between clinical assessment and the panel's photographic assessment when the same scale was used. Second, the study has shown the good intra-rater and inter-rater consistency of photographic assessments of change. Third, the number of experts can be restricted to as few as five to obtain consistent and valid estimations.

Apart from the demonstrated metrological qualities, the use of photography presents several advantages—whether for daily practice or for clinical research—that should be emphasized: (1) *communicability*: digital photography data can be easily and swiftly transmitted to peers; (2) *transparency*: digital photography allows easy storage and subsequent monitoring of the data on disease severity, an advantage that may improve *post hoc* evaluations; (3) *better blind assessment*: during clinical trials, the clinical evaluation of a patient may provide verbal or nonverbal (including behavioral) clues—for instance, the occurrence of side effects—that may facilitate investigators' deduction of the treatment group; photography tends to lessen this risk of information bias; (4) *independence of assessments*: in clinical research trials, the

use of photographs avoids the risk of communication between a given investigator and either patients or peers, and therefore preserves the independence of investigators' assessments; (5) *homogeneity*: in a multicenter or multi-investigator trial, photographic assessment can be centralized, allowing all subjects to be evaluated by the same panel of judges. In a clinical research perspective, these advantages reduce both variability (centralization of evaluations) and bias (''blinded'' evaluations). Meta-analytical evaluations can also provide insight by reassessing efficacy homogeneously from photographic archives of patients from several trials.

This pilot study was conducted with a single photographer and a small number of patients. Sample size was calculated to evaluate the internal validity of our photographic method. The applicability of the whole process and the scope for generalization of our findings should be discussed, however. Photographer training and workload for taking pictures (5 minutes per patient) were minimal, but admittedly more time consuming than performing a simple clinical PGA. The technical quality of the photographs was good for all patients. The intervention of several photographers could, however, lead to greater variability in quality of pictures. Perfect standardization leaving no room for personal initiative is the key to address this point. Being photographed proved acceptable to all patients, but it should be noted that our patients were made aware of the constraints when they agreed to participate. Also, our small sample size did not allow us to study our photographic assessment method across various skin types. As this pilot study has been completed, we have implemented a large multi-investigator study to test acceptability, applicability, technical quality, and performance in subgroups of patients in a more accurate manner.

We studied a group of patients presenting different treatments and evolutions during the 1-month follow-up. This was intended to reflect real-life conditions. Large inter-patient variability tends to increase ICC values, however. Conversely, estimating the validity of our photographic method by computing the difference between final and baseline static PGAs led to lower ICC values, because computing a difference on a narrow scale can only capture obvious changes and not more subtle ones. Assessment of validity is always difficult to undertake when the variable to evaluate is a subjective one. This is a particularly prominent problem in Dermatology (Chren, 2000). Assessment of validity should ideally refer to an unequivocal gold standard instrument, but such reference is rarely available. Although change in clinical PASI scores is used as the main outcome efficacy measure in clinical trials, the PASI score has been criticized for its inaccuracy in patients with low PASI values (that is, under 10). Also, because PASI and PGA scores are different scales with different ranges, clinical PASI score could not be used as a reference to test validity of the PGA-based photographic scoring.

There are intrinsic limitations in using photographs to assess severity and change. In psoriasis, thickness, for example, is certainly difficult to appreciate on photographs. The consequences of this limitation on a global photographic evaluation of severity and change were not specifically addressed in this study. Another limitation is the noncomprehensiveness of the skin area photographed. The nine poses selected for our study covered about 95% of the body surface area, but important areas such as the scalp were not evaluated. We believe that adding poses would increase validity but would reduce applicability.

In conclusion, we present a pilot study on the photographic assessment of change and severity in psoriasis. Although photography has long been used in Dermatology, photographic evaluations have been made easy to implement only since the development of digital photography. New possibilities offered by this technical and economic breakthrough need to be explored. This proof-of-principle study yielded encouraging results regarding validity and consistency in psoriasis patients. Photographic assessment using standardized photographs might be extended in the future beyond psoriasis toward a large variety of skin conditions.

MATERIALS AND METHODS

Patients

Between December 2005 and February 2006, 51 consecutive adult patients being treated for psoriasis were approached in the outpatient and phototherapy clinics on one specified day each week. Inclusion criteria were as follows: (1) clinical diagnosis of plaque psoriasis of any duration; (2) agreement to be photographed at two visits 1 month apart. None was reluctant to have standardized pictures taken. Thirty patients agreed to participate. Severity was assessed clinically at both visits using standard scales (see ''Scales and measurements''); change between baseline and follow-up was assessed at the follow-up visit 1 month later. A standardized set of photographs was taken at each visit. The study was conducted in accordance with the Declaration of Helsinki Principles. According to the French law (*Code de Santé Publique, article L1121-1*), an authorization from the Ethics Committee was not required for this type of study. Patients signed an informed consent form stating that they agreed to have their photographs taken, analyzed, and included in a scientific publication.

Standardization of photographs

We used a nonprofessional 8-million pixels Canon Eos 350D ''Rebel'' digital camera with a fixed 35 mm focal length Canon lens. Settings were standardized: manual mode, autofocus on, shutter speed of 1/50 seconds, diaphragm aperture of f5.6, 400 ASA sensitivity, JPEG format, integrated camera flash on. Pictures were taken in a room with artificial neon light. Patients were placed in front of a light-blue papered wall. A set of nine standardized poses was adapted from Halpern *et al.* (2003). These are presented in Figure 2. Time for taking a full set of photographs was about 5 minutes. Men wore different styles of underwear, but were routinely asked to roll up their underwear to increase the visible surface of their buttock. Women were asked to remove bras consistently. A full set of nine photographs were obtained for all 30 patients. Photographs were taken by a single dermatologist (DF), without formal training.

Presentation of photographs

Unedited JPEG pictures of each patient were transferred into an Adobe Portable Document Format file. Each file was made up of nine A4 format sheets in landscape orientation. Each sheet presented

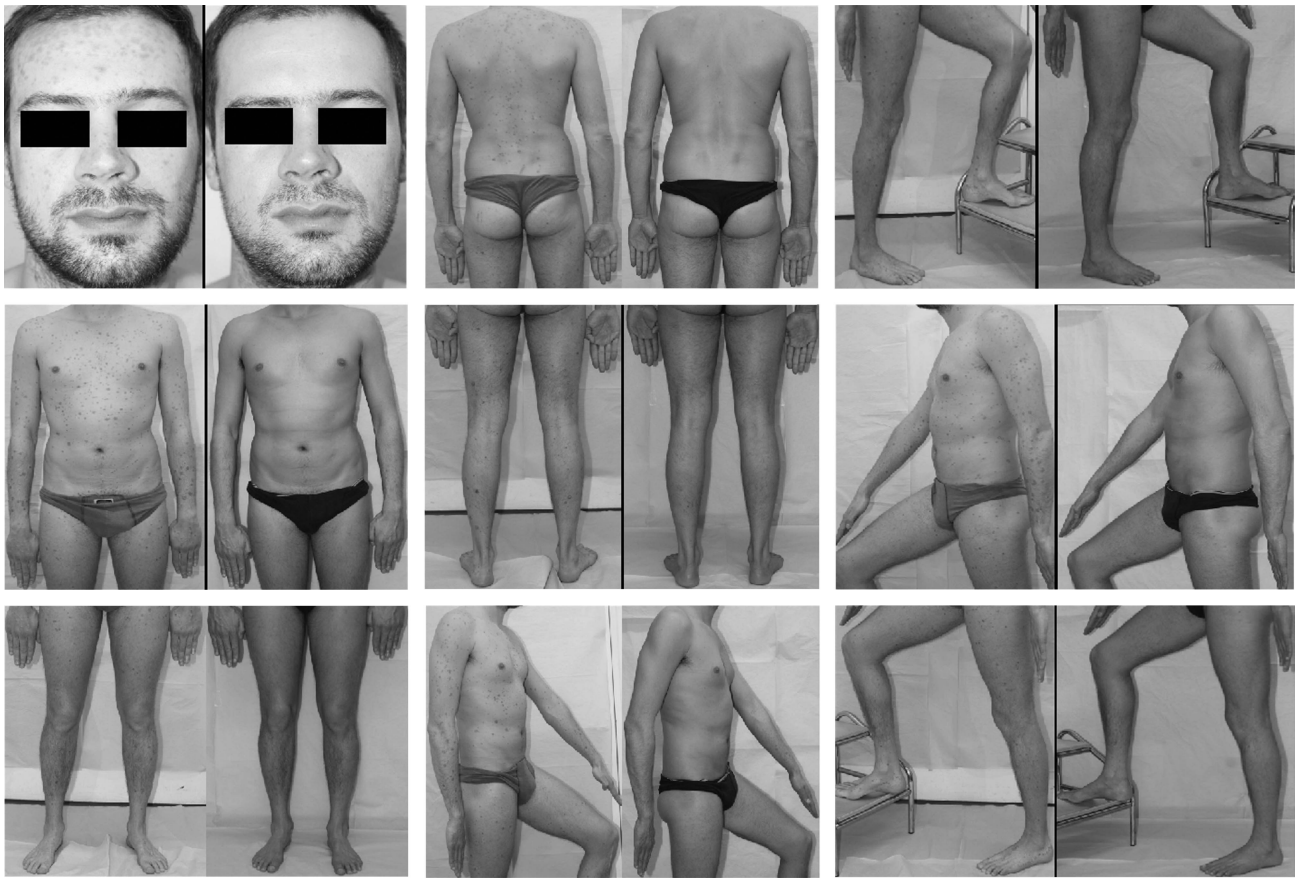

Figure 2. Nine standardized photographs applied twice at 30-day intervals to 30 patients with chronic plaque psoriasis. In this patient, mean static photographic PGA scores from the five experts was (4.0) at baseline (individual scores: 3, 4, 4, 4, 5) and 0.8 at day 30 (individual scores: 0, 1, 1, 1, 1); mean dynamic photographic PGA score from the five experts was 4.2 (individual scores: 1, 5, 5, 5, 5).

two pictures of the same pose: the baseline picture on the left-hand part and the follow-up picture on the right-hand part.

Scales and measurements

We firstly aimed to focus on global assessment of change between two sets of photographs taken one month apart in psoriasis patients. Assessments were performed by a panel of five experts. Each expert rated change for each patient using the two sets of photographs. We have used the standard term of ''dynamic PGA'' to refer to this scale, and to specify that the assessment was made from photographs, the term ''photographic dynamic PGA'' is used hereafter.

Secondly, together with the assessment of change, severity was also scored. For this purpose, a scale known as the ''static PGA'' was used. When assessed from photographs, the score is referred to below as the ''photographic static PGA'' (severity), to distinguish from the ''dynamic'' (change) score. When the clinician assessed the patient rather than the photographs during a visit, the severity score is referred to as the ''clinical static PGA''. As there was only one clinician involved in the clinical part of the study, there was only one ''clinical static PGA'' score for a given patient at the baseline visit and another at the follow-up visit, whereas there were five experts rating from photographs, yielding five different ''photographic static PGA'' scores for the baseline visit and five for the follow-up visit, for each patient.

To obtain a synthetic assessment from the panel of the five experts, for each patient, the mean of the five ratings was computed. These scores are referred to as the ''panel photographic dynamic PGA'' (change) and ''panel photographic static PGA'' (severity). The scales are presented in Figure 3. A scoring summary is presented in Table 2.

Assessors and sessions

Photographs were independently assessed on computer screens by a panel of five senior dermatologists. They had at least 4 years' experience in dermatology, were practicing in an academic hospital, and three of them had been specialized in psoriasis management for more than 3 years. All photographic assessments were blinded to clinical data. Raters were explicitly requested to use the scales in the following order: ''dynamic PGA'', then ''static PGA'' for the baseline set, and finally for the follow-up set. Two similar rating sessions were organized at a 1-month interval (''test'' and ''retest'' sessions) on the same set of the photographs of the 30 patients, by the same five experts. Between the ''test'' and the ''retest'' sessions, raters did not look at the photographs, did not reread their first scorings, and did not exchange any information about the study. This ''test–retest'' design was used to evaluate intra-rater consistency (see ''Consistency'' below).

a

Dynamic PGA:
In each slide, photographs on the left represent baseline, photographs on the right were taken 1 month later. Rate the *change* of psoriasis severity from –5 (considerable deterioration) to +5 (considerable improvement), for each patient, using Dynamic PGA:

(+5) Very large improvement/cleared (+90 to +100%)
(+4) Large improvement (+70 to 89%)
(+3) Moderate to large improvement (+50 to 69%)
(+2) Moderate improvement (+30 to +49%)
(+1) Mild improvement (+10 to +29%)
(0) No or minimal change (–10 to +10%)
(–1) Mild deterioration
(–2) Moderate deterioration
(–3) Moderate to large deterioration
(–4) Large deterioration
(–5) Very large deterioration

b

Static PGA:
Rate psoriasis severity at *each time point*, for each patient, using Static PGA: first, for all left-hand photographs (baseline), then for all right-hand photographs (taken 1 month later).

(6) Severe psoriasis
(5) Moderate to severe psoriasis
(4) Moderate psoriasis
(3) Mild to moderate psoriasis
(2) Mild psoriasis
(1) Psoriasis almost cleared
(0) Clear (no lesion)

Figure 3. (**a**) Dynamic PGA and (**b**) Static PGA, as presented to assessors in our study.

Table 2. Scoring

	Clinical assessment	Photographic assessment
Change (between baseline and follow-up visits)	Not assessable[1]	Photographic dynamic PGA (× 5) Panel photographic dynamic PGA (x1)
Severity		
Baseline	Clinical static PGA (× 1)	Photographic static PGA (× 5) Panel photographic Static PGA (× 1)
Follow-up	Clinical static PGA (× 1)	Photographic static PGA (× 5) Panel photographic static PGA (× 1)

Each of the five raters composing the panel was blinded to clinical data, and independent to other raters. For the appraisal of intra-observer consistency, all photographic assessments were repeated at a second (''retest'') session. ''(× *n*)'' refers to the number of results for the variable.
[1]Comparison of two clinical states in a given patient cannot be made simultaneously and can only rely on memory of the baseline state when seeing the patient at the follow-up visit.

Statistical analysis

Intra-class coefficients. ICCs were used to assess both validity and consistency (Muller and Buttner, 1994). For ''static PGA'' scales, only scores for baseline visits were used.

ICCs were estimated by the percentage of total variance that results from patient effect, using a two-way random effects analysis-of-variance model (Muller and Buttner, 1994):

$$ICC = \sigma^2_{patient}/(\sigma^2_{patient} + \sigma^2_{rater} + \sigma^2_{residual})$$

Generally, agreement can be qualified as ''almost perfect'', ''substantial'', or ''moderate'', for ICC values in the (0.80–1.0), (0.60–0.80), and (0.40–0.60) ranges, respectively (Landis and Koch, 1977). Agreement for ordered categorical variables can also be estimated with weighted Kappa statistics (Falissard, 2001). Weighting is used to penalize large disagreements (for example, ''1'' vs ''5'') more strongly than small ones (for example, ''1'' vs ''2''). There is no consensus on which weight should be used, 1/N and $1/N^2$ being the most widely used. Because $1/N^2$-weighted Kappa statistics lead to the same values (± 1%) as ICCs, only results for ICCs are presented here.

All analyses were conducted with R software, release 2.3.0 (R Development Core Team, 2006). The use of an analysis-of-variance model for 95% confidence interval (CI) estimation and ICC comparisons would be based on the hypotheses of normality and homoscedasticity. These hypotheses were not confirmed by checking ICC graphical distributions in our data set. Therefore a nonparametric bootstrap method was implemented to estimate ICC 95% CI and to compare ICCs. All test formulations were two tailed. $P \leqslant 0.05$ was defined as the significance threshold. ICCs are presented with 95% CI.

Consistency of photographic assessment

Consistency of experts' scores from photographs. Consistency (or reliability) is the ability of a measurement process to yield the same result when the measurement process is repeated by the same observer (intra-rater consistency) or by another observer (inter-rater consistency) (Feinstein, 1987). For each expert, intra-rater consistency was estimated by the ICC between the ''test'' and the ''retest'' session measurements.

Assessment of inter-rater and intra-rater consistency was also conducted with a Bland and Altman graphical method (Bland and Altman, 1986); this analysis corroborated the results yielded by the ICC analysis (data not shown).

Consistency of panel ratings from photographs. Intra-panel consistency was assessed by computing the ICC between the ''test'' and ''retest'' sessions for the 30 patients. Inter-panel consistency could not be directly assessed, as only one panel of five experts was available. We estimated inter-panel consistency with the Spearman–Brown formula (Lord and Novick, 1968): the inter-panel consistency for the ''panel photographic dynamic PGA'' was estimated for different numbers of experts in the panel.

Validity of photographic assessment

Validity is defined as how closely the result of a tested procedure conforms to the results obtained with a reference procedure (''Is there agreement between the results of the tested tool and the reference tool?''). This concept has been also termed accuracy (Feinstein, 1987). The score of interest in our study was the ''panel photographic dynamic PGA''. It was not possible to compare ''panel photographic dynamic PGA'' against a reference clinical score, as there is no such reference clinical score: simultaneous comparison of two 1-month-apart states cannot be made, and memorization of

baseline state at the time of follow-up evaluation cannot be considered reliable enough to be used as a reference clinical score. Therefore, we used static rather than dynamic assessments to determine validity. The validity of photographic assessment was ascertained by assessing agreement for two comparisons: (1) agreement between ''panel photographic static PGA'' and ''clinical static PGA''; (2) agreement between ''clinical'' and ''photographic'' Delta-static PGA, Delta-static PGA being defined as ((follow-up static PGA score) − (baseline static PGA score)). ICCs were used to assess the agreement between scores.

Sample size calculation

Sample size calculation was based on the required precision of intra-rater ICC estimations. It has been shown that ICCs estimate precision increases, that is, the ICC 95% CI narrows with the number of patients and/or assessors. Under the normal hypothesis, for a given ICC, the range of the ICC 95% CI can be determined by the number of patients and assessors included. Previous studies have shown that, in terms of 95% CI precision, the benefit of including more than five assessors is small (Giraudeau and Mary, 2001; Bonett, 2002). In addition, an intra-rater ICC value of 0.80 for photographic dynamic PGA was expected, and we decided that a 95% CI span of 0.20 would be satisfactory. Thus, using a previously published approximation (Bonett, 2002) of the mathematical relationship between ICC, 95% CI width (w), number of patients (n), and number of assessors (k), the sample size required for this study was estimated: if ICC = 0.80, $w = 0.20$, $k = 5$, and $z_{\alpha/2} = 1.96$, then $n = 29$ patients. We therefore decided to include 30 patients and 5 raters.

CONFLICT OF INTEREST

Authors declare neither conflict of interest nor financial disclosure. Wyeth-France provided the funding for the camera used in this study, but did not offer any personal grant for the authors nor had any influence on the study analyses or results.

ACKNOWLEDGMENTS

We thank Drs Amélie Arsouze, Edouard Bégon, Justine Dautremer, Laurence Fardet, François Durupt, Sophie Georgin-Lavialle, Fabien Guibal, Simon Jacobelli, Annabelle Lévy, Antoine Petit, Stéphanie Régnier for their assistance, and Angela Swaine Verdier for reading and correcting the manuscript. The authors state that this work has not been published in another journal. This work has been partly presented at the *36th European Society for Dermatological Research* congress (7–9 September 2006, Paris, France), and at the *Journées Dermatologiques de Paris* congress (6–10 December 2006, Paris, France).

REFERENCES

Ashcroft DM, Wan Po AL, Williams HC, Griffiths CE (1999) Clinical measures of disease severity and outcome in psoriasis: a critical appraisal of their quality. *Br J Dermatol* 141:185–91

Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–10

Bonett DG (2002) Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 21:1331–5

Chren MM (2000) Giving ''scale'' new meaning in dermatology: measurement matters. *Arch Dermatol* 136:788–90

Falissard B (2001) La précision de mesure. In: *Mesurer la subjectivité en santé*, (Masson ed), Paris, 155–7

Feinstein AR (1987) *Clinimetrics*. New Haven: Yales University Press, 272 pp

Feldman SR, Krueger GG (2005) Psoriasis assessment tools in clinical trials. *Ann Rheum Dis* 64(Suppl 2):ii65–8

Finlay AY (2005) Current severe psoriasis and the rule of tens. *Br J Dermatol* 152:861–7

Giraudeau B, Mary JY (2001) Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med* 20:3205–14

Halpern AC, Marghoob AA, Bialoglow TW, Witmer W, Slue W (2003) Standardized positioning of patients (poses) for whole body cutaneous photography. *J Am Acad Dermatol* 49:593–8

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–74

Langley RG, Ellis CN (2004) Evaluating psoriasis with psoriasis area and severity index, psoriasis global assessment, and lattice system physician's global assessment. *J Am Acad Dermatol* 51:563–9

Lord FM, Novick MR (1968) *Statistical theories of mental test score*. Reading, MA: Addison-Wesley

Muller R, Buttner P (1994) A critical discussion of intraclass correlation coefficients. *Stat Med* 13:2465–76